NATURALLY WE

A PHILOSOPHICAL STUDY OF COLLECTIVE INTENTIONALITY

Submitted by Mattia Luca Gallotti to the University of Exeter

As a Thesis for the Degree of

Doctor of Philosophy in Philosophy

In September 2010

……..……

Abstract

According to many philosophers and scientists, human sociality is explained by our unique capacity to 'share' the mental states of others and to form collective intentional states. Collective intentionality has been widely debated in the past two decades, focusing especially on the issue of its reducibility to individual intentionality and the place of collective intentions in the natural realm. It is not clear, however, to what extent these two issues are related, and what methodologies of investigation are appropriate in each case.

In this thesis I set out a theory of the naturalization of collective intentionality that draws a line between naturaliz*ability* arguments and theories of collective intentionality naturaliz*ed*. The former provide reasons for believing in the naturalness of collective intentional states based on our commonsense understanding of them; the latter offer responses to the ontological question about the existence and identity of collective as distinct from individual intentionality. This model is naturalistic because it holds that the only way to establish the place of mental entities in the order of things is through the theory and practice of science. After reviewing naturalizability arguments in philosophy, I consider an influential research program in the cognitive sciences. On the account that I present, the irreducibility of collective intentionality can be derived from a theory of human development in scientific psychology dealing with phenomena of sociality like communication, recently refined by Michael Tomasello.

List of Contents

List of Figures

Preface and Acknowledgements

Several years have passed by since I first realized that society is the most astonishing construction of the human mind, but it is only when I was writing my master dissertation at the LSE in 2006 that I came across John Searle's *The Construction of Social Reality*. Before long I was captured by the lucidity of the ideas in *The Construction*. It is by trying to penetrate the depths of Searle's thinking, and despite the apparent ease of his writings, that I came to philosophy. I have since felt the urge to find my own way to the meaning of philosophy by observing, learning from, and sympathizing with Searle philosophizing. I hope these few words testify to the profound intellectual debt and to the importance of his influence on me.

The research presented here is the product of times and places, of which I would like first to convey some insights. After a slow and confused start, I became persuaded that the sense of a Ph.D. is not just a matter of how valuable it is as a professional training, so to speak, but of how far one can go in turning it into a *life* experience. Soon I have begun to confront the highs and lows of doctoral condition in more existential fashion, which helped me gain awareness of my potentialities as a writer and of my imperfections as a person, as well as strengthen my conviction in the virtues of willpower. For sure, I could never navigate through this alone, or without the mentorship of Francesco Guala.

Since the first time I met Francesco as a fresh graduate in economics in September 2004, before planning to embark on a Ph.D. in philosophy, I have been increasingly drawn to his moral and intellectual authority in ways that go much beyond mere academic supervision. With patience and persuasion, Francesco has succeeded in instilling in my mind the idea that the only way to find out what I think, is to write. And how effective he was in prompting me to think harder and write simpler has become all the more apparent after he left Exeter for good in 2008. The wording of this preface cannot express all my gratitude for teaching me how to do philosophy almost from scratch, and for having supervised me with intact enthusiasm and devotion in the past two years. His

reassuring presence and inspirational figure are the *sine qua non* of my work. Without doubt, I could not think of my best philosophical thoughts but in his possession – this thesis is dedicated to Francesco.

When it occurs to me how fortunate I have been to be 'raised', philosophically speaking, under the influence of John Searle and Francesco Guala, I always think of two lessons. One is that, to steal Colin McGinn's words (2002), the best way to avoid ideological bias, meaning the denial of obvious facts, is always a good deal of common sense. The other is that it is indeed possible to try to enliven the Analytic paradigm of logical rigour and conceptual clarity with some sort of 'grand-theory' Continental-style touch.

Nigel Pleasants took on a primary tutoring role after Francesco's departure from Exeter. With his gentle and subtle understanding of what it feels like to go through the write-up stage, Nigel has turned our supervisory meetings in thoughtful and enjoyable conversations. Above all, I have learned a lot from his insights on how to develop my own intimate relationship with writing. I am also extremely grateful to Paul Griffiths, who acted as my second supervisor during his yearly research stays in Exeter during fall. Paul's comments on some parts of my thesis not only largely improved their overall quality, but gave me a clue of his superb argumentative style. Inside and outside Exeter, I want to express my thanks to Alexander Powell, Mauro Rossi and Deborah Tollefsen for helpful discussions.

This research project has led a travelled life in progress, and I am grateful to the many people who contributed to enrich it over the years. The thesis started life *seriously* during my first semester as a visiting scholar at the University of California at Berkeley in 2008. I cannot fully convey the importance of those six months, which changed me philosophically, or my taste of 'Berkeley spirit' - a mixture of intellectual thrill and passion, and loneliness, which pervaded me during such a momentous phase of American politics. I owe a great debt to Jennifer Hudin, for making me always feel an important part of the growing 'family' of the *Berkeley Social Ontology Group*, which she

manages with grace and verve. Her generosity and friendship brought me back to Berkeley again in fall 2009. Many of the ideas of this thesis incubated in discussions with Jennifer and with Klaus Strelau, and probably took shape in Berkeley at the *Brewed Awakening*. Among my Berkeley friends who gave aid and advice are Sophia Krzys Acord, Camilla Bernardi, Frederick Eberhardt, Joe Landon and Jonas Schneider.

In spring 2009 I had the opportunity to visit Michael Tomasello's social cognition lab at the Max Planck Institute for Evolutionary Anthropology in Leipzig. During the stay I worked on the structure of the second part of this thesis, and I developed a strong interest in the philosophy of psychology. I am especially grateful to Mike for bringing the experimental research program on collective intentionality to my awareness, and to Henriette Zeidler for assistance throughout the stay. Having daily conversations with scientists with different background and expertise was a challenge for me to try to get more 'real' and sound less philosophical. I thank, in particular, Malinda Carpenter, Felix Warneken, Hannes Rakoczy and Federico Rossano for sharing their thoughts with me.

I'd like to thank two persons whose support was vital throughout the Ph.D. I still remember when I first came into contact with Franco Donzelli, and his precious advice when I started to figure out how to fund graduate studies in philosophy, rather than in economics. Working as the research assistant to Enrico Giovannini at the Organization for Economic Cooperation and Development (OECD) in Paris, in the summer of 2007, has significantly widened the scope of my studies and career prospects, and given me the unforgettable feel of what it means to turn ideas into action. I owe to Enrico and Franco another important lesson – drive for change is mostly a matter of self-motivation and intellectual honesty.

With a number of people I not only discussed but, most importantly, enjoyed the fact that sociality is such a *natural* ingredient of our lives. To be sure my journey would have been less sparkling without: my Morley-Road-fellows Daniele Carrieri, Samuel Jones and Pierre-Olivier Méthot, who

made the gestation period of writing-up less burdensome; Khalid Almezaini, Valeria Cinaglia, Cara George, Ana León Mejía, Michiru Nagatsu and Andrea Rota.

The support of my family is something I simply cannot, and don't want to, confine to letters. If there is one way to try to express my gratitude to zia Angela, zio Pietro and zia AnnaMaria, Jimmy and Nicoletta, it is by recognizing that my thesis is partly theirs, and that without their presence I could not make it through. Above all, my deepest debt is to my parents - 'forever' Kiko and Katy – whose dedication has no equal in the language of love, and is the key to my inner self.

To Francesco Guala

Mentor and Friend

One

Introduction

Sociality is a characteristic trait of humanity. At the core of sociality lies the human capacity to share attitudes of any kind: cognitive, like beliefs; conative, like intentions; and affective, like emotions and sensations. 'Sharing' can be given a variety of meanings and serves a variety of functions. Yet, some basic insights can be drawn from the analysis of everyday-life situations of social interaction. I see John pointing enthusiastically to the work of a famous living artist, despite his distaste for contemporary art; so I would not engage with him in the way he expects me to if I did not share an important piece of information with him: he is well-acquainted with the painter. In another context, I sit with John on the grass when he glimpses at his watch and looks frightened into my eyes. No word conveys the feeling of urgency and fear that I have experienced in a similar circumstance when I was about to be interviewed for a job, and asked him to accompany me.

As these examples show, what is needed for any two persons to interact successfully is that they understand and experience things *together*, so to speak. And such behaviour is typically underlain by the kind of 'meeting of minds' that people establish when they share a piece of information, or emotional state, about the action scene and about their personal history. Despite its intuitive strength, however, this idea has made its way into academic and public debates only in the last decades of the twentieth century. Here is how *The New York Times* columnist David Brooks has captured the fundamental shift in the way the issue is currently framed:

> Over the past 30 years, there has been a tide of research in many fields, all underlying one old truth – that we are intensely social creatures, deeply interconnected with one another and the idea of the lone individual rationally and wilfully steering his own life course is often an illusion. Cognitive scientists have shown that our decision-making is powerfully influenced by social context – by the frames, biases and filters that are shared subconsciously by those around. Neuroscientists have shown that we have permeable minds. When we watch

> somebody do something, we recreate their mental processes in our own brains as if we were performing the action ourselves, and it is through this process of deep imitation that we learn, empathize and share culture. Geneticists have shown that our behaviour is influenced by our ancestors and the exigencies of the past. Behavioural economists have shown the limits of the classical economic model, which assumes that individuals are efficient, rational, utility-maximizing creatures. Psychologists have shown that we are organized by our attachments. Sociologists have shown the power of social networks to affect individual behaviour (Brooks, 2008).

For philosophers, the key to human society is the capacity for collective intentional behaviour. 'Intentionality' is a technical term of philosophical jargon which does not mean just intending something; it stands for the capacity of the mind to be aware *of* things in the world. So, if you desire a coffee, or wonder whether you have time for it, or fear that the train leaves soon – your mind is 'directed toward' something. Desires, dreams, fear and love, doubts, and thoughts in general are all *about* objects or states of affairs in the world. Intentional behaviour is thus a form of behaviour performed with a certain 'reaching-out' attitude. What makes it collective is that the participants in a joint activity conceive of individual actions as oriented to a shared goal.

Suppose John and I come across a friend in difficulty and unwittingly offer help. It seems plausible that what we mean in order to bring about joint assistance, we do it together. In other words, it is because John and I see each other as being part of the same 'group', that we understand John doing his part, and I doing mine, only as part of our doing it together. More generally, when people gather and act as a group intentionally, the fact that they do something together implies that no member of the group does it 'on her own'. Hence, the intentionality of collective behaviour does not consist only in the instrumental fact that individuals engage in interaction, but the fact that they do so by intending and enacting things together.

This propensity to think and act as the members of a collective is called 'shared' or '*collective intentionality*'. In spite of the ease with which we commonly think of people in interaction as

capable of transcending their individuality and seeing things from a 'we-perspective', the concept of 'collective intentionality' is relatively new. It entered the philosophical scene only in the nineteen-eighties[1], when it became clear to philosophers like Raimo Tuomela, Margaret Gilbert and John Searle that there is a *conceptual* difference in the understanding of collective as opposed to individual action. The problem, then, is how to account for the specific attitude that underpins collective intentional behaviour.

In general, we can break this problem into two sets of questions. There are questions concerning the existence and the identity of collective, as distinct from individual, intentional states; and questions about the conditions for having knowledge of them. How the relation between the former (ontology) and the latter (epistemology) is conceived of has had remarkable importance in the systematization of collective intentionality theory. In the past two decades, philosophical research has mostly focused on the problem whether the concept of collective intentional behavior can be decomposed into the concepts that we already deploy in understanding individual behavior. And the answers, by and large, fall in two camps. For non-reductivists like Searle (1990/2002), there are plenty of counterexamples for the idea that thinking-as-a-group, or group-thinking, requires some primitive 'sense' of sociality. Reductivists like Bratman (1993), in the other camp, hold that all is needed to share attitudes is that the mental states of the individual agents be properly connected and supplemented with mutual knowledge. Although there are arguments in support of either way of tackling the irreducibility question, none has proved decisive in settling the question of the *nature* of collective intentional states.

The opposition between reductivists and non-reductivists hides a remarkable consensus on how best to interpret the demand for the conditions of reduction of collective to individual mental states. For thinkers like Searle, among the others, realism about collective intentionality is justified by the impossibility of individuating the conditions of reduction of collective to individual mental states in non-circular fashion. Searle concludes, therefore, that collective intentionality is a '*biological*

[1] Some of the intuitions behind the rise of collective intentionality theory can also be found in the phenomenological tradition (Schmid, 2009).

primitive phenomenon that cannot be reduced to or eliminated in favor of something else" (1995: 24; emphasis mine). Alternatively, were reduction to succeed the opposite conclusion would be warranted: 'there is no fact of the matter' for the ability of people to think as a group. But this is equivalent to saying that conclusions concerning the alleged irreducibility of collective intentionality are reached in both camps by the same method. Questions like 'Is there a fact of the matter that justifies realism about collective intentionality?' are addressed by exploring the folk-psychological attributes of collective intentionality in everyday language. And the result of this analysis is taken as 'evidence' to settle *ontological* questions concerning the place of collective intentionality in the natural realm.

Although my own sympathies are with Searle's primitivist account, in this thesis I want to take another route to investigate the problem of collective intentionality. Instead of drawing existential conclusions from an analysis of the uses of collectivity concepts in everyday discourse and social science research, I pursue a natural-scientific approach to issues of reduction. According to 'naturalism', roughly, it is up to science to establish whether collective intentional behaviour can be given a reductive explanation at some level of biological explanation. Whilst collective intentionality philosophers are all dedicated naturalists in principle, they have rarely pursued this line of research in practice. My choice to pursue this route responds to a specific motivation then: the failure to give a convincingly naturalistic account of the irreducibility of collective intentionality is likely to have profound consequences in the way the research in the foundations of society is currently undertaken. More in detail, there are two problem areas that can be enlightened by a scientific understanding of collective intentionality: the possibility of an empirical social ontology, and the process of sharing mind states that enables communication.

Social ontology is the study of how persons relate to one another and to the social facts they constitute. Social-institutional entities are constitutively created by people's intentional attitudes – beliefs about beliefs, as it is often said – towards themselves and the other members of their group. It is because people collectively intend a piece of paper to be money, for example, that money *exists* (Searle, 1995). This might be interpreted as suggesting that social ontology can, and should, be

approached by analyzing the concepts of sociality. But what matters for ontological considerations is simply that there might be something that enables individual agents to think and act collectively, and not what they *think* there is. This is why conceptual analysis cannot take you very far in the ontological investigation of social reality. This, of course, is not to say that *a priori* intuitions have not contributed important insights into the philosophy of collective action and, ultimately, social science. The point is that the question of what there is in society, including perhaps irreducible collective intentional states, is an empirical question to be settled with factual evidence.

The area of communication studies offers another example of the importance of a naturalistic theory of collective intentionality. As it is widely argued inside and outside of philosophy, communication seems to be *logically* impossible on various grounds. This is known as the 'problem of reference', the problem of how any two persons can know that they mean the same thing in communicative exchanges, be they linguistic or pre-linguistic. Nonetheless, people do communicate with success, and this is something that can easily be ascertained by noticing that miscommunication is the exception rather than the norm in everyday interaction. The literature on reference has, more or less implicitly, acknowledged that mutual understanding requires some sharing at the mental level. However, considerable less work has been done on the issue of what constitutes the relevant sharing, and how this can illuminate the problem of reference in practice.

Let us, then, assume that the naturalness of collective intentionality poses a serious challenge to classic approaches in the philosophy of society: How would a natural-scientific approach settle the question of the irreducibility of collective intentional states, exactly?

1.1 The Naturalization Route

One of the prominent doctrines of contemporary analytic philosophy, 'naturalism' is an ill-defined concept used in various philosophical and scientific circles to mean distinct though overlapping phenomena. In its most general characterization, naturalism is a label for the idea that natural science is 'the ultimate measure of things' (Sellars, 1956), and sets the best approach to

philosophical investigation. More precisely, commitment to naturalism comes in an ontological and a methodological form (De Caro and Macarthur, 2004).

The *ontological* version holds that nature, as the subject matter of the natural sciences (notably physics, but also chemistry and biology), exhausts the reality of what there is, allowing no place for 'supernatural' entities of any sort (Papineau, 2007). In spite of the fact that we might not have direct access to every entity postulated by the scientific theories, what science at its best[2] tells us is approximately true of the constituents of the natural realm, no matter whether animate or inanimate. Note that belief in this ontological picture is not justified on *a priori* grounds concerning the alleged primacy of one 'culture' over another, that is, the natural sciences as opposed to the humanities and the social sciences. Scientific research, in fact, advances by trial and error, and is pervasively affected by social-cultural habits and historical contingencies. Current theories which are highly valued by the scientific community might be overtaken in the future by more reliable ones. The commitment to naturalism is rather justified by the success of natural science in accounting for and in predicting natural phenomena; it dictates that the only real mind-independent properties that there are, in the sense of not being conditional upon our theories and descriptions of them, are those that science may discover.

The *methodological* version of naturalism focuses on the epistemological component of the relation between philosophical practice and scientific investigation. It emphasizes the commonality of aims and methods adopted in philosophy and science, by stressing the significance of an integrated inquiry into the structure of the natural order. Indeed, much of the debate about the viability of naturalism in philosophy has sprung from belief in methodological naturalism. While this seems to suggest a weaker commitment to naturalism than the ontological view, the methodological implications for the autonomy of philosophical practice are more threatening. Ontological naturalism is still compatible with a division of intellectual labor between philosophers and scientists, where the former are mostly devoted to the clarification of purely conceptual issues

[2] 'Science' is to be intended in the narrow sense of the body of the most highly confirmed and reliable theories as for explanatory and predictive power. The term does not generically refer to all theories put forward by alleged scientists.

arising from scientific discourse or in discussing the general implications of specific scientific findings. Methodological naturalism in contrast holds that, in order to come up with substantive and informative philosophical theories, one has to adhere to the language and practices of natural science. Philosophers should 'cross the line', as it is often said, and become a bit more scientists if they want to contribute substantially to the understanding of reality.

It is often highlighted that philosophical research differs from the practice of the natural sciences in substantial and inevitable respects, ranging from the level of generality of the questions addressed to the driving motivations for answering them. Nevertheless, there are good reasons for thinking that the seeming differences between philosophy and science are less sharp than one might think at first glance. For its relevance in strengthening attempts to achieve progress in knowledge of the world, the interdependence of philosophy and science broadly conceived has dominated the philosophical agenda for most of the twentieth century with deep consequences for the ongoing project of naturalizing the facts of the mental. A naturalistic theory of the mind is one that explores the continuity between its domain and that of one of several neighboring natural sciences.

One specific naturalistic project tries to give an argument for the *naturalizability* of collective intentionality. For collective intentionality to be naturalizable is for us to believe that the difference between the intentionality of the first-person plural and that of the first-person singular is a natural attribute of the world. But there is a remarkable difference between the argument that collective intentionality is naturalizable and a theory of collective intentionality *naturalized*. While the question whether there are good reasons to endorse realism about collective intentionality is a metaphysical question, it is in scientific theory and practice that we find a reliable answer to the question whether there actually is any sound explanation which succeeds in meeting the criteria for the naturalization of collective intentionality.

In order to keep these questions separate, I shall structure the thesis in two parts. In the first part, which includes chapters 2, 3 and 4, I shall discuss the philosophers' meaning of naturalization through John Searle's and Raimo Tuomela's defense of realism about collective intentionality. Although they disagree on why we believe that collective intentional states are real, *i.e.* true of the

reality that we inhabit, both theories are motivated by similar considerations and employ the same method of investigation. Two linking themes arise from these analyses of collective intentionality. The first is the logical structure of collective intentional states - what it means for people to have intentional states shared with others; the second concerns the conditions of existence and identity of collective as distinct from individual mental states. The former reflects an action-theory approach to the structure of collective intentions *qua* intentions; the latter takes into account foundational issues concerning the ontology of the mind. Overall, Tuomela and Searle, as well as most social theorists and philosophers, examine the naturalness of collective intentionality via conceptual analysis.

In the second part of the thesis, I shall focus on one theory that treats collective intentionality as a problem of empirical social ontology: the research project on the cognitive roots of sociality set up by psychologist Michael Tomasello. Drawing on the conceptual resources of collective intentionality to interpret the findings from research in primate cognition, developmental social cognition and language acquisition, Tomasello is the first scientist to engage critically with the research paradigm developed by philosophers' intuitions concerning group-thinking.

The subject of chapters 5 and 6 will be those aspects of Tomasello's theory of sociality which shed light on the corners of the collective intentionality debate left unexplored by philosophers. As for the problem of the nature of collective intentionality, Tomasello leans towards a primitivist account of the capacity for 'we-mode' thinking and acting. As for the mechanisms that underlie collective intentional behavior, he points to a set of pro-social inclinations and inferential skills for sharing mental states. What is most remarkable of these arguments, and the main motivation for choosing Tomasello's comparative approach as the best candidate for a theory of collective intentionality naturalized, is that they are developed and based on a large set of empirical data. This is a paradigmatic example of a natural-scientific approach to issues of social ontology that confronts philosophical problems by invoking continuities with the natural sciences and by treating the phenomenon at stake in a testable manner.

1.3 Summary of the Chapters

I shall proceed as follows. In chapter 2, I shall present the problem of collective intentionality. Philosophers of society postulate collective intentionality to make sense of episodes of everyday interaction where individuals intend to do something together. The key insight is that our common understanding of collective intentional behaviour is not exhausted by concepts of individual intentionality. Indeed, the literature has mostly focused on how to capture the conceptual distinction between the two, while leaving unclear *why* we should approach the problem of collective intentionality from the perspective of the irreducibility question. In this chapter I shall fill this gap by locating the theory of collective intentionality in the context of previous work in the subject, in order to highlight its innovative contribution to classic debates in the philosophy of society. Then, I shall argue that the irreducibility issue presupposes the broader question of the nature of collective intentionality, which asks for a principled distinction between two general meanings of reductive explanation. One is associated with philosophers' preferred method of investigation, linguistic analysis and intuition, whereas the other construes reduction in natural-scientific terms and allows for the possibility of empirical evidence to settle questions about the ontology of collective intentionality. With this distinction in mind, I shall make a distinction between naturalizability arguments in philosophy and theories of naturalization in science, and suggest possible levels of biological explanation of collective intentional behaviour.

From chapter 3, I shall tackle the question whether we have good reasons for thinking that collective intentionality is a natural attribute of reality. This claim has been first put forward by Searle in the context of his account of the construction of social reality (1995). In its most general and controversial formulation, collective intentionality is defined by Searle as a biologically primitive phenomenon of the minds of individuals that cannot be analyzed as the summation of individual intentional states plus mutual knowledge. This definition encompasses the two pillars of Searle's realist approach to the ontology of mind and society: internalism, the thesis that intentionality is an intrinsic property of the biology of the brain; and individualism, the view that society is nothing over and above its individual components. Both views, especially the former, have gained Searle a number of attacks. Contrary to these critiques, however, I shall argue that the

problem with Searle's account does not lie in its assumptions, which instead offer convincing reasons for endorsing realism about collective intentionality. The problem is that Searle treats the biological nature of collective intentionality as a self-evident 'fact', which asks for more cautious elaboration and empirical check.

In chapter 4 I shall consider a social-constructivist explanation of the nature of collective intentionality. Social constructivists like Raimo Tuomela hold that fundamental aspects of human life, including meaning and intentionality, are contingent upon communal social-cultural habits. Although almost all accounts of social ontology subscribe to a very general conception of 'construction', according to which social facts are constituted and maintained through collective acceptance, Tuomela stands out among the major collective intentionality theorists as the proponent of a full-blown constructivist response to the question of the existence and identity of collective intentional states. A core idea of constructionism, it is that research should aim at showing that socially-constructed entities are under human (social, cultural) control, rather than the control of natural factors. Whereas it is very common in the social-constructivist literature to find arguments against the very idea that science should be treated as a successful – if not the ultimate - source of knowledge about the world, my goal in this chapter is to show that Tuomela's account of social ontology is consistent with the tenets of methodological naturalism. I shall proceed by contrasting a famous interpretation of the so-called 'rule-following' problem – more generally: the problem of the understanding of thought and language - with considerations drawn from the philosophy of mind and language of Wilfrid Sellars. Sellars' naturalism offers decisive arguments to counter the radicalism of radical constructivists who view collective intentionality and agency as nothing but social constructions.

In chapters 5 and 6 I shall articulate the second part of my thesis, which is devoted to illustrating Michael Tomasello's program of research in psychology as the most advanced theory of collective intentionality naturalized. The gist of this project, which I shall refer to as the 'Shared Intentionality Hypothesis', is that the complexity and variety of social-cultural phenomena depend on a species-specific cognitive and motivational 'infrastructure' for sharing mental states. In chapter 5 I shall

provide a thorough scrutiny of the Shared Intentionality Hypothesis in light of Tomasello's vast research with infants and their nearest primate relatives, such as chimpanzees. It is worth keeping in mind that my goal is to emphasize the methodology whereby conceptual and empirical issues are jointly tackled and illuminate each other. The point, which I shall emphasize over and over again, is that the evidence of the naturalness of group-thinking, which emerges from research on phenomena such as joint attention in early cognitive development, is no longer the outcome of commonsense and *a priori* intuitions, but rather of experimental practice.

In the cognitive sciences, the Shared Intentionality Hypothesis is a highly regarded response to the question of what sets human cognition apart in the animal kingdom. The intuitive appeal of the theory, as well as the remarkable body of empirical findings in support of it, does not save Tomasello's conclusions from a number of criticisms, however. Some of them arise from the vagueness of certain central concepts which would therefore require more conceptual work on Tomasello's side. But, as I shall argue in chapter 6, most critiques presuppose a wrong-headed interpretation of Tomasello's philosophical position, with unfortunate consequences on the overall assessment of his hypothesis. Contrary to the internalist reading put forward by many commentators, I shall present Tomasello's as an externalist theory of the nature and acquisition of reference.

The underlying motivation is that, by facing issues regarding the emergence and development of shared intentionality through the lens of the voluminous literature on the problem of reference, it will become clear that the problem of communication has significant bearings on the neighboring disciplines including the area of communication studies. Yet, to describe the Shared Intentionality Hypothesis as an externalist theory of reference leaves open the question of what specific construal of externalism Tomasello subscribes to. Since the motivation for proposing this hypothesis is to identify the actual psychological factors that ground reference in the context of interaction, I shall criticize the tendency of most commentators to interpret the Hypothesis as a semantic theory of reference. If there are reasons to believe that Tomasello leans towards externalism in general, this is because he confronts the problem of reference from a pragmatist, instead of semantic, standpoint.

As I shall argue in the last chapter, the assertion that mutual understanding of reference requires individuals to construe the action scene as one of shared intentionality illuminates the core idea of the irreducibility thesis: collective intentionality is prior to individual intentionality because the sharing of mental states is *developmentally* prior and *causally* necessary for reference-fixation. This conclusion is crucial to assess the significance of Tomasello's research in discussions of the irreducibility problem in philosophy, along with the challenges facing his theory. As we shall see in the Conclusions, the next task is to develop the account of the causal influence of group-thinking in setting the conditions of possibility of individual intentional states, so as to make a step forward in the naturalization of collective intentionality.

Two

Naturalizing Collective Intentionality

Philosophers of society postulate collective intentionality in order to make sense of episodes of everyday interaction where individuals intend to do something together. The key insight is that our common understanding of collective intentional behaviour is not exhausted by the concepts of individual intentionality. In this chapter I discuss the question of the naturalness of collective intentionality by examining the gist of conceptual *versus* natural scientific methods of reductive explanation and their consequences on matter of social ontology. I conclude by distinguishing arguments for the alleged naturalizability of collective intentionality from accounts of collective intentionality naturalized which make appeal to distinct levels of biological explanation. Such distinction sets the stage for the two-part discussion of collective intentionality in the chapters to come.

2.1 Introduction

Intentionalism in the philosophy of social science is the view that a theory of human society should be built on the intentional attitudes of the individual agents. As Searle (1995) has famously shown, for example, it is because people collectively intend a piece of paper to be money that money *exists*. The view that facts of the social cannot be ontologically constituted unless people exhibit a certain intentional attitude towards themselves and the other members of their social group, as well as towards the facts they contribute to form, has prominently figured in most 20th social science research (Gilbert, 1989). It is only in the nineteen-eighties, however, that one of its cornerstone assumptions was brought to light and subjected to thorough philosophical scrutiny: collective intentionality.

The starting point is the observation that the 'collective' nature of such phenomena as coordination, cooperation and communication cannot be fully captured by the concepts that we

deploy in understanding individual intentional behavior. For philosophers like Raimo Tuomela, Margaret Gilbert and John Searle, phenomena of sociality underlie a particular mode of thinking, exemplified by notions like 'collective intention' and 'plural subject', which causes individuals to share attitudes of various kind – be they cognitive (beliefs), conative (intentions) or affective (emotions) (Schmid, 2009). Shared, or collective, intentionality is thus a label for the idea that individuals have a propensity to think and act as the members of a collective when they engage in joint action. It is because you and I 'see' each other as being part of the same group that *we* intend to do something together. Yet, how can we prove that this ability is a natural feature of human cognition?

In the twenty years since its initial formulation, research in the nature of collective intentionality has mirrored the more general concern of philosophers to identify the place of the mind in the natural realm. Naturalists like Searle have interpreted the central question of collective intentionality as a demand for the conditions of reduction of collective to individual mental states. Based on the impossibility to individuate such conditions by means of linguistic analysis and intuition, Searle concludes that collective intentionality is a biological primitive form of mental life. What can justify talk of collective intentionality as a natural feature of human psychology? And among those who hold a reductivist stance about the idea that there may be a fact of the matter for the tendency to share attitudes, how is reduction effected? Clearly the argument for the irreducibility of collective intentionality belongs to a family of questions of broader scope which concern the meaning of naturalization and the role of conceptual analysis in philosophy.

In this chapter I shall discuss the irreducibility thesis as the clue to the problem of the naturalness of collective intentionality. The notion of irreducibility lends itself to a twofold interpretation in the present context. There are questions concerning the existence and the identity of collective, as distinct from individual, intentional states; and questions about the conditions for having knowledge of them. How the relation between the former (ontology) and the latter (epistemology) is conceived of, I argue, is of remarkable importance in the study of the nature of collective intentionality. In the work of the founding fathers of the subject, ontological conclusions

are drawn upon analyses of the uses of collectivity concepts in everyday discourse and social science research. On a natural scientific approach to questions of reduction, on the contrary, it is up to science to establish whether collective intentionality can be given a reductive account at some level of biological explanation.

The chapter is structured in five sections. In §2.2 I shall illustrate the rise and development of the research program in collective intentionality on the background of classic debates in the late-twentieth-century philosophy of society. In §2.3 I shall focus on the third and final step of the sequence of elements that figure in the collective intentionality literature, *i.e.* the irreducibility thesis, and discuss the method by which philosophers derive existential conclusions from it. In §2.4 I shall consider an alternative approach which construes the meaning of reduction in natural scientific terms and allows for the possibility of empirical evidence to settle questions about the ontology of collective intentionality. Finally, in §2.5 I shall lay out the conditions for a naturalistic theory of collective intentionality by distinguishing *naturalizability arguments* from *theories of naturalization*, and by singling out the levels of explanation that articulate the biological account of collective intentional behavior.

2.2 The Rise of Collective Intentionality Theory

In philosophy, the central problem of collective intentionality is whether collective intentional states are irreducible to individual intentional states. As we saw in the Introduction, 'collective intentionality' and 'irreducibility' are technical notions, involving concepts that ask for detailed analysis. For clarity, here I shall illustrate the theoretical framework in which the problem of collective intentionality arises as a sequence of three steps, dealing with: the motivation for the theory of collective intentionality; the notion of collective or 'we-intentions'; the irreducibility thesis. In this section I shall examine each step in a diachronic perspective, to illustrate the rise of collective intentionality theory on the background of some classic debates in the philosophy of society.

The first step in the 'standard' characterization is concerned with the *motivation* for the theory of collective intentionality. Since its appearance, collective intentionality has been given a prominent role in accounts of the foundation of human society and, especially, of collective action[3]. In social theory and philosophy, action is a piece of intentional behaviour. The first difficulty that one encounters concerns the meaning of 'intentionality'. This idea is captured by the Latin word '*intentio*' meaning a 'directing towards a target': intentional behaviour is a form of behaviour oriented to the pursuit of goals. Thus, action is always performed with a certain intention. Yet, the emphasis is not on the instrumental nature of intentions, meaning that action depends on one being motivated to act and plan to do so. Whether the intention is formed in advance or materializes with the bodily movements is not the point at stake; what matters is the 'reaching-out'[4] of intentions, the idea that they aim at something, which distinguishes intentional from purely reflexive behaviour. Neither is action identified by behaviour performed with intentions only. Any attitude including intentions, beliefs, desires and more generally thoughts, is intentional in that it directs behaviour to some object or state of affairs in the world.

Collective action can thus be defined roughly as action undertaken by two or more agents who purport to do something together. How persons relate socially to one another and to the social facts they constitute is the problem of social ontology – what there is in the social arena (Pettit, 1993). There are of course a number of interesting theories and questions in the philosophy of collective action. In broad terms, we can say that the problem of collective action is the problem of how individual agents can come to intend and enact things collectively. What does 'collective' mean from the individual point of view? Let us begin with the notion of collective action. First, not every episode of interaction between at least two persons can be classified as collective intentional behaviour. Actions that involve more than one agent, each acting on her own, are merely accidentally, not intentionally, collective. So, we may want to restrict the inquiry to collective actions performed by people intentionally. Yet, as we mentioned, the intentionality of collective

[3] The meaning of 'collective action' is not confined to political action or to a course of action to be chosen in coordination games.

[4] *Intentio* derives from the verb '*intendere*' which literally means to stretch (Crane, 2001: 9).

action is an attribute of generally purposive, rather than merely instrumental, behaviour. Hence, we are faced with a second difficulty about the concept of acting together.

In order to get a clear grasp of the problem, consider the following story. Suppose that you and I go running together every now and then. One day we decide to step up and register for the next London Marathon. We show up on the race day, run the entire Marathon and reach exhausted the final line. At the end, you shout at me something like: "We did it! You and I ran the Marathon!" There is an obvious way to understand this expression in terms of individual intentions which I shall call the *distributive* reading: action is predicated over the individuals, so the 'we' in the expression refers to you and me running the race individually. Indeed, on this reading, it would not have made any difference if we had not planned to embark on the Marathon in advance, and just met by chance at the starting line of the race. We would still have run it in the (distributive) sense that you did it, and I did it. So we want our interpretation of collective action to capture the difference between cases of this kind and genuine case where we not only happened to run together but we did it jointly.

One way to understand what is left out of the distributive reading is to look closely at the intentional attitude that the agents display when they intend and do something with others[5]. If we ran the Marathon with the intention to do it together, 'as a group' so to speak, then of course it is true to say that each of us ran the Marathon; but the opposite is not. The claim that you and I ran the Marathon does not imply that we did somewhat together, and we did it intentionally. This is the sense of the 'we-as-a-group' interpretation that the distributive reading fails to capture. I shall call this the *collective* reading. Let me enliven this point with another example. The Pompidou Centre in Paris, one of the world-famous museums of contemporary art, was designed by architects Renzo Piano ad Richard Rogers. The sentence 'Piano and Rogers designed the Pompidou Centre' obviously expresses the idea that each gave his own contribution to the final creation. But it means that they did it *jointly*, that the project was a truly collective outcome resulting from the two

[5] Bardsley (2007) and Ludwig (2007) offer similar, although not identical, reconstructions of collective intentional behaviour.

architects acting as a group. Thus we interpret the claim collectively, as opposed to the meaning of 'Piano, like Rogers, designed parts of the Pompidou Centre' which suggests that action predicates are distributed over the individuals.

The general approach should be clear enough: when two or more agents come together and act as a group in achieving a collective goal intentionally, the fact that they do something together does not mean that any member of the group does it on her own (Barsdley, 2007). It takes two to tango, so to speak, in the sense that the individual contributions to the joint performance cannot be partitioned over the single dancers if the jointness of the collective action is to be captured. This is the same as saying that, on the collective reading, we understand what we do in terms of you doing your part, and I doing mine, only as part of our doing it together. It follows that joint-action sentences are to receive different analyses depending on whether they are understood in the distributive or the collective sense. The problem, then, is how to conceptualize, and account for, the specific attitude that underpins collective intentional behaviour.

Prima facie two responses are conceivable. The first response characterizes the jointness of collective action as a feature of the bearer holding the relevant attitude. For somebody to hold a 'we-as-a-group' attitude means that there *is* a group to ascribe the attitude in the first place. This argument builds on a theory that has been around for more than a century[6], and has received an influential formulation in the work of Durkheim (1953). The theory accounts for the jointness of collective action by postulating plural agents over and above the individuals engaged in the action. That is, the subject of the joint action has a specific ontological referent distinct from the individual subjects. This is also to say that the 'we' understood collectively points to an existent entity that is not reducible to the sum of first-person singulars. Consider the case of corporate organizations, for instance. The argument goes that, in saying that organizations are legally and morally responsible

[6] The theory inspires at least one facet of the individualism, or micro-macro, debate concerning the nature of social phenomena: *ontological* individualism. Ontological individualists are committed to the view that macro-phenomena are nothing over and above their micro-parts: every social entity is actually an attribute of individual agents. Holists oppose this argument by defending the ontological irreducibility of the social. I don't mean to offer a comprehensive review of the debate, but for an introduction see O'Neill (1973) and Lukes (1973).

for their actions, we don't just refer to the actions of their individual members *as if* they were a group. There are convincing reasons for arguing that we hold a realist, rather than purely figurative or metaphorical, stance towards organizations (Tollefsen, 2002b). We attribute intentional behaviour to them because they are minded *i.e.* intentional agents. Neither should these attributions be considered false, an argument easily dismissed by the evidence of their explanatory and predictive success in various social science research programs like rational choice theory (Tollefsen, 2002a; List and Pettit, 2006).

The view that there are irreducible 'collectives' is vulnerable to a number of critiques, though. Among the others, one problem is to make sense of the idea that collectives have their own attitudes emergent from those of their individual constituents. This problem is related to the limits of emergentism[7]. At any rate, even if a solution was put forward it would not be decisive. In fact, provided that collectives intend and do what their members intend and do, how could we account for the 'we-as-a-group' attitude of collectives unless we know what it means for persons *qua* individual agents to have such an attitude in the first place? To posit existent supra-individual agents shifts the burden of explanation to another level without actually meeting the initial challenge - what it means for individuals to intend and enact things together. Therefore, appeal to the ontology of collectives is not a suitable response to the problem of collective intentional attitudes.

The second response takes into consideration the psychology of collective action. In the early days of collective intentionality theory, philosophers started to confront the problem of collective action by investigating the type of attitude that persons display when they intend and do something with others. Back to the Marathon story, the difference between running the race by one's own and

[7] There is a voluminous literature on the concept of emergence and its use in social theory (starting from Fodor, 1974; for a review see Sawyer, 2001). Briefly, the lesson of emergentism is that, if we want to explain the peculiarity of the properties that arise at the macro level, we must explicate how they emerge from micro-level properties. The question of emergentism is how to reconcile the two senses of 'emergent social phenomena'. On the one hand, it is said that social phenomena emerge from individual intentional attitudes in the sense that there must be an explanation of how the former are grounded in the latter. On the other hand, the collective outcome is independent of those attitudes, meaning the former cannot be epistemically reduced to the latter.

as a group, might be a feature of the way in which each of us *represents*[8] this action in her mind. Instead of having a thought expressed by the words "I intend to run the Marathon with you", you and I may entertain the following representation: "*We* intend to run the Marathon together". Depending on how each understands the 'we' of the relevant intention in his/her mind, we would intend to engage in the Marathon as a collective intentional effort - as a group - or simply as individual runners. The intuition that behaviour can be guided by intentions that are also collective, as distinct from first-person singular intentions, motivated philosophers like Raimo Tuomela and Margaret Gilbert in the nineteen-eighties to found their analyses of social reality on the notion of collective or '*we-intentions*'.

2.2.1 The Early History

The first articulate characterization of the notion of collective intention can be found in "*We-Intentions*", a paper that Tuomela co-authored with Kaarlo Miller in 1988 and which is now widely recognized as the first self-contained piece of collective intentionality theory. The importance of the paper, however, is not motivated by the novelty surrounding the concept of we-intentions. In fact, the notion had appeared about twenty years earlier in scattered remarks by Wilfrid Sellars concerning the nature of norms, which Tuomela elaborated and brought to completion in his own theory of social ontology. Let us, then, begin with the conceptual background of 'We-Intentions'.

In his 1963 "Imperatives, Intentions and the Logic of 'Ought'", Sellars provided a reductionist account of moral reasoning to practical thinking. What one ought to do, expressible in 'ought-statements', is analyzable in terms of the conclusions of practical reasoning, which are expressions of one's intentions to do something. Sellars however noticed that the universality of moral

[8] The standard approach to the ontology of the mind in philosophy and cognitive science takes the mind to be a representational system. In broad terms, for people to have intentions and thoughts – more generally, intentional states - is for them to represent the aspects of the world those states are directed at. What a mental representation is, and why we ought to consider the mind as a representational 'machine', is the subject of a huge interdisciplinary literature. It is not possible to discuss all the positions at stake in the space of this thesis, so the representational theory of the mind will be assumed as a default position. For a recent critique of representationalism, see Garzon (2008).

principles, the fact that they are applicable to different agents while retaining their inner 'force', is not exhausted by intention-based discourse in the first-person singular. For Sellars there is some inherently *normative* relation linking individual ought-statements with collective intending: I ought to do my part if *we* intend to do something. He then proposed to capture the inter-subjective bond of moral norms by means of ought-statements that indicate 'we-mode' rather than egocentric intentions (Sellars, 1963: 205):

> We have argued that moral consciousness is a special form of *we*-consciousness, and, in effect, that one who does not intend in the we-mode, *i.e.*, has no 'sense of belonging to the group', cannot be said to have more than a 'truncated' understanding of thought (Sellars, 1963: 205: emphasis in original).

Hence, Sellars' notion of we-mode thinking and acting is meant to characterize the ability of an individual to intend and act as the member of a group. In his language, "*intending-as-one-of-us*" is the logical precondition of the actual sharing of intentions (Sellars, 1963: 204-5; emphasis in original).

There are two aspects of this theory that have had a lasting influence on the systematization of collective intentionality theory. The first is the emphasis on the modality of thought and action that underlies collective intentional behaviour. Although it had been around for twenty years, this insight of Sellars was elaborated and subjected to thorough philosophical scrutiny in the nineteen-eighties in Tuomela's analysis of collective intentional behaviour. The second aspect worth of attention is the normative character of the relation between individual intentionality and the collective character of phenomena like moral norms, which illuminates the meaning of intending-as-a-group. This idea is the central feature of the approach of Margaret Gilbert, another founding figure of collective intentionality theory. Although she declares to have become aware of Sellars' work via the scholarship of Tuomela (Gilbert, 1989: 493), Gilbert arrives on independent grounds at conceptualizing the normativity inherent in the "semantic phenomenon involving the pronoun 'we'" (1990: 8). Let us consider Tuomela's and Gilbert's views in detail.

In *A Theory of Social Action* (1984), Tuomela develops some of Sellars' intuitions concerning the conceptual structure of human social action[9]. The central thought is that persons are social in that they believe that each other is social too. To elaborate on this idea, Tuomela formulates a series of 'holistic social concepts' to explain how single agents act collectively out of individual intentional attitudes. One of these concepts is introduced as follows: "We claim that the 'sociality' or 'social relatedness' central to people's acting together in a central sense comes from or even consists in their relevant we-attitudes" (1984: 12). We-attitudes constitute a class of attitudes that individuals exhibit when they intend and enact things with others. For this reason, Tuomela considers we-attitudes as "the 'carriers' of collective intentionality" (Tuomela, 2002: 17).

Tuomela identifies two main aspects of collective intentionality. In the general sense of intentionality, people acting together are collectively intentional in that they show "social relatedness" (Tuomela and Miller, 1988: 370). In a narrower sense, joint actions are collectively intentional in that they are performed specifically for some collective purpose. Tuomela's account involves collective intentionality in the first sense but not necessarily in the second sense. In other words, human social action is taken to be meaningful (intentional) although it might be performed with no purpose. An interesting feature of this view is that Tuomela takes collective intentionality to be foundational with respect to the ontology of the social world, although he tends to explain collective intentionality by appeal to other concepts concerning social-cultural practices. For example, he refers to "different kinds of collective intentionality" (2002: 17) whose "common denominator" is shared we-attitudes[10] (*ibid.*).

[9] "Given an adequate notion of we-intention (involving the notion of mutual belief) the notion of an intentional joint action can be formulated. (…) With the help of we-intentions, mutual beliefs, and (intentional) joint actions, one can characterize social norms. Given the notion of social norm, social roles can be analyzed. Next, with the help of roles and we-intentions, one can define a strong, normative notion of a social group. From social groups one can proceed to social organizations, institutions, and finally to the notion of a social community" (Tuomela and Miller, 1988: 369).

[10] I leave this point unexplained until chapter §4 where I will provide a comprehensive reconstruction and evaluation of Tuomela's standpoint. Central to his account is the idea that human actions are social "in the wide sense that they conceptually presuppose the existence of other agents and of various social institutions" (Tuomela and Miller, 1988: 369). Collective intentions arise as plans that result not from the aggregation of mere individual we-intentions but from negotiation and discussion among the members of the group.

Thus, intentional collective behaviour is performed by people we-intending to act together. In the introductory chapter of (1984), Tuomela remarks that we-intentions underlie "a mode reflecting the concept of group ('us') on the level of an individual" (1984: 13). Intending in we-mode, in other terms, implies that there is a 'we' or group to which intentions refer to. We-intentions are of a motivational or action-prompting kind, as opposed to standing intentions where the emphasis is on the referred-to thing in the 'aboutness' sense (Tuomela and Miller, 1988: 378). A further aspect is the subject that holds we-intentions. Tuomela states that joint actions are performed by a plural subject or a many-person agent. In this respect, the focus is put on group-thinking where a number of agents act together with the aim to achieve a common goal. Hence we-intentions are also called *group intentions*.

At first glance, it seems plausible to assume that what distinguishes collective from individual actions is precisely the joint intention shared by the agents. After all, the distinguishing feature of we-intentions is that the individuals believe that they are cooperating: each agent must know that the other participants in the joint action are also committed to do their part. So the question is whether the collective intention corresponds to the belief that the others will do the same. Tuomela introduces a distinction that prevents us from accepting this solution, namely the difference between we-intentions and *joint* intentions. In Tuomela's words, "an agent's we-intention (…) is his 'slice' or part of the agents' joint intention, and conversely a joint intention can, upon analysis, be said to consist of the participants' mutually known we-intentions" (2005: 333). This passage introduces a concept, mutual knowledge, which will be discussed in detail in the remaining part of this section.

For the time being notice that, on Tuomela's reading, the agent's intention to perform part of the action does not imply the specific belief that the joint action will produce a certain outcome. On the one hand, then, we-intentions represent the agent's willingness to do something together and can be called 'aim-intentions'. On the other hand, joint intentions are 'action-intentions' in that they entail the direct performance of the action. We-intentions and joint-intentions differ also in another respect. As Tuomela correctly points out, if by joint intention we mean the general intention of the group towards the collective goal to be achieved, we may be tempted to identify the belief of the

group members with the joint intention in question (Tuomela and Miller, 1988: 330). How could we define the conceptual presuppositions of we-intentions except by pointing to the goal of the group? In other words, the agents we-intend to do something together and this leads to the formation of the joint intention as a plan of action. But the belief that the others will also participate in the action seems to be a presupposition of the plan of action (joint intention). So, where does the concept of sociality – the idea that agents we-intend on the basis of the belief that the others will do the same - originate?

Margaret Gilbert answers this question by proposing a solution which closely resembles Sellars': normativity. In *On Social Facts* (1989), Gilbert puts forward a thorough examination of 'intentionalism', "the view that (...) individual human beings must see themselves *in a particular way* in order to constitute a collectivity" (Gilbert, 1989: 12; emphasis in original). The facts of the social cannot be ontologically constituted unless people assume a certain intentional attitude towards themselves. The theory of Gilbert thus develops the idea that our everyday concepts of sociality – concepts like mutual belief and intention, social group, social convention - are *plural subject* concepts (Gilbert, 1989). By 'plural subject' Gilbert means the subject to which the unity of action and the psychological attributes of the 'we' that unifies the individual attitudes are ascribed. "Is there a collective agent here" – asks Gilbert (2006: 12):

> There is reason to find an affirmative answer attractive. Consider the following. On this account, what does 'We' refer to, in 'We are doing A'? It refers to the jointly committed individuals as such. Thus it implies the *real unity* – in Hobbes's phrase - that a joint commitment creates. To echo Hollis and Sugden, we constitute a *supra-individual unit*. Further, in Rousseau's terms now, the joint commitment that unites us creates *a single moving power*. In a more modern phrase, it provides a single *locus of control* for the movements of each (*ibid*.; emphasis in original).

This analysis links the notion of plural subject to that of joint commitment. Plural subject-hood, in other words, is a normative phenomenon. The parties to a joint activity think of themselves as

members of a group as a consequence of holding promises and obligations towards each other. To get a grip on the problem, Gilbert invites us to consider a paradigmatic example of a social phenomenon construed around two people who engage in a walk together (Gilbert, 1990). To go for a walk together is one of a list of 'shared', or 'joint', or 'collective' activities of a special kind which are performed by individuals intending their action to be expression of a view that may properly be referred to as "of one mind" (1990: 10). Joint commitments are formed when each party expresses his or her willingness to participate in the activity together with another. Once this happens, a 'pool of wills' is established with the effect that obligations and entitlements are now 'out in the open'. Such 'common knowledge', in accordance with David Lewis' (1969) formulation of the notion, empowers the agents with rights and reasons to act in a way that accomplishes the plural subject's, or collective, intending.

So, the plural subject emerges from the binding together of individual wills. The bond is not just the unilateral expression of one's promise to meet another's, but it is a form of 'conditional commitment' that requires everybody to be equally committed. "Once this willingness to form the plural subject of the goal in question has been expressed on both sides, in conditions of common knowledge, the foundation has been laid for each person to pursue the goal *in his or her capacity as the constituent of a plural subject* of that goal" (Gilbert, 1990: 7; emphasis in original). Hence, the pool of wills plays a foundational role in the grounding of collective intentionality. But what comes first in the explanation - the obligations and entitlements or the very concept of a plural subject?

This issue is analogous to the question arising from Tuomela's concept of we-intentions. Tuomela's account, as well as Gilbert's notion of plural subject-hood, serves the function to explain what it means for individuals to think and act in a collective way. But for reasons that I shall elucidate in the last part of this section it seems that both accounts fail to render the jointness of collective intentional behaviour in a non-circular way. This consideration introduces the third feature of the framework of collective intentionality: the irreducibility thesis.

2.2.3 The Irreducibility Thesis

The theory of collective intentionality has grown around the question: Are collective intentional states *irreducible* to individual intentional states? One of the central concepts in contemporary philosophy and science, reduction indicates the process by which entities of any kind (theories, propositions, facts, individuals, properties, behavioral patterns, etc.) are redefined in terms of other entities. Procedurally, reductive explanations are construed by laying down a set of necessary and sufficient conditions for the target entity to be reduced to 'base' entities which do not themselves comprise the target. Before distinguishing among types of reductive explanation, I shall focus on the meaning of reduction in the standard theoretical framework of collective intentionality.

Let us consider a simplified version of Tuomela's account of we-intentions (Tuomela and Miller, 1988; Tuomela, 2005):

> Given a joint action X and a certain number of agents forming a collective G, each member *we-intends* to perform the action if and only if the following conditions are satisfied: *(a)* the agent intends to do his part of X; *(b)* the agent has a belief to the effect that the joint action opportunities for X will obtain and that a sufficient number of members of G will do their parts of X; *(c)* the agent believes that there is a mutual belief among the members of G to the effect that the joint action opportunities for X will obtain; *(d)* condition *(a)* holds in part because of *(b)* and *(c)*.

For Tuomela, each agent's representation of the joint action's purpose presupposes that the others will do their parts to achieve it; furthermore, not only does the agent have a belief representing the outcome as jointly achieved, but she also knows that all agents know that this is the case. In "*Collective Intentions and Actions*" (1990), the article in which the term 'collective intentionality' was coined, John Searle charges Tuomela's account of circularity (1990)[11]. More in detail, Tuomela construes we-intentions to explain what it is for an agent to intend and act as a group, namely to represent an action's target as something that can only be achieved by the group as

[11] Here I am only concerned with Searle's critique of Tuomela; references to his theory of collective intentionality will thus be limited to the essential. The reader must wait until chapter 3 for a comprehensive evaluation of Searle's stance on the irreducibility thesis.

a whole. Yet, on his account, we-intentions presuppose the belief that there already are other agents with the same kind of intention. Clearly, if the very notion of we-intention entails that one must believe that there are others who intend and act as a group, the analysis is circular because it resorts to the very concept in need of explication.

Therefore, in Tuomela's theory, there is no effective reduction of collective intention to more elementary concepts that capture the social dimension of thought and action. Collective intention is irreducible, or *primitive*. In fairness, Tuomela's own thinking about the structure of we-intention has evolved throughout the years in ways that substantially depart from the initial characterization. In his most recent contribution to the subject, he clearly acknowledges that any account of the jointness of collective intentional behaviour is faced with the question of how to construe the relevant 'we' in such a way as to avoid the charge of circularity. "We-mode mental states and actions typically are joint states and actions in a strong sense involving an irreducible, thick 'we' (that is, a 'we-together'), and this makes the ontic 'jointness' level central for the construction of the social world" (2007: 10). Hence, either one takes the concept of thinking and acting in we-mode as not decomposable into more basic components *i.e.* primitive, or one is to accept that the analysis is likely to be circular – though perhaps not viciously so.

Similar considerations apply to the notion of plural subject-hood, the key concept of Gilbert's analysis. For a plural subject to come about, "its members must correctly understand their situation *in a certain way* and their behaviour must be explicable in terms of *this understanding*" (Gilbert, 2006: 12-3; emphasis mine). But isn't this form of understanding – the individuals' special way of thinking of themselves as members of a plural subject - the concept that the very notion of a plural subject is introduced to clarify? By arguing that such form of understanding consists in "a grasp of a subtle conceptual scheme, the conceptual scheme of plural subjects" (Gilbert, 1989: 416), Gilbert seems to treat the concept in need of explanation as a primitive notion, too, thus leaving it unspecified (Tollefsen, 2002). If people did not have unmediated and direct understanding of the scheme of a plural subject, it would be impossible to explain where the joint commitments originate without ending up running in a circle. Moreover, it is not satisfactory to just reply that the notion of

a plural subject is a 'technical' term, and that people need not be able to master it when they think in we-modality. It is precisely because the notion is very technical that it is problematic to introduce it on grounds that do not require explicit access on the side of the people engaging in a joint activity (Tollefsen, 2004: 12).

These criticisms show that Gilbert and Tuomela encounter the same conceptual difficulty in characterizing collective intentional behaviour: they aim to analyze it by resorting to a prior understanding of the concept in need of explanation. So, neither does a normativity-based view of shared intention nor one based on mutual beliefs and common knowledge succeed in spelling out the base of collective intentionality in a non-circular manner. For this very reason, Searle (1990) concludes that collective intentionality must be a primitive feature of the mind, *i.e.* one that cannot be analyzed as the summation of individual intentional states and their interrelations (resulting in a state of mutual knowledge). Aside from the specifics of Searle's theory, philosophers usually look with suspicion at theories that are explicitly built upon irreducible or primitive new notions. If it is accepted that collective intentionality figures at the foundation of social reality, it is an essential part of any theory of society that the concept of collective intentionality be further explained. If, on the contrary, the irreducible notion constitutes the very *explanandum* of the theory, there are good reasons for discarding the theoretical framework as unsatisfactory.

This conclusion, however, trades on the ambiguity of 'irreducibility'. As we pointed out, to say that collective intentionality is primitive is the same as saying that collective intentional states cannot be understood (described, explained, analyzed) in terms of the concepts which we already deploy in understanding individual mental states. Notice that the point of this definition is not that collective intentionality cannot be explained at *any* level: an intentional predicate is primitive in a domain when it cannot be conceptually reduced to simpler constituents of the *same definitional domain*. That is, the concept of collective intentionality is said to be primitive not in 'absolute' terms, so to speak, but with respect to the specific framework in which it is theorized. In such framework, collective intentionality appears as primitive; yet, on a distinct conceptual background,

say one that does not involve intentional predicates whether in I-mode or we-mode, it might well be the case that collective intentionality can be given a reductive explanation.

An obvious corollary is that, whenever we evaluate the 'classic' project of collective intentionality analysis, we should keep in mind that it draws on the framework of *folk*, or *intentional*, psychology. This framework assumes that, in Tuomela's words, "persons are thinking, experiencing, feeling, and acting beings capable of communication, cooperation and following rules and norms" (2007: 6). It is within the conceptual scheme of intentional agency that the question of the irreducibility of collective to individual intentional states has been formulated and debated. Thus, in order to evaluate the irreducibility thesis and, more generally, the nature of collective intentional behaviour we need to look at how the founders of collective intentionality theory construe the relevant framework of analysis. Their privileged methodology exploits experience-based expertise and commonsense intuitions to disentangle the conditions of reduction of collective to individual intentional predicates. This methodology employs one of a spectrum of possible reductive explanations, namely conceptual analysis.

2.3 *A Priori* Knowledge and Conceptual Analysis

Reductive explanations fall by and large in two categories: conceptual and scientific. The crucial difference between conceptual and scientific reductions is a matter of how relevant conceptual analysis is in deriving existential commitments about the nature of the things concepts are about. Conceptual reductions purport to state the conditions for something to satisfy the meaning of a concept using terms that are different from those designating the target concept. Scientific reductions in contrast move from the assumption that the conditions of existence of an entity cannot be adjudicated purely on conceptual grounds. In this section we will deal with conceptual reductions for which philosophical reflection is an integral and indispensable component of the inquiry into the nature of social reality. An example of this approach is Bratman's reductionist theory of shared intention which we will be discussed in the second part of the section.

For most of the 20th century, philosophy has been primarily concerned with the analysis of concepts, although the significance of conceptual analysis has faced important challenges in recent times. Until modern science and the scientific method have established themselves as the primary source for achieving putative true knowledge of the world, the prevailing view was that reductive definitions convey the nature, or essence, of everyday concepts[12]. Reductive definitions were thus designed to disentangle the meaning of common but somewhat obscure predicates by setting out *a priori*, exceptionless and intuitively acceptable conditions for their application. But this claim was soon overtaken by the now received view that philosophical research ought to be undertaken in close relation with – if not as part of - science, as the discussion of methodological naturalism in the Introduction has clarified. The consequence is a weaker endorsement of *a priori* analysis in philosophizing, one that recognizes its role as a reliable source of substantive knowledge only insofar as the analysis is integrated into the construction and assessment of empirical (synthetic) theories of the world (Papineau, 2007).

However, according to a leading contemporary school of thought inspired by the work of David Lewis and Franck Jackson, known as the 'Canberra Plan', conceptual analysis is still a necessary requirement for drawing substantial *existential* commitments concerning the nature of entities like intentional predicates. According to Jackson (1998), whenever we try to achieve informative knowledge about the nature of intentional predicates, we are faced with the tension between the folk and the scientific properties associated to these predicates. Hence, any whole-hearted naturalist needs to address the problem of "when and whether a story told in one vocabulary is *made true* by one told in some allegedly more fundamental vocabulary"[13] (Jackson, 1998: 28; emphasis mine). According to the Canberra school this preliminary effort of conceptual clarification is essential to reach conclusions concerning the ontology of the non-conceptual world. In other words, it is an indispensable requirement of any attempt to naturalize intentional concepts by making appeal to

[12] The difference between accounts that give conceptually, as distinct from scientific, necessary and sufficient conditions for something to be what it is, is illustrated by the example of colour (see Crane 2001 for an extensive formulation of this point).

[13] According to Jackson, it is not required to give this 'bridging' explanation by setting out necessary and sufficient conditions in *physical* terms (*ibid*.: 62).

synthetic theories that one first identifies their role in folk psychology. Conceptual analysis turns out to be constitutive of the metaphysical agenda then, in that it identifies the properties to be reductively naturalized.

Unfortunately, most attempts to fix the meaning of various concepts in philosophy and psychology by means of reductive definitions have produced scarce results (Stich and Lawrence, 1994). Objections to the Canberra Plan can be divided in two groups. One line of criticism concerns the status of primitive concepts in folk psychology. We have examined earlier what it means to conceive of a term as primitive relative to the theoretical framework to which it belongs. Now the question is: If we cannot explain collective intentionality using other intentional terms, how might we achieve a fuller understanding of the nature of the phenomenon within the framework of folk-psychology? Recall that the point of the Canberra Plan is not to analyze the intentional predicate against a *different* conceptual framework, but to provide a causal-functional account of it. Yet, from the presupposition that irreducible terms acquire meaning against a given network of related concepts, it does not follow that there are no laws in the basic or the special sciences that invoke these very terms (Stich and Lawrence, 1994).

The second line of objection is that it may be the case that not all common sense concepts can be reduced by way of the same conceptual procedure. If multiple descriptions of the very same concept are on offer, which one is best suited for fixing the nature of the entity at stake? Such criticism has led many to question the real value of conceptual analyses in the economy of naturalization projects. Conceptual descriptions might be "useful as a general guide to identifying something, but they do not settle what it is for a thing to be a thing of that kind" (Grayling, 1997: 199). This argument finds support in the case of those predicates that are defined in terms of natural kind[14] classifications, as it is typically the case in scientific reductions.

Before we turn attention to this second class of reductive explanation, I shall discuss Bratman's theory of shared intention as an example of the conceptual reduction of collective to individual intentional states. One might correctly object that Bratman is concerned with analyzing the structure

[14] For a comprehensive survey of the concept of natural kind see Bird and Tobin (2008).

of collective intentionality rather than naturalizing it, in the sense that he does not believe in the existence of a fact of nature for the collective mode of reasoning. Nevertheless, the point of discussing Bratman's theory in this context is that it is a chief example of the kind of methodology that mixes ontological with conceptual matters. I shall focus on the analysis of the method by which, from an allegedly successful reduction of collective to individual intentional terms, he concludes that groupthink does not *exist*.

2.3.1 Sharing Intentions

Bratman stands out among collective intentionality philosophers as a fierce critic of the view that postulates some irreducible collective attitude as the prerequisite for sharing intentions. His individualistic account 'in spirit' aims at providing a non-circular account of shared intention that avoids recourse to a primitive capacity (Bratman, 1993). For this reason, it might be objected that his contribution to the debate over the naturalization of collective intentionality is marginal. This is true insofar as the contribution is assessed from within the debate on the irreducibility of group-thinking. Yet, not only does Bratman's theory proceed from a direct attack on the irreducibility thesis, by arguing that collective intentional states are nothing over and above the set of interrelations between the individual intentional states of the participants in a joint activity. More significantly, the motivation for discussing his theory in this context is to offer an example that critically addresses the irreducibility question by way of a conceptual reduction. In fact, Bratman's existential conclusions about the naturalness of collective intentionality are entirely confined to a discussion of how best to construe the interrelations of individual intentional states.

To begin with, notice that Bratman's concept of shared intention and the notion of we-intention, which philosophers like Tuomela and Searle inherit from Sellars, is significantly different. Bratman rejects the idea that shared intentions are attitudes of a certain kind that depend on the existence of a 'we' in the head of individuals. Ordinary attitudes, like beliefs and intentions, may involve either the activity of a singular or a plural subject and, yet, this does not entail a shift in the nature of intending. What is distinctive of shared intention is its content, which differs from that of the

individual intentions that constitute it depending on how these intentions are contextualized and interrelated. "Both Tuomela and Searle want to allow that there can be a we-intention/collective intention even if there is in fact only one individual (...). In contrast, it takes at least two not only to tango but even for there to be a shared intention to tango" (Bratman, 1993: 103). But how can two people intend to tango if each person does not realize that this is what the other intends to do *together*? In other words, how does a purely individualistic account of shared intention that does not appeal to an irreducible capacity face the threat of circularity?

Let us consider a simplified version of Bratman's account (Tollefsen, 2004):

> Given the joint action J, we intend to J if and only if: *(1)* (a) I intend that we J; (b) you intend that we J; *(2)* I intend that we J in accordance with and because of *1*(a) and *1*(b), and meshing sub-plans of *1*(a) and *1*(b); you intend the same; *(3) (1)* and *(2)* are common knowledge between us.

The conditions *(1)* and *(2)* evidently show that the agents involved in the joint action must be responsive to each other. The concept of mutual responsiveness is a central component of the characteristic functioning of a shared intention (Bratman, 1992: 328). According to Bratman, in shared activities each of us is responsive to the intentions and actions of the other as well as to the collective end. Yet, Bratman invites us to take a 'neutral' stance in evaluating the 'we' that figures in the condition *(1)*, namely in the content of the intentions of each participant (Bratman, 2008). This component can be taken as referring to the joint activity that these intentions give rise to only as part of the "web of attitudes" that unifies and coordinates individual states (Bratman, 1993: 108).

The threat of circularity is therefore avoided by analyzing mutual responsiveness in terms of 'meshing sub-plans' and 'interlocking intentions'. For my and your intentions to be shared, each must intend that every participant performs the joint activity in accordance with sub-plans that mesh, in the sense of being co-realizable. And this requires that the relevant intentions be interlocked so as to create some 'semantic interconnection' (Bratman, 2008). I intend that we J in part by way of your intention that we J. The state of intending that we J is, in sum, a state of shared

intentionality which results from each person having an intention that is interrelated with another's in the right way, rather than from a *sui generis* kind of intending.

Moreover, as Bratman points out, all this would not be possible if the participants did not have common knowledge[15] as well. When the participants in the joint activity plan to act together, they in some sense know of the fact of the shared intention, including aspects of treating the others as co-participants and of interweaving sub-plans as required. Bratman, though, does not give a precise formulation of the kind of epistemic access underlain by the concept of common knowledge (Bratman, 2008). As we have already remarked, Bratman's analysis is silent with regard to the charge of circularity that emerges from specifying the content of mutual belief. Yet, what reasons can be given for remaining neutral about the problem of circularity?

At first glance, it is reasonable to make appeal to certain aspects of sociality – concepts of interlock, mesh, interdependence, etc. - to decompose the 'we' that appears in the content of the individuals' states. But how is it possible to evaluate the collective attitude as reducing entirely to individual attitudes linked in the appropriate way, if the very 'we-concept' figures in the content of each participant's intentional state? To disentangle the content of shared intention in terms of features of individual states seems to shift the problem on a further level of conceptualization. Moreover, it has recently been suggested that Bratman's analysis is also limited in one important respect, which will become clear in the second part of the thesis where attention will be drawn to the naturalistic theory of shared intentionality in developmental social cognition.

Pacherie and Dokic (2006) maintain that Bratman's model is too cognitively sophisticated in that it describes the mechanisms of sharing intentions as involving the kind of conceptual resources and conscious planning that are fully observed only in adults. The problem with this characterization is that it cannot explain why infants as young as one-year olds prove able to engage in meaningful episodes of shared intentionality with their caretakers (Tomasello, 2008). There are robust results in cognitive psychology showing that these episodes occur far before any 'theory of mind', the term

[15] I use the term as a technical notion without digging deeper into it. Greater attention to the relation between common knowledge and collective intentionality will however be paid later.

used in developmental social cognition to denote the child's socio-cognitive abilities for interaction, is established. Therefore, it is problematic to ground the account of shared intention on the capacity for rational deliberation that Bratman considers a pre-condition for sharing intentions. Furthermore, the objection of cognitive sophistication can also be read as an implicit attack on the idea that the underpinnings of collective intentionality must all be cognitive. Bratman is aware that two persons engaging in a joint action raise distinctive obligations towards each other. None of them can opt out, in other words, without the other's permission. But he also contends that the normative aspect is not a foundational ingredient to the sharing of intentions (Bratman, 2008). Along with his peculiar reading of the notion of shared intention, this is yet another way for Bratman to depart from Sellars' intuition of the inherently normative nature of collective intending.

To sum up, the lesson of Bratman's reductivist account is that there is no primitive collective thinking and acting under an appropriate construal of individual mental states and their interrelations (including mutual beliefs). Although this conclusion is interesting on its own as a reasonable critique of the irreducibility thesis, the aspect of interest for our discussion is the way by which it is achieved. Bratman argues that there is no capacity for collective intentionality that cannot be reduced to its elementary units and their connections, where the reduction is entirely supported by conceptual considerations alone.

2.4 Scientific Reduction and Conceptual Irreducibility

The second family of reductive explanations is built upon the notion of scientific, instead of conceptual, definitions. The aim is still to reduce the pre-theoretical sense of intentional terms to a set of necessary and sufficient conditions, except for the fact that reductions are now cashed out in scientific terms. To vindicate the reality of intentional predicates is thus no longer considered the result of an intuition-driven, *a priori* analysis of the concepts associated with them. For semantic properties to be reduced to their basic underpinnings, the latter must be natural kind terms which can only be detected by doing the appropriate sort of science. Scientific reductions do not form a

monolithic group but come in different forms: bridge-law reductions[16], identity reductions, and functional reductions (Kim, 2006: 276). Identity and functional reductions will be examined together in chapter 3 when the main presupposition of Searle's naturalism - that mental phenomena are individuated by their causal roles and can, subsequently, be reduced to neurobiological states - will be elucidated.

The idea that the meaning of natural kind terms can be identified by means of conceptual analysis has come under the attack of Saul Kripke and Hillary Putnam in the 1970s (Putnam, 1975; Kripke, 1980). Their criticisms are classified in arguments from ignorance and error, and modal arguments. In spite of the fact that both forms of argument exploit linguistic intuitions to show that intentionalist concepts do not give information about the 'naturalness' of the entities designated, thus making indirect appeal to the method that those arguments are meant to reject, the Kripke-Putnam argument has had lasting consequences on the debate of the naturalization of the mind. More specifically, modal arguments are thought-experiments in which we are asked to question properties that we would never imagine real-world entities could lack[17]. The point of these thought experiments is that we can use an expression to refer across counterfactual scenarios without knowledge of the features that constitute the extension of the entity referred to.

Analogous considerations characterize arguments from ignorance and error. Here the point is that, although people rely on the information that they possess when they think and talk about something in the world, empirical research can always prove that this information is actually true of something else, or perhaps nothing. Many examples from the history of science show that we are prepared to learn new facts about the way we think about things. These are compelling reasons for thinking that descriptions of the concepts of natural kinds is inessential to 'settle', in the sense of coming to know, the truth-conditions of their extension. So, without empirical check, it is likely that

[16] Since the founding work of Ernest Nagel in nineteen-fifties (see Nagel, 1961), the bridge-law model of reduction has remained the standard reference for the reduction of scientific *theories*. The standard form of such reductions is inter-theoretic, such as in the classic model of gas temperature-pressure laws reduced to statistical molecular physics.

[17] The *locus classicus* is the Twin-Earth parable (Putnam, 1975), which I shall elucidate in chapter 6.

descriptions 'in the mind' of people will pick out things that do not belong to the extension, or will exclude things that do belong (Margolis and Laurence, 1999: 22).

The relevance of the Kripke-Putnam picture[18] is to turn light on the 'classical' conflation of metaphysics with epistemology (Rey, 1983). A prominent assumption of philosophical discourse, the distinction between metaphysics and epistemology separates issues concerning *what* there is in the world, and *how* we describe, classify, infer or know about it. Kripke and Putnam argue that what makes a tiger an entity of that kind, or George Washington the entity designated by the name 'George Washington', is not a matter of what we know about them. The general point is to deny that what we ('internally') know about the entities identified by proper names and natural kind terms determines what entities those are. Whatever conditions support our use of an expression does not constitute necessary and sufficient conditions for identifying the extension of it. This is entirely a metaphysical problem, a fact about the world rather than a fact about our beliefs about it (Rey, 1983: 291).

The moral is that it must take something other than the analysis of internal concepts to *justify* a realist attitude towards the nature of some entity. Or at least this can only be done as a result of an empirical investigation that falls under the province of scientific inquiry. It is through science alone that we achieve reliable knowledge of the predicates of intentional language. In other words, "if our commonsense views (…) may be seriously mistaken, then the (alleged) fact that common sense imbues intentional states with scientifically unacceptable features entails nothing at all about the scientific respectability of intentional states" (Stich and Lawrence, 1994: 178). The upshot of this argument for our discussion is that, whether there is any natural kind that vindicates the place of groupthink in the natural realm, is up to natural science to discover. The clue to the naturalization of

[18] This line of argumentation has paved the ground for the emergence of 'externalism' in the philosophy of mind, the view that the existence and identity of the meaning or content of thoughts is entirely a matter of the existence and identity of the real-world entities thought about. This intuition has hugely swayed projects of naturalizing intentionality, on the presupposition that the intentional content of thoughts is whatever they are related to in the world *in the appropriate way*. Programs thus differ to the extent in which they give specific characterizations of the causal relation that connects mental contents to the entities in the world which they are about.

collective intentionality is to identify the correspondent of the irreducibly collective dimension of intentional states in some lower-level processes describable in the vocabulary of, say, neurobiology. If there are detectable correspondences between psychological occurrences and neuro-physiological events, then realism about collective intentionality is justified. So, to say that collective intentionality is irreducible within the schema of intentional agency is not to say that there is *no* scientific explanation at all of it.

What 'language' can be used to describe the collective intentionality of mind and action outside of psychology? Samuels (2002) argues that, in scientific as opposed to folk psychology, a trait is psychologically primitive if there is no explanation of the process through which the trait is acquired. A theory of acquisition results from scientific theorizing and aims at explaining how an organism has come to possess a given trait, as opposed to commonsense explanations that predict and explain the trait in intentional terms. But the conclusion that the concept for a given trait is primitive if there is no scientific theory of acquisition does not exhaust the scientific understanding of it. Samuels convincingly makes the point that a cognitive feature that stands undefined relative to scientific psychology might be conceived as no longer primitive at another level of theorizing. Though collective intentionality is treated as a psychologically primitive feature, it can nonetheless be given a specification in terms of proximate or ultimate causes; this can be done in neurobiology or molecular biology for instance, or via a psycho-developmental explanation.

2.4.1 Fitness

Before we distinguish among levels of biological explanation, it is worth reminding that the scientific inquiries into the nature of entities that turn out to be primitive on some level of conceptualization are a common issue in the philosophy of science. Philosophers of science have long debated the role and 'empirical' meaning of theoretical entities in such textbook examples as Newtonian mechanics or Darwin's theory of natural selection. The notion of 'fitness' is a good example in this respect, as it helps single out the kind of issues that we encounter in exploring the naturalness of collective intentionality in science.

'Fitness' is a key explanatory concept of the theory of natural selection. It is used to express the Darwinian thesis that evolution is driven by the differential capacity of biological organisms to adapt to their environment. However, the notion of fitness has also raised a host of questions related to the explanatory power and testability of Darwin's theory. Alexander Rosenberg (1983; 1988), in particular, has cast himself in the last twenty five years as the main proponent of the view that in order for the key concepts of natural selection to have scientific legitimacy, fitness should be given a non-circular interpretation. If this is not possible, the only way not to trivialize the theory is to treat fitness as a theoretical entity, namely "a primitive or undefined term *with respect to* the theory of natural selection" (Rosenberg: 1983: 463-4; emphasis in original).

Two interpretations of fitness have polarized the debate thus far. The classical interpretation identifies the relation between two individual organisms, one of which is fitter than the other, in terms of rates of reproduction. This is an operational definition that gives a measure of an organism's fitness based on its number of offspring. On the view that considers ensembles instead of individual organisms and analyzes evolution through the lenses of population genetics, "the theory of natural selection is then treated as a set of claims about how populations' and sub-populations' sizes change over time as a function of differing reproductive rates at some initial time, holding environments constant" (Rosenberg, 2008: 3). Some of the proponents of populational interpretations are also advocates of the second interpretation of fitness as a probabilistic disposition. The fitness of an organism under this definition does not depend on the actual number but on the propensity to have a certain number of offspring. So in order to define what a propensity to have a certain number of offspring consists in, one must focus on the disposition's causes and effects. These remain nonetheless conceptually distinct from the actual behavior –an organism can have the propensity, but never actually riproduce - and this would save the definition from the charge of circularity resulting from direct reference to rates of reproduction.

However, Rosenberg's view is that both interpretations suffer from significant flaws which only have the effect to trivialize the theory of natural selection. The only way to retain the explanatory potential of the theory of natural selection is, on his account, to give a definition of its causal

variable – fitness - on independent grounds, namely without making appeal to some of its determinants (*i.e.* differential reproduction). Hence, Rosenberg proposes to treat 'fitness' as a primitive notion. This is not equivalent to saying that fitness is simple on *any* account. If there is no conceptual room for an independent definition on the most appropriate axiomatization of the theory of natural selection available at the time[19], then 'fitness' should be treated as a theoretical primitive (Rosenberg, 1983: 464). But this of course does not rule out the possibility that there may be another axiomatization of the theory, waiting to be elaborated by scientists, which succeeds in giving a reductive explanation that avoids the charge of circularity. This is a point of great importance to understand why a primitive feature in the framework of intentional agency such as collective intentionality can be given a natural-scientific, reductive explanation at the biological level.

2.5 Prospects of Naturalization

To be realist about collective intentional states is to assume that they are real, *i.e.* true of the reality that we inhabit. What is it for something to be real? Since for a naturalist only scientific theory and practice provide a reliable answer to what there is in nature, arguments for the reality, or naturalness, of collective intentionality are the object of naturalization programs. A naturalistic program is one that explores the continuity between its domain and that of one of several neighboring natural sciences (Sperber, 1996). In the following I shall construe the notion of naturalization as entailing two distinct meanings, philosophical and natural-scientific, and discuss their role in the economy of the thesis.

First, the naturalization of collective intentionality is the project that aims at giving an argument for the *naturalizability* of collective intentionality. For collective intentionality to be naturalizable is for us to believe that the difference between the intentionality of the first-person plural and that of the first-person singular is a natural attribute of the world. If it belongs to the basic 'fabric' of reality, collective intentionality can then be given a reductive explanation which falls in the

[19] Rosenberg refers to the axiomatization proposed by Mary Williams (1970).

scientific domain. As we will see in the following two chapters, there are two realist arguments in the collective intentionality literature – John Searle's and Raimo Tuomela's - which advocate distinct conceptions of what makes us think of collective intentionality as part of the natural realm. These arguments give reasons for naturalizing collective intentionality; yet, for it to be *naturalized* is a matter of scientific investigation.

The second meaning of naturalization is the view that there is a remarkable difference between the argument that collective intentionality is naturalizable and a theory of collective intentionality *naturalized*. The question whether there are good reasons to endorse realism about collective intentionality is a metaphysical question. Hence, it is separate from the question whether there actually is any viable scientific theory which succeeds to meet the criteria for the naturalization of collective intentionality. Standards of success, after all, will primarily depend on specifying what it is to naturalize the subject matter. And this question does not contemplate a unique set of naturalistic conditions, depending on the target of naturalization. How are these conditions to be set out? In this final section I will discuss several candidates for a scientific explanation of collective intentional behavior, and set the stage for the analysis of one in particular that will take place in the second part of the thesis.

In general, research programs on collective intentional behavior tend to exhibit high eclecticism in their methodology, due to the inter-disciplinary nature of the subject and the availability of tools across various fields. Outside of the humanities and social sciences, a paradigmatic natural-scientific approach to collective intentionality is to discover some natural mechanism that explains aspects of the phenomenon in a testable manner. To help distinguish among possible naturalistic programs, let us start from one influential approach to the scientific study of behavior, which builds on the distinction between ultimate *versus* proximate causes (Mayr, 1961). Ultimate causes can be succinctly described as those concerned with 'why-questions', that is, why a trait came to be in an organism; in contrast, the pursuit of proximate causes purports to answer 'how-questions', concerning the way the trait operates in the organism. Building upon this analysis, the biological study of behavior is nowadays interpreted as asking four questions, which are known as

Tinbergen's 'four questions of ethology' after Nikolaas Tinbergen elaborated them in his programmatic paper "On the Aims and Methods of Ethology" (1963)[20].

The four explanatory areas that structure the study of behavior patterns are: causation, function, ontogeny and evolution (Sterelny and Griffiths, 1999; Griffiths, 2008). Questions of *causation* aim at a proximal explanation of the mechanism in charge of triggering and controlling this behavior; proximal causes can be detected at various levels of complexity including the cognitive, the physiological, or the chemical level. Questions of *survival value* ask for an adaptive explanation of the role, or function, that the behavior currently plays on the chances of survival and reproduction of the organism. *Ontogenetic* questions fall generally into the scope of developmental psychobiology, the study of how the pattern of behavior revealed by causal analyses emerges in the organism and changes with age. Finally, *evolutionary* or phylogenetic questions confront the ultimate issue of how and why this pattern evolved the way it did, and are routinely answered by comparing similar patterns of behavior in related species. Proximate (causal and ontogenetic) and evolutionary (functional and phylogenetic) analyses drive most programs of naturalization of the mind, and appear to characterize also the state of art of naturalistic programs of collective intentionality.

The proximate causation of collective intentional behavior is perhaps the most debated and publicized facet of the ongoing project of naturalizing the facts of the mental. When social theorists and philosophers debate the naturalness of collective intentionality, most often they refer to the question whether there is any successful reduction of first-person plural intentional predicates to states of the brain, the character of which is undoubtedly 'scientific'. The target of scientific reductions, as we said, is to reduce the mental to the neurological by showing that any description of a phenomenon in folk-psychological terms could be translated into the language of neurobiology. At present there are various research programs in cognitive neuroscience which draw, more or less explicitly, on the conceptual resources of collective intentionality theory to identify the neural bases

[20] These questions are the key to Tinbergen's vision of ethology, which he contributed to found on solid objectivistic and naturalistic grounds along with his long-term collaborator and friend Konrad Lorenz (Burkhardt, 2007).

of social behavior[21]. Nevertheless, many philosophers inside and outside of this sub-field have expressed relevant doubts on the soundness of reductionism as a viable naturalistic route.

A perfect 'translation' of the psychological categories of collective intentionality theory into neurological categories encounters the problem of the multiple instantiation of (collective) intentional states. The concept of 'multiple realizability' was introduced by Jerry Fodor in the 1970s as a critique of *type*-identity theory, the view that each type of mental state is identical with some type of neural state[22]. Fodor showed that confidence in the project of reducing intentional predicates of a certain type into purely natural terminology – which would involve stating a set of necessary and sufficient conditions for the application of those predicates - is misplaced because most higher-level mental properties can be multiply instantiated in lower-level physical states (Fodor, 1974). This problem applies to the 'collectivity' of mental states, too. Suppose that Carrie is playing with her one-year old son Paul in the house garden when she points to dad parking the car as an invitation to welcome him back home. We can view this as a case of collective action where the goal is common, with Carrie and Paul sharing collective intentions (the intention that 'we welcome dad home'). Since it is plausible to assume with Fodor that their intentions would be instantiated in different neuro-physiological states, what is it about these states that make Carrie and Paul have intentions of the same type, *i.e.* we-intentions?

Problems like the multiple realizability of collective intentional states enlightens the current trend among philosophers to assume a rather 'liberal' stance towards the issue of the neurological causes of collective intentionality. Those who defend a realist stance about collective intentionality tend to assume that every token of mental state held in we-modality is a neurological, hence natural, phenomenon *in principle*. So, naturalness is granted on more liberal grounds than reductionism (in

[21] Among the others, see Walter *et al.* (2004), Adenzato *et al.* (2005), Pacherie and Dokic (2006), Rilling (2008a; 2008b).

[22] Multiple realisability of the mental is usually given as the reason that urged philosophers to direct attention towards forms of token-identity theory (the view that each particular instance, or 'token', of mental life is identical with a brain state) and functionalism (according to which mental states are distinguished by their functions, or causal roles, in relation to behaviour and other mental states) as the current orthodoxy in the philosophy of mind (for an introduction see Botterill and Carruthers, 1999).

its physicalist fashion[23]) would allow: it is accepted that there is a correlation between collective intentional states and brain states which would be discovered by neuroscience. Meantime, however, it need not be necessary to provide a successful type-type inter-theoretic reduction for those states to be shown to be real.

Evolution, broadly conceived, singles out the other wide project of understanding the place of the mind in nature. The project is carried out in terms of evolutionarily-driven biological and psychological explanations which are developed on two distinct, though interrelated, layers of conceptualization: the relationship between function (adaptation) and evolution, and the phylogeny of a species. Questions concerning the selective advantage of behavior patterns in relation to issues of survival and reproduction have informed discussions of evolutionary psychology. A prominent research paradigm in the philosophy of social science that pursues this methodology in studying the origin of cooperative behavior is evolutionary game theory (for an introduction see Alexander, 2008). With regard to collective intentionality more specifically, although there has been some debate on the evidence available on the allegedly evolutionary underpinnings of collective intentionality (Vromen, 2003), this naturalization route has not been pursued in a systematic way until a few years ago.

Since the late nineteen-nineties, the prospect of naturalizing collective intentionality has been significantly revitalized by the rise of the experimental program in cognitive science that has its theoretical foundation in the work of psychologist Michael Tomasello. By conducting threefold comparative research in primate cognition, developmental social cognition and language acquisition, Tomasello is the first researcher to have enriched the collective intentionality tradition with a rich battery of experimental findings that illuminate, and help articulate, some of the central philosophical issues of the subject. Tomasello's program ideally covers all four areas of the explanation of collective intentional behavior. It then stands out as the most mature account of collective intentionality *naturalized*, and will therefore be carefully disentangled and discussed in the second part of this thesis.

[23] I will come back to physicalism in the discussion of Searle's view of naturalism in chapter 3.

2.6 Concluding Remarks

The irreducibly collective nature of phenomena like communication and cooperation cannot be fully captured by the concepts that we use in understanding individual intentional behaviour. In this chapter I have shown how this intuition has motivated the rise of collective intentionality theory in the late 1980s. The theory has since grown around the question whether collective intentional states are irreducible to individual states. As I have argued, the notion of irreducibility lends itself to various meanings, and it is by drawing a line between conceptual and scientific reduction that I have set the ground for the discussion of the naturalization of collective intentionality.

On the one hand, philosophers have tackled the irreducibility question by analyzing collectivity concepts in ordinary language. Their analysis of we-intentions, in particular, have contributed important insights into the structure of collective action and provided a number of reasons for being realist about collective intentionality. Naturalizability arguments, however, are insufficient to settle questions concerning the existence and identity of collective as distinct from individual intentional states. In fact, while a feature can be treated as a theoretical primitive relative to a given theoretical framework, this is not equivalent to saying that the feature is simple, or that it cannot be given an account at another level of explanation. Whereas naturalizability arguments give reasons for naturalizing collective intentionality, for it to be naturalized is a matter of scientific investigation.

I have then elucidated various prospects of naturalization, according to which the scientific study of collective intentional behaviour ought to investigate proximate and ultimate causes. In recent years, significant results concerning the foundations of collective intentionality have been achieved across various sub-disciplines in the cognitive sciences like experimental psychology and social neuroscience; yet the collective intentionality literature in philosophy has hardly engaged with those lines of inquiry. Hence, I have discussed various issues arising from a scientific treatment of collective intentionality and directed attention to the research program which I shall consider in the second part of the thesis.

Three

The Sense of Collective Intentionality

Collective intentionality is the bedrock of John Searle's philosophy of society. In this chapter I shall illustrate Searle's realist approach to the ontology of mind and society: internalism, the thesis that genuine intentional states are structurally independent of how the world is like; and individualism, the view that society is nothing over and above its individual components. Collective intentionality is thus defined by Searle as a biologically primitive phenomenon of the minds of individuals that cannot be analyzed into more elementary units. Against common critiques of this 'primitivist' conception, I shall argue that the problem with Searle's account does not lie in its assumptions but in that he treats the biological nature of collective intentionality as a self-evident 'fact', which however asks for more cautious elaboration and empirical check.

3.1 Introduction

Collective intentionality lies at the foundation of John Searle's philosophical construction of social reality. The term was coined by Searle in his seminal paper on collective intentions and actions (1990/2002); a few years later, in *The Construction of Social Reality*, he defined any fact involving collective intentionality as a *social* fact (1995: 172). Searle has since been a very influential figure in establishing collective intentionality as one of the central tools to deal with the ontology of the social world (2010).

The key to understanding Searle's theory of collective intentionality is to situate it in his overall philosophical project. Although it culminates in the third stage of his research, the study of human society brings to completion a long-standing investigation rooted in earlier work in the philosophy of language and mind. Unlike other 'founding fathers' of the theory, in fact, Searle's enthusiasm for collective intentionality is motivated by broader interests than the analysis of everyday collectivity concepts in the philosophy of social science. As he often reminds us, the overarching question is to

explain how phenomena like intentionality and consciousness, as well as cooperation and the rise of institutions, find their place in "a universe consisting entirely of physical particles in fields of force" (Searle, 2010: 3). The purpose of Searle's 'Grand Philosophical Theory' (Smith, 2003) is thus to justify the place of social ontology in the more comprehensive ontological framework of the natural sciences.

Collective intentionality is one of the 'building blocks' of this project, spanning from physics to society (Searle, 1995). By claiming that collective intentionality has its roots in the biology of the brain, Searle aims at showing that the facts of the social are ultimately grounded in the human mind, thus reconciling them with the facts of physics, biology and chemistry. In its most general and controversial formulation, collective intentionality is defined as a biologically primitive property of the minds (brains) of individuals that cannot be analyzed as the summation of individual intentions (Searle, 1990/2002). They key-words of this definition are 'primitive' and 'biological'. The former refers to the two constraints that Searle wants his analysis to satisfy: all kinds of intentionality lie in the heads of individuals and cannot be reduced into more elementary units. The latter has to do with the fact that both constraints must be compatible with biological naturalism.

However, Searle takes these constraints as well as the alleged biological nature of collective intentionality as self-evident 'facts'. In claiming that all intentional phenomena are intrinsically natural, Searle endorses a form of realism about the mental: there are facts of the world that make it the case that there are intentional mental phenomena[24], be they individual or collective. So, there are plenty of aspects concerning Searle's naturalistic approach to collective intentionality that need further elaboration. In particular, it is unclear how collective mental phenomena arise from physical phenomena, and the sense in which it is argued that all forms of intentionality are compatible with (biological) naturalism.

For Searle, these are empirical questions that should be left to scientists to answer. Some questions, however, concern the method of philosophizing by which Searle makes claims

[24] As Searle writes elsewhere, in reply to Fodor, "aboutness (*i.e.* intentionality) is real, and it is not something else" (Searle, 1992: 51).

concerning the ontology of collective intentionality. We are told that "often when philosophers talk about 'naturalizing intentionality' (…) they take 'naturalizing' to mean denying the existence of the phenomena in question. So, for example, naturalizing intentionality would consist in showing that there really is no such thing as irreducible, ineliminable intentionality. (…) That is not the sense of naturalization that I am talking about" (Searle, 2007: 19). What is Searle's sense of naturalization, then?

In this chapter I will answer this question by illustrating and discussing Searle's theory of collective intentionality against the background of current debates concerning the understanding of language and mind[25]. I shall proceed as follows. In §3.2 I shall discuss various alternatives in response to the question of what makes collective intentions – or, more generally, collective intentional states - irreducible to individual states. In §3.3, I shall complement the irreducibility thesis with the second feature of Searle's account, the view that all intentionality is held in individual brains, and defend it from a number of wrong-headed criticisms. In §3.4 I shall take into consideration the notion of the Background to explain what makes Searle believe that collective intentionality is biologically primitive. Finally, I shall focus on the status of the claims that 'intrinsic' collective intentional states are not only conceptually irreducible but also biologically primitive features of individual brains. Where these two claims stand relative to each other, and whether conceptual analysis paves the way for existential conclusions about the nature of collective intentionality, will be addressed in §3.5.

3.2 Kinds of Intentionality

Human mental life manifests itself in a vast array of forms. Consider the list of mental states that Searle provides to show how pervasive the presence of intentional predicates is in everyday language and thinking:

[25] The inquiry into collective intentionality, in fact, closes a circle that began with *Speech Acts* (1969), the analysis of linguistic social phenomena, and continued later with the effort to ground the study of all intentional phenomena in the minds of individuals. In this regard, *Intentionality* (1983) offers the conceptual apparatus to examine the structure of intentionality.

> Belief, fear, hope, desire, love, hate, aversion, liking, disliking, doubting, wondering whether, joy, elation, depression, anxiety, pride, remorse, sorrow, grief, guilt, rejoicing, irritation, puzzlement, acceptance, forgiveness, hostility, affection, expectation, anger, admiration, contempt, respect, indignation, intention, wishing, wanting, imagining, fantasy, shame, lust, disgust, animosity, terror, pleasure, abhorrence, aspiration, amusement, and disappointment (Searle, 1983: 4).

The unifying feature of all these predicates is *intentionality*, which we have defined in chapter 2 as the property of some mental states whereby they are directed at states of affairs in the world (Brentano, 1874). Despite many subtle differences in their conceptions of intentionality, by and large philosophers tend to split in two opposing schools of thought. On the one side are those who endorse the idea that intentionality is not a feature of mental states *per se* but rather exists only under appropriate descriptions of it[26]. On the other side are philosophers like Searle who believe in *original* intentionality – a view according to which there is always a fact of the matter to discover about what a person means independently of any description or interpretation of it (Dennett and Haugeland, 1987). The view that intentionality is an intrinsic property of the mind is prominent in Searle's philosophy of mind and society.

Searle's general theory of intentionality inherits its structure from the theory of meaning that was proposed in the nineteenth century by Gottlob Frege (1892/1980) in the context of natural languages. Motivated by interests in the foundations of communication, Frege's theory countenances the classic picture of meaning that semanticists call 'the Millian view' after John Stuart Mill (1867). The Millian view holds that the meaning of proper names like 'Obama' is the bearer that this term denotes. Despite its simplicity, the so-called referential theory of meaning raises a number of questions that have kept the debate alive in the philosophy of language for much of the twentieth century (for a review see Reimer, 2009). Some of these questions concern the

[26] Dennett's notion of the 'intentional stance' exemplifies this strand of thought by emphasizing the constitutively interpretive nature of intentionality: states of mind that refer to something in the world can only be ascribed to a person by somebody taking the appropriate intentional stance.

identity-conditions of meaning in the contexts in which the referential role of certain linguistic expressions proves insufficient to adjudicate it. The cases at issue comprise: identity statement between co-referring expressions, existence statements, empty names in meaningful statements, and propositional attitude attributions (Devitt and Sterelny, 1999).

Faced with the cognitive and epistemic shortcomings of the Millian view, Frege introduces the notion of 'sense' to settle questions about reference in uncertain contexts. As he claims in "On Sense and Reference" (1892), there are two dimensions to the meaning of a linguistic expression. There is *reference*, which is the entity that the expression denotes; and there is *sense*, the 'mode of presentation' in which the referent is thought about. If two expressions pick out the same referent in the world, identity of meaning will depend on the sense in which each expression is presented to thought. Or, in other words, to know the meaning of a word *is* to grasp its sense (Margolis and Laurence, 2007: 565). The relation between sense and reference can be conceived of as one of 'mediation': one understands the meaning of an expression by the medium of sense, where grasp of sense consists in recognizing what the expression refers to.

It has become customary in the literature to read this claim as saying that sense and reference contribute to meaning somewhat separately. This objection is partially true, as the claim that sense 'determines' reference is ambiguous in one major respect (Evans, 1981). It suggests that there is a positive theory of sense that one ought to work out prior to identifying reference. Yet, there is no sharp line to be drawn between sense and reference: sense is to be conceived of as a way of thinking of, rather than determining or fixing, reference. "We should not expect to be given the sense of an expression save in the course of being given the reference of that expression" (Evans, 1981: 294). For example, if one person is asked to explain how she understood a certain expression, the intuition suggests that she would end up describing what is it that makes it the case that the expression has the meaning that it has, namely its referent. Of course the *description* of what the expression refers to will be from her own standpoint, which is to say that perspective will enter the meaning of the expression. But perspective characterizes the way in which the person describes the referent and not her representation of it.

This 'unified' reading of sense and reference finds its most convincing formulation in Searle's conception of *intentional content* (Searle, 1983). The idea is that all intentional states exhibit 'aspectual shape' in that the referred-to object is presented to individual minds under a certain aspect. Thus, one thinks of some referent not 'as it is' but in a certain manner (Crane, 2003). For example, there is an obvious difference between two beliefs that represent London as, respectively, the capital of the United Kingdom and the site of the Tate Modern. While the referred-to object (London) is one, the thoughts pick out two distinct aspects of London. This is to say that a difference in the intentional content of mental states reflects a difference in the way in which these contents are accessed in thought[27].

Along with intentional content, there also are different forms, or '*intentional modes*'[28], in which a state of mind can be directed at the world. As the long list of mental states showed at the outset of this section, the object of intentional states can be re-presented in different forms depending on whether it is, say, hoped for or feared by the subject. That is, the difference between hope and fear highlights forms of intentionality which present the object differently to the mind. So, in the previous example one person can believe that London is the site of the Tate Modern or, alternatively, she can hope that London rather than, say, Paris is the site of the Tate Modern. The object is the same while the mode in which persons think about them is different.

Intentional content and intentional mode are the two features that distinguish mental states which are directed to the world from those which are not[29]. However, one feature of intentional states that

[27] The notion of intentional content, in particular, has become shorthand for 'intentionality' in contemporary programs of naturalization of the mind. Briefly, to say that mental states are intentional is to say that they bear content about the world, where 'content' is roughly synonymous with 'meaning'. The upshot is to conceive of intentionality as a semantic property along with meaning and reference, with the effect that a naturalistic account of the mind is an account of intentional content naturalized.

[28] Frege's notion of the mode of presentation of a mental state is not the same as Searle's notion of 'intentional mode'. Following Searle, the former is a feature of the *content* of intentional mental states, while the latter identifies the *form* of intentional mental states.

[29] Note that, on a certain definition of intentionalism, Searle is *not* an intentionalist philosopher. 'Intentionalism' in the philosophy of mind is also a label that identifies those who think that all mental states are intrinsically intentional. A further distinction within this camp is between weak and strong versions of intentionalism, depending on whether conscious states are said to fall under the definition of intentionality or not (Byrne, 2001; Crane, 2007). On this conception of intentionalism, Searle rules out those instances of

this theory, as well as almost all accounts of intentionality in the philosophy of mind, has long ignored is *how* individuals access the contents of intentional states. The point is that philosophers commonly analyze the structure of 'belief-desire psychology' by investigating the conditions for the existence and identity of *individual* intentional states. By 'individual' I mean that it is implicitly assumed that people represent intentional contents in the first-person singular: states of the world are 'thinkable' insofar as they are given to minds in an individual, or '*I*-mode', perspective. So, the standard way to express linguistically the belief that, say, today is sunny is by means of the propositional attitude '*I* believe that today is sunny'. As we saw in the previous chapter, this assumption has come under the attack of Gilbert's and Tuomela's critiques of intentionalism in the late 1980s. But it is against the background of Searle's general theory of intentionality that this line of argumentation has received its most explicit and convincing formulation in the early stages of collective intentionality theory.

3.2.1 Collective Intentionality as a Primitive

In 'Collective Intentions and Actions' (1990/2002), Searle provides an argument for the idea that individual intentionality is just one of two kinds of intentionality. The other kind, which he illustrates by way of a counterexample known as the 'Business School Case', exhibits features that are irreducible to individual intentionality. In the following I shall take the Business School Case as the point of departure to discuss the specificity of collective intentional phenomena and their irreducibility to other kinds of intentionality.

We are asked to imagine a bunch of fresh graduates who leave business school after being exposed to Adam Smith's theory of the 'invisible hand'. In its popular characterization, the theory says that it is by pursuing their self-interest that people frequently promote the interest of society more effectively than when they intend to promote it. Graduates will then benefit humanity just by "being as selfish as each of them possibly can and by trying to become as individually rich as they

---

mental life that are not directed at something other than themselves, like forms of elation or nervousness (Searle, 1983). Searle's intentionalism should therefore be defined more loosely as a form of realism about intentional psychology in accounts of the social reality.

can" (Searle, 2010: 47-8). This interpretation makes perfectly clear that each graduate may form an intention to pursue personal interests for the sake of the public good and, yet, there is no true cooperation or collective action in spite of their having the same goal. After all, according to Searle (2010), Smith's metaphor of the invisible hand is a telling example of a selfish 'ideology', which shows that people can indeed pursue the collective goal of benefiting humanity without intending to do so in a cooperative way but purely on individualistic grounds.

Consider now the exact same case, except that now business school graduates meet the day after graduation and agree upon a special deal. They engage with each other in the promise to make the most out of Smith's lesson by helping humanity through selfish behavior. At a certain time, one of them drops out and decides to work for the Peace Corps. If this scenario happened in the first case, we are told, there would not be any relevant consequence on the conduct of the others. But in the second case such a move breaks the texture of obligations and commitments established by the 'collective' promise. If each graduate acted on the same kind of intention in both cases, that is, if *all* intentions were individual – what would bring about that texture of social norms that make the second business school case an instance of 'genuine' cooperation?

Since the bodily movements are exactly the same in both scenarios, the difference must lie in the mental component. For Searle, this component is captured by a difference in the *way* graduates intend the common target of their intentions. While in the first case they act upon individual intentions which nonetheless are directed at the same target, to make sense of the intentions that generate cooperative behavior Searle introduces a *sui generis* type of mental states: collective, or 'we', intentions. The advantage of collective intentions is to reconcile the intuition that business school graduates intend the goal of their own thoughts and actions differently across the two scenarios with the observation that their behavior is apparently the same. Besides, while Searle's proposal aligns with that of the other collective intentionality theorists on the specificity of collective as opposed to individual behavior, what distinguishes his account is the claim that collective and individual intentional states differ in *kind*.

What is the kind of collective intentional phenomena? Although he has largely elaborated on the irreducibility of collective to individual states over the years, Searle does not clearly state what aspect of the structure of mental states separates distinct kinds of intentionality. In his writings, however, there are sufficient clues to specify the structure of collective intentionality and to provide a general answer to our question. In this regard, it has been suggested that there are three possible interpretations of what makes collective intentions an instance of a special kind of intentionality (Pacherie and Dokic, 2006). The first is that they exhibit a *content* which differs from that of individual intentions. The second resorts to the *type of entities* to which collective intentions are assigned. Finally the third interpretation is that 'we-intending' underlies *a mode of thinking and acting* that diverges from the 'I-mode' of individual intentions.

Let us begin with the first option. As we said, each person entertains a state of mind with a content that represents the reference in a certain fashion. Suppose that a bunch of people gather together to perform an action together in the same way as the business school graduates do when they make the promise to achieve the collective good by acting selfishly. What would the content of their thoughts be in the case in which they intend to fulfill their engagement by acting as a plural subject 'we'? Clearly each person understands her own thought and action as directed to a goal to be achieved collectively. But how are the two 'perspectives', the individual and the collective, represented by the single agent in one and the same mental state? There seems to be a tension between the representation of the goal that each person intends to pursue by doing her own part, and that of the plural subject that undertakes the collective effort as a whole.

Searle tackles this problem by claiming that the intentions of the individual participants in the collective action are related to the overall goal in the same way as singular intentions relate to certain actions as means to ends. The typical example is that of firing a gun by pulling the trigger, in which the intention of firing a gun *is* fulfilled by pulling the trigger, so there is only one intention and one action (Searle; 1990/2002: 99). Similarly, in the collective case, the intentions of each person constitute one whole with the goal to be pursued together. Insofar as the features of the content of collective intentional states are already present in that of individual states, the specificity

of collective intentional phenomena is not a matter of intentional content. As Searle claims: "In collective intentionality I have to presuppose that others are cooperating with me, but the fact of their cooperation is not part of the propositional content of my part of the collective intentionality; rather, it is specified in the form of the collective intentionality" (2010: 53). Although the thinking subject is plural, the content of collective intentions is still 'individual' in that it represents the reference of each person's thought and action.

This conclusion suggests that the *sui generis* form of collective intentions may be a feature of the subject to which they are ascribed. If I-intentions are typically assigned to individuals, we-intentions would fall under the scope of 'collective subjects'. But this ontological possibility sounds unreasonable in light of Searle's commitment to the tenets of individualism. In giving the criteria that any account of collective intentionality must satisfy, Searle holds that it must be consistent with the fact that "society consists of nothing but individuals" (1990/2002: 96). Therefore, it is not ontologically acceptable that collective intending is the deliverance of Hegelian world spirits or group minds of some sort (Searle, 1995: 25). The subjects of collective intentions must be the individual subjects.

Rather than invoking some collective entity, one could claim that the bearers of collective intentions are special because their individual *minds* relate in particular ways with each other. The specificity of collective intentionality would then consist in the kind of connections that the single agents form in the act of sharing intentions. We have already encountered a version of this theory in chapter 2 while discussing Bratman's individualist conception of shared intentions. Recall that, for Bratman, first-person singular intentions must be interrelated and supplemented with mutual knowledge for their bearers to collectively intend to do something together (Bratman, 1993). Since Searle and Bratman are both committed individualists, let us evaluate whether collective intentionality arises from the interrelation between I-intentions plus mutual beliefs.

Back to the Business School Case, suppose that each graduate pursues her self-interest knowing not only that the others have been exposed to the same lesson, and therefore will act in the same way, but that *this* is mutual knowledge among all of them. There are at least two reasons to believe

that the final state is not one of truly collective intentionality. To begin with, Searle holds that individual intentions supplemented with mutual beliefs do not amount to the 'sense of collectivity' that accompanies we-intentions. But there is a better-informed, cognitively-based, reason for discarding the possibility that individual subjects need only share mutual knowledge for collective intending: the mutual knowledge-approach ends up over-intellectualizing the mind.

In a two-person situation, in fact, the mutuality of knowledge consists in a state that both persons entertain when 'I know that you know that I know that you know that...' about the target of the joint action. The alleged state of collective intentionality would thus be reached asymptotically through an inferential and iterative chain of high-degree epistemic states. But human cognition, as decades of research on the so-called 'bounded rationality' hypothesis confirm, is limited in ways that make it impossible for people to handle propositional states for more than three or more degrees of processing. "The mere presence of I-intentions to achieve a goal that happens to be believed to be the same goal as that of other members of a group does not entail the presence of an intention to cooperate to achieve that goal" (Searle, 1990/2002: 95).

Thus, the third and last option available is that collective intentions underlie a distinctive *mode* of thought and action. For individuals to intend in 'we- modality' is equivalent to represent aspects of the world from an intrinsically collective perspective. Notice that 'intrinsic' entails that collective intentional states are irreducible to individual states. But Searle also uses another term to characterize the irreducibility thesis: he claims that collective intentionality is a *primitive* psychological trait (Searle, 1997). Since the notion of primitive is paramount in Searle's theory of the social reality, some clarification is necessary.

'Primitive collective intentionality' is open to a twofold interpretation. The main motivation for believing in the irreducibility of collective intentions stems from the failure of reductivist accounts like Bratman's to make sense of the inter-subjectivity of human thought and action in a non-circular way. Accordingly, collective intentional states cannot be assembled out of more elementary units, notably individual states and their interrelations (including I-mode mutual beliefs), so they are primitive. However, Searle seems to suggest a more challenging motivation for the view that

collective intentional phenomena underlie a primitive form of mental life: "The crucial element in collective intentionality is a sense of doing (wanting, believing, etc.) something together, and the individual intentionality that each person has is derived *from* the collective intentionality that they share" (Searle, 1995: 24-5; emphasis in original).

A better formulation of the relation of 'derivation' between collective and individual intentional phenomena would be to say that, when people represent aspects of the world in the 'we-mode' they access the contents of their representations in the first-person plural *prior* to first-person singular representation. But the concept of priority is hardly mentioned in Searle's work. It would however be very important in turning his argument against reductivist accounts into a *positive* theory of the irreducibility of collective intentionality. So there is no convincing support for this argument except for the basic intuition: "Intuitively, in the collective case the individual intentionality, expressed by 'I am doing act A', is derivative from the collective intentionality 'We are doing act A'" (Searle, 1990: 92). One might then ask why Searle does not spell out the idea as he should.

In any case, Searle pursues another route in spelling out the meaning of 'primitive'. The irreducibility thesis is not the only defining feature of Searle's project of 'individualizing' collective intentionality (Meijers, 2003).

3.3 Collective Intentionality without Collectivity

The second condition that Searle proposes for his account of collective intentionality is that:

> 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, 2002: 96).

This constraint brings to the fore a central element of Searle's philosophy: 'methodological solipsism'[30]. This is the view that all intentionality persons can 'have', no matter whether individual or collective, is not only primitive but also internal to the brains of individuals. Internalism has stimulated a heated debate in the philosophy of mind in the last quarter of the twentieth century, which revolves around the conditions for the existence and identity of intentional contents. The consensus nowadays, especially among those philosophers who aim at providing naturalistic accounts of the mind, is to embrace forms of externalism with regard to specific mental phenomena including collective intentionality. Searle's internalist stance has thus been criticized and discarded on several counts. Before we turn our attention to these critiques, let us briefly review the general problem.

Consider the sense-reference distinction: Which one among the actual object referred to and its aspectual shape individuates the intentionality of a mental state? In other terms: when people represent a state of the world, what feature of the state individuates the content of the representation – the actual object or the aspect under which the content appears in mind? *Internalists* hold that intentional contents are 'narrow': the existence and identity of contents do not entail the existence of their intentional object. So, when we think about something in the world, it is not necessary for us to represent the content of our thought that the thing actually exists in the world. *Externalists* challenge this view with the argument that intentional contents are 'wide': the existence of the intentional object is necessary for the existence and individuation of thought-contents. We cannot think about something if there is *nothing* to be thought about. This way to formulate the problem shows that the divide between internalist and externalist intentionality is primarily about the conditions of 'thinkability' of a thought about a particular object, what makes its content available in mind.

In *Intentionality* (1983), Searle provides a thorough defense of internalism by drawing on arguments from Gottlob Frege, Bertrand Russell and Peter Strawson. The idea is that, for a thought

[30] But note that, despite this expression in Searle (1990/2002), solipsism is an ontological, rather than methodological, thesis about the nature of the mind (Schmid, 2009).

to be about any entity or state in the world, it is not necessary that the entity actually exists – as it is the case with non-existent entities which are nevertheless 'thinkable' like, famously, unicorns or the King of France. What is crucial to identify the contents of mental states is the aspect under which they are presented to individual minds. Searle calls this aspect 'descriptive content' and defines it as the totality of mental content that is made available to the subject by simply representing the referred-to object. But it is not necessary for somebody to associate a description with a given expression that the content be given 'in words'. Such a description must not be intended in the sense of a linguistic description; it rather includes the set of necessary and sufficient conditions that individuate the entity referred to by the relevant expression. In this respect, Searle's internalism is influenced by issues of linguistic meaning[31] but, in fact, it is by no means confined to the philosophy of language only.

3.3.1 Brains in Vats Thinking Collectively

Internalism is the second condition that any theory of collective intentionality should satisfy. It is often characterized as the 'brain in a vat' constraint to emphasize that all intentionality can be had by a brain in a vat even if it happens to be radically mistaken about the world. This is to say that genuine intentional states are structurally independent of what the world is like: for individuals to refer to something 'out there', as it were, it is not necessary that their brains be in any relation with anything external. All intentionality is inside the minds (brains) of individuals.

This view has a striking consequence in the case of *we*-mode intentional states. It implies that the latter have their individuation-conditions set out independently of the existence of the *real*

[31] Another name for internalism is *descriptivism*: the sense of a proper name or natural kind term is the description that uniquely determines its reference when associated (by the speaker) with the name (term). The philosophical research on reference has long taken proper names as its paradigmatic case study. Proper names form a 'genuine' class of referring expressions because they refer to, or purport to refer, to particular objects and individuals (Reimer, 2009). While it is important to distinguish between types of linguistic expression in dealing with issues in the philosophy of language, discussions of the metaphysics of meaning are generalized to linguistic expressions of *every* semantic category. Along with proper names, natural kind terms have received most attention: natural kind terms are expressions that purport to refer to objects and properties that typically fall under the inquiry of the natural sciences.

persons that form the plural subject 'we'. It might be argued that brains in vats cannot think as a team, because we need others for sharing thoughts and actions whenever we entertain first-person plural states (Schmid, 2003). Therefore, since the main justification of collective intentionality is to explain social phenomena that involve cooperation among individuals, ranging from communication to the nature of institutions, it is very challenging to make sense of the idea that collective intentionality exists without collectivity.

Many philosophers find this view unintelligible precisely because of the brain-in-a-vat constraint. How might ever be possible that any individual can think as the member of a group if there is not at least another individual that forms the group? Searle is categorical on this point: "I could have all the intentionality I do have even if I am radically mistaken, even if the apparent presence and cooperation of other people is an illusion, even if I am suffering a total hallucination, even if I am a brain in a vat. Collective intentionality in my head can make a purported reference to other members of a collective independently of the question whether or not there actually are such members" (1990/2002: 97). How are we to understand this claim? Since the internalism-externalism debate is too broad to be settled in the short space of this chapter[32], and because my own sympathies are with internalism, I will only analyze arguments that help establish the *coherence* of Searle's view of collective intentionality. Against the critics that take this view as inconsistent, I will argue that collective intentionality can perfectly be understood in an internalist way, while remaining neutral on the issue of whether this reading must be privileged to the externalist one.

The literature offers two lines of interpretation, which spring from a weakness of Searle's account. Unsurprisingly, the point under attack is the idea that brains in vats can be radically mistaken in thinking to have we-intentions when, in fact, there is no plural subject to which they can be ascribed. In discussions of individual intentionality, to say that an individual is mistaken means that she may be wrong in thinking that something is the case. Namely, there are no truth-conditions in the world that satisfy the content of her thoughts. But the specificity of collective

[32] I will come back to this point more extensively in chapter §6.

intentional states lies in the kind of thinking rather than in their content. So, there is room for the possibility that individuals can be mistaken *both* at the level of the content *and* at the level of the psychological mode of intentionality. Searle clearly acknowledges this possibility but leaves it somewhat unexplained (1990/2002: 98), thus exposing his theory to a host of externalist attacks.

One line of interpretation can be subsumed under the argument elaborated by Elisabeth Pacherie (2007). On Pacherie's construal, the internalist defends the view that brains in vats can have intrinsic collective intentional states. A tension arises, however, between the claim that one can be radically mistaken about having 'we' states, namely one *thinks* to have them when it is not the case, and the claim that there are such states in people's heads independently of how things are in the world. If the brain in a vat is wrong in thinking that it has a collective intentional state, then clearly it does not have it (Pacherie, 2007: 161). Hence, *either* the brain is wrong to the effect that it has no 'we' state *or* it has it. But to conclude that one can be left with the illusion to have collective intentional states is self-defeating for the internalist who believes in the existence of genuine 'we' states. Notice that the whole critique is founded on the presupposition that the internalist wants to show that collective intentionality is a fact of the matter in individual brains.

This construal reveals a misunderstanding of Searle's internalist stance. As we mentioned before, internalism is about the conditions of 'thinkability' of thoughts, including we-thoughts. That there exists such a 'thing' in the world as collective intentionality in the case of a veridical thought is purely incidental to the very act of thinking and acting in *we*-modality. So, it is not correct to relate the claim that one may be radically mistaken about having collective intentional states with the question whether this proves that they are, or not, inside the brains of people. This is beyond the internalist lesson, which is to explain what makes it the case that people *can* think in *we*-modality, no matter how veridical the thoughts about that are. To put it differently: the point is *logical* in that it concerns the conditions of possibility of collective intentionality, rather than the truth-conditions that must be satisfied for people to have genuine collective intentional states in the head.

This is an aspect of great relevance in understanding internalism, though it is too often neglected by critics. Searle would never deny that real interaction with the others and the world is essential to

the recognition and discrimination of thought-contents. And this requires sharing background skills and experience at the socio-cultural level as well as being connected with the natural environment in the appropriate way. But this is an empirical fact, which lacks *logical* necessity. The explanatory force of the brain-in-a-vat constraint is most evident precisely in the circumstance in which none of the conditions of satisfaction for having collective intentional states are actually fulfilled. While externalists like Pacherie are interested in exploring the conditions for the existence of collective intentionality in this world, internalism is a thesis about the identity-conditions of intentional mental phenomena across possible worlds or counterfactual situations (Crane, 2001).

The main proponent of the second line of interpretation, Anthony Meijers (2003) is well aware of the point about the 'logical' nature of internalism. Like the other externalists, his critique makes a start from Searle's intuition that collective intentional states can be mistaken in two ways, but he develops the point in a novel direction. In fact Meijers belongs to the group of collective intentionality philosophers who closely follow Gilbert (1989) in claiming that collective intentionality is not entirely a matter of cognitive attitudes. What distinguishes the irreducibly collective dimension of behavior is the fact that social phenomena arise from normative attitudes between the participants. So, the structure of collective intentionality consists in the social relations that only get formed when the subjects agree upon them. And this requires that the bonds that manifest in binding claims and obligations have a 'foundation' in the reality. "Having a foundation means that the intentional states are one-sidedly dependent upon *two* or more participants. In case these participants do not exist in the real world, there is simply no collective intentionality" (Meijers, 2003: 179; emphasis in original).

This argument is a defense of externalism against accounts that purport to give the conditions of collective mental phenomenon in abstraction from their truth-conditions. But Meijers correctly proves that Searle's view is defective in a more subtle way: it cannot explain the structure of social reality unless it first clarifies what it means to share intentional states. In other words, Searle's theory is a theory of shared rather than *sharing* intentionality; one that tells what it is to have states

in *we*-modality rather than how they get shared[33]. It follows that, *if* a theory of sharing intentionality was spelled out internalistically, it would certainly fail to make sense of the external, interpersonal relations that ground the final state of collective intentionality.

But the question, then, becomes: What is a theory of sharing intentionality? Meijers seems to believe that it must be a theory of shared content along the lines proposed by Bratman (1993) or Velleman (1997) among others (Meijers, 2003: 175*ff*.). On the interpretation discussed in §3.2, intentional or mental content is the content of the thoughts that people entertain when they refer to something in the world. Contents would then be shared when individuals have access to the same, or similar, representations of the reference of their thoughts, as a basis for making interaction and cooperative behavior possible. Is this what Searle means by his theory of collective intentionality? As we said, Searle would never deny that the reference of intentional states is recognized, and shared, in interaction with the others and the world. In fact, as we argued above, what must be internal to individual minds is not the conditions by which people share contents, but the first-person plural *mode* of behavior that makes the sharing logically possible in the first instance. So, the problem with Meijers' analysis is not lack of appreciation of the logical character of Searle's analysis, but the unmotivated emphasis on the aspect of content instead of *we*-mode.

In conclusion, Searle's internalist approach to collective intentionality fares better than most externalist-leaning critics maintain. In this section I have contrasted reasons for giving up internalism with arguments that restore the intelligibility of the claim that collective intentionality can exist without collectivity. Yet, critiques of internalism bring to the fore some evident limits of Searle's analysis that need more discussion. If our capacity to think of individuals – both ourselves and others - in the first-person plural is independent of there being an actual 'we', what does it rely on? The general dissatisfaction towards Searle's response is due to the contention that collective intentionality is just an intrinsic feature of the mind; this claim appears question-begging if it is not

[33] The picture on page 26 of *The Construction of Social Reality* is a telling proof of the synchronic dimension of Searle's analysis: it represents the heads of two individuals containing we-intentions. I agree with Meijers that the picture is also misleading in that it shows *two* persons having a collective we-intention, when the point is just to show that this must not be the case (Meijers, 2003: 182).

developed into an explanation of what brings about the very capacity of collective intentionality. To address this problem, we turn to the final aspect of Searle's account: the Background.

3.4 The Background

Searle does not refer to collective intentionality just as a primitive feature of the mind. He defines it as a *biologically* primitive form of mental life. What does he mean by 'biological'? First notice that this is not an attribute of collective intentional phenomena in general but of the *capacity* to engage in collective behaviour, which consists in "something like a pre-intentional sense of the 'other' as an actual or potential agent like oneself in cooperative activities" (Searle, 1990/2002: 102-3). The capacity for collective intentionality is thus a natural tendency towards cooperation that, according to Searle, can be observed in other animal species as well and is "biologically *innate*" (Searle, 1995: 37; emphasis mine). By 'innate' Searle intends that this sense of community is the outcome of processes of biological evolution rather than of cultural or linguistic acquisition, and has its roots in brain structures that function causally in enabling us to engage in social endeavors.

According to Searle, the 'pre-intentional' sense of community is part of the *Background*[34]. The Background is a technical notation in Searle's philosophy, which was originally introduced to settle issues concerning the understanding of meaning, starting with the literal meaning of a sentence (Searle, 1978; 1983). The 'literal (or sentence) meaning' differs from the so-called 'speaker meaning' – the message the speaker wants to get across by uttering a certain sentence - in fixing what the sentence means if understood literally. Obviously, the single components of a sentence may be insufficient to reconstruct the meaning of the sentence at large. In particular, once we get rid of all the possible sources of ambiguity such as metaphors and indirect speech acts, the sentence is still open to various interpretations because the "sentence meaning *radically* underdetermines the content of what is said" (Searle, 1992: 181; emphasis in original). What is it needed to reveal the

[34] I follow Searle in using the capital letter when I refer to the theory of the Background to distinguish this use from the ordinary conception of background.

full content of a sentence, then? Searle suggests that understanding of language requires us to *take for granted* a number of aspects which are not overtly present to the mind. These aspects literally stand on the background of our thoughts and, yet, they enable grasp of sentence meaning with no apparent interpretative effort.

The idea that there is an underlying layer of competences that enable intentional phenomena to function is the core claim of the Background theory. But it is important not to misunderstand Searle's words here. In fact, it is precisely the causal role played by background capacities that has undergone a wealth of critiques (Stroud, 1991). What aspect of intentionality does the Background enable exactly? A great deal of confusion arises from the fact that, as Searle makes it plain in *Intentionality* (1983), the inquiry into the nature of language is part and parcel of the study of the mind. So, linguistic phenomena have their intentional character grounded in the intentionality of the mind. This is to say that understanding of linguistic meanings and understanding of mental contents are closely tied. The question, then, is what determines the content of intentional states.

This question can be tackled in two different ways. One is to explain what makes it logically possible for people to access the contents of their thoughts. This approach, which animates the internalism-externalism controversy, traces back to Frege's discussion of the problem of communication. For the time being let me point only to one feature of this debate. For internalists like Frege and Searle the content of intentional states is construed in terms of the aspect under which the mind picks out its referents. Yet, if reference can only be individuated via the grasp of sense, where to grasp is an intentional act and then inherently perspectival, communication will logically be impossible if not by happenstance. Frege interprets this problem as a demand for a mind-independent conception of sense: senses are abstract, truth-bearing entities with a separate ontological status from mental entities. If senses are external to minds, something objective that people can grasp inter-subjectively, communication becomes possible despite the perspectival nature of intentional contents. But Frege's proposal is unsuitable on various grounds[35]. As it is clear,

[35] The motivation for Frege's Platonism is *not* to explain grasp of sense in terms that do not involve the aspectual shape of intentional content. It is rather to avoid the threat of 'psychologism', the idea that

the problem is not whether people succeed to refer to the same entity in communicative exchanges, which is easily fixed by simply noticing that miscommunication is the exception rather than the norm in everyday interaction. The problem is how one person can ever be in the position to know what is going on in another mind on an 'objective' basis. We seem to be trapped in what Searle has often called an intentional circle (1983).

The Background theory is *not* a response to this problem. Background capacities are causally relevant in that they enable the functioning of intentional phenomena, but they do not fix the way in which each individual understands the contents of her mind. As we mentioned above, the meaning that each person 'attaches' to her thoughts depends on the interaction with the others as well as with the environment. In an important passage, Searle shows awareness of this distinction and claims that "the Background functions causally, but the causation in question is not determining. In traditional terms, the Background provides necessary but not sufficient conditions for understanding, believing, desiring, intending, etc., and in that sense it is enabling and not determining" (Searle, 1983: 158). So, the Background does not stop the regress of interpretation that derives from the intentional nature of understanding; "the only thing that blocks those interpretations is not the semantic content but simply the fact that you have a certain sort of knowledge about how the world works, you have a certain set of abilities for coping with the world" (Searle, 1995: 131). This peculiar kind of knowledge that makes intentional contents immediately intelligible is procedural and constitutes the Background.

3.4.1 The Sense of the Other

---

understanding of meaning ultimately depends on some private, idiosyncratic, mental representations (Frege's conception of mental representation differs in relevant respects from the notion in use in contemporary philosophy of cognitive science).Why suppose that an account of senses as public and objective is not jeopardized by psychologism? Senses must be grasped in order to be shared, in fact. But Frege's theory falls short of explicating what kind of cognitive relation holds when distinct minds grasp senses under the same aspect. If senses are mind-independent no less than any other entity external to the mind, then it will be unclear in what respect recognition of senses should be different from grasp of mind-independent entities. In the same vein, we ought to have different modes of presentation for any given sense (Margolis and Laurence, 2007). Thus, Frege's view does not break into the intentional circle, unless grasp of sense is *itself* spelled out in terms that do not resort to any attribute of intentional psychology.

In *The Rediscovery of the Mind*, Searle defines the Background as a set of "mental capacities, dispositions, stances, ways of behaving, know-how, savoir faire, etc, all of which can only be manifest when there are some intentional phenomena, such as an intentional action, a perception, a thought, etc." (1992: 196). As such, the Background consists in a bunch of brute physical, causally-defined capacities at the brain level which enable mental and, then, derivative (*i.e.* linguistic) intentionality.

The relation between the Background and collective intentionality is not only one of neuro-physiological causation, though. Searle argues that there is also another characterization of the proximate cause of collective behavior to be cashed out in accordance with the theories and the findings of evolutionary biology. It is highly controversial whether there is robust evidence in support of the claim that the sense of collective intentionality is immediately responsible for the emergence of such phenomena as cooperation and altruism on the evolutionary scale (Vromen, 2003; Rakoczy and Tomasello, 2007). For the time being, however, the task is not to assess Searle's claim in light of the evidence available[36]. I am rather focusing on what follows from the claim that the sense of collective intentionality belongs to the (neuro-physiological, evolutionary) Background of cooperative behavior. Two aspects are worth of consideration in particular.

The first aspect concerns the thesis that the Background underlies collective intentional phenomena and, hence, the ontology of social reality. Provided that the Background consists in causal capacities, Searle seems to suggest that cooperation and the very nature of sociality escape intentionalist explanation. What would justify use of collective intentionality as the central tool of social ontology, if it cannot be given an explanation in folk-psychological (*i.e.* intentionalist) terms (Pacherie, 2007)?

This problem is partly due to the vagueness surrounding the notion of Background since it first appeared in scattered remarks by the late Wittgenstein (1953/2001). More in detail, the problem is that we only have at disposal an intentionalistic vocabulary to account for something which

[36] Focusing on the research program carried out by Tomasello and his collaborators in developmental and evolutionary psychology, this issue will be explored in detail in the second part of this thesis.

allegedly stands outside of the framework of intentional psychology. What we would need is, on Searle's account, a second-order non-intentionalist vocabulary allowing description of the Background in non-folk psychological terms. Besides, the problem seems to be that the Background is not an intentional phenomenon itself. However, physical know-hows lie on the background of the mind not because they are intrinsically non-intentional[37], but because they are *pre*-intentional. To say that background capacities are not intentional is not equivalent to denying that they can be represented. The intentionality of the Background is 'potential' depending on the fact that it rises to the level of conscious processing. When this happens, and we therefore become aware of the existence of the Background, background capacities have already entered the content of mental representations.

Let me clarify the point by way of a simple example. Driving a car is a very demanding and multifaceted task, although it is often not experienced as so complex once the rules of conduct slip into the background of thought. People usually perform a number of other things while driving, from entertaining thoughts to eating to engaging in conversation. In a way, we don't seem to be *aware* of driving a car when, in fact, we do it. In this context the Background represents a reasonable explanation of why we perfectly succeed in getting to the final destination. As Searle claims: "The Background not only shapes the application of the intentional content – what counts as 'driving to work', for example; but the existence of the intentional content in the first place requires the Background abilities – without a terrific apparatus you can't even have the intentionality involved in 'driving to work', for example" (Searle, 1992: 195).

The second aspect that deserves attention concerns Searle's naturalism. The Background theory helps us understand in what respect collective intentionality is said to be biologically primitive. But this hardly amounts to a theory of collective intentionality *naturalized*, namely one that gives the conditions for naturalizing collective intentionality. In fact, this is not what the theory of the Background is designed for. An argument that postulates a set of neuro-physiological as well as evolutionary conditions enabling collective behavior is not equivalent to one that actually gives the

[37] Like episodes of elation or nervousness according to Searle's definition of intentionality (1983).

specific conditions that are responsible for bringing about collective intentionality. For naturalistic philosophers like Searle, after all, this is precisely the job of natural scientists. The concept of Background is only aimed at shedding light on what it *means* for people to grasp intentional contents in *we*-modality from 'within' the intentionalist framework. And this, obviously, does not amount to identifying the proximate causal roots of intentionality, both individual and collective.

It is because the theory of the Background points to some mechanisms which are irreducible to intentional predicates, and are therefore primitive in the general theory of intentionality, that the very same mechanisms can further be specified in a language that does not involve intentional psychology. As we discussed in chapter 2, these considerations are supported by the very meaning of irreducibility in the sciences of the mind: a trait is primitive relative to the theoretical framework in which it is postulated. The next question is, therefore, why Searle believes that there may be an explanation in the natural sciences for the idea that collective intentionality is a biological fact.

3.5 Conceptual Analysis and Scientific Reduction

What is the motivation for claiming that collective intentionality is a biologically primitive phenomenon? In Searle's writings, this assertion is always accompanied by the proof that collective intentional states cannot be reduced to or eliminated in favour of something else (1995: 24). Why does Searle relate the irreducibility thesis with claims about the alleged naturalness of collective intentionality?

This question is relevant to assess a very common criticism against Searle's theory. Philosophers want to know what justifies placing collective intentionality among the brute facts of the brain. But since Searle takes this 'fact' to be commonsense along with his commitment to the irreducibility thesis and the argument for internalism, most critics cast serious doubt on the cogency of this line of reasoning. In the absence of empirical evidence, so the argument goes, any conclusion about the place of collective intentionality in the natural realm remains "magical" (Hornsby, 1997: 432). What seems unmotivated, in particular, is the explanatory role of conceptual evidence in drawing

naturalistic claims, *as if* the existence and identity of collective intentional states were postulated on the basis of linguistic analysis alone.

This critique is not totally unjustified. Recall from the discussion of naturalization projects in chapter §2 that conceptual analysis comes in two forms, depending on its role in philosophical practice. On the 'thin' interpretation, conceptual analysis is deployed as part of scientists' job of constructing and assessing empirical theories. On the 'thick', Canberra-style interpretation, conceptual analysis is essential to setting the agenda for drawing metaphysical conclusions about reality. Which interpretation does suit Searle's approach? At first glance, although Searle's philosophizing underlies belief in the continuity between conceptual analysis and scientific practice (Searle, 2007), it is ordinary language analysis that imbues his theory of intentionality.

There are a number of reasons for assimilating Searle's project to the thick view of conceptual analysis. Some have to do with his intellectual biography[38], while others emerge from his characterization of intentionality. With regard to intentionality theory, Searle holds that in order to have a full comprehension of the mind one has to address two kinds of *prima facie* independent questions: the "logical/philosophical questions (for example, What exactly is the logical structure of intentionality?)" and "the biological questions (for example, How exactly are intentional states caused by brain processes?" (Searle, 2007: 7). Unfortunately, whenever Searle refers to collective intentionality, these two kinds of questions are conflated in ways that lead critics to doubt the soundness of his conclusions.

It is precisely the conflation of logical and biological questions that has gained Searle the charge of inconsistency. To address this challenge we need to specify in more precise terms whether Searle believes that analyses based on our ways to conceive of, and describe, collective intentionality tell us something about its *actual* nature and structure. One way to proceed is then to examine separately where Searle's attitude towards conceptual analysis, stand relative to his naturalistic

[38] This aspect reflects Searle's own background as a philosopher of ordinary language under the mentorship of British philosophers J.L. Austin and P. Strawson. Ordinary language analysis defines a certain way of philosophizing that developed in Oxford during the nineteen-fifties, stressing the relevance of common sense speech over more abstract approaches to traditional philosophical problems. The core thesis is that philosophy is to be done by focusing first on how words are used in everyday language.

claims in order to detect any possible relation. Searle's stance on the issue of reductionism, in fact, exhibits peculiar features. Earlier we saw that reductionism (in both its conceptual and scientific variants) has established itself as the privileged method of naturalization in modern philosophy. But reductionism has various meanings: to identify which one Searle hangs on becomes essential to understand what it means for him to address the logical and the physical/biological issues of collective intentionality.

In general, reductions are divided in those that purport to eliminate the target of the reduction and those that re-define the target without eliminating it. The former are used to designate the program in the philosophy of mind called 'eliminative physicalism' whereas the latter indicate 'reductive physicalism'. Although both are forms of physicalism, their proponents' attitudes toward the ontology of the mental are significantly different. The thrust of eliminative physicalism, as exemplified in the works of Paul and Patricia Churchland[39], is that folk-psychological patterns of classification of mental facts are fundamentally *wrong*: mental properties and events do not exist. Hence, the only things one can find in nature for the 'mental' are neural states.

Reductive physicalism, instead, is elucidated by the psycho-neural identity theory[40] according to which the mental is identical with the physical[41]. Unlike eliminativists, the identity theorists do not deny the existence of the mental. "If any identity claim 'A = B' is to be true, then A and B must both exist" (Crane, 2001: 53). But, then, when the target entity (A) is proven to be identical to another entity (B) in the sense that A is reduced to B, what is the sense in which we still refer to A *and* B? The answer points directly to the difference between the issues that concern the logical

[39] The classic reference is to Paul Churchland's "Eliminative Materialism and the Propositional Attitudes" (1981).

[40] The identity theory has its champions in C.C.J. Smart (1959), David Lewis (1966), David Armstrong (1968) and Donald Davidson (1970) among the many others.

[41] The identity between the mental and the physical is thought of as coming in at least two forms: 'type-identity' theory, when the identity is postulated at the level of mental properties, and 'token-identity' theory, when it is postulated at the level of mental events (instances of properties). The distinction does not bear crucial relevance for our purposes. Also notice that my use of the term 'identity' is broad as I have clarified in §4, in that I refer to identity *and* functional reductions together (see footnote 12).

structure of intentional mental phenomena and those about their realization in the biology of the mind. I examine this difference in two steps, beginning with Searle's view of naturalism.

3.5.1. Deconstructing Biological Naturalism

In the contemporary debate on naturalism Searle stands out as a committed physicalist. At the beginning of *Making the Social World*, he reminds us that the overarching question of his lifetime investigation is how to explain that in a universe of physical particles in fields of force there can be such things as "consciousness, intentionality, free will, language, society, ethics, aesthetics, and political obligations" (2010: 3). However, physicalism is such a broad and problematic concept to encompass significantly different ontological claims. Searle's stance is unique in this respect, as he clearly acknowledges when he claims that "in developing a naturalistic philosophy we can begin by rejecting both the reductionism and the eliminativism of traditional materialism" (2007: 26). The result is a form of naturalism dubbed '*biological naturalism*', which is undoubtedly inspired by, but not restricted to, physicalism.

One way to bring the thrust of biological naturalism into light is by answering the question how naturalism has become the dominant position in philosophy nowadays. A close look at the history of post-seventeenth-century science shows that belief in naturalist doctrines has evolved in response to "the received scientific opinion about the range of causes that can have physical effects" (Papineau: 2007: 4). In other words, it is the long quest for understanding the causes of natural things, what produces spatiotemporal effects in the real world, which has fostered commitment to the view that the most truthful picture of nature is the one provided by the most successful science. It then turns out, interestingly, that the default naturalist doctrines held until the twentieth century recognized a pluralist range of causes as the source of physical effects. One example is *sui generis* causes like vital forces in the biological realm.

Yet, during the last century a novel consensus has emerged in the scientific and philosophical community, which posed more restrictive constraints on the scope of causal influence. The range of constraints on naturalist categories was thus limited to strictly physical causes. Ontological

naturalism has then come to signify the view that anything that makes a causal difference in the reality must be physical (Papineau, 2007: 6). This shift in the conception of naturalism is particularly evident in the sciences of the mind where the problem *par excellence*, *i.e.* the mind-problem problem, has come to be conceived of as the problem of mental causation.

Psychophysical causation is loosely referred to as the relation of 'making something happen' in the world. So, the point about mental causation is that mental phenomena make things happen by bringing about physical effects in the world. Besides, the causal picture of the mind tells us that human agency is prompted by the activity of the mental. But in virtue of what should intentional states be picked up as the causes of bodily movements? What kind of causes are they? As we said with regard to the history of modern science, in principle we would not be prevented from thinking of special mental facts as non-physical, 'rationalizing' reasons for action. But the acceptance of *sui generis*, read non-physical, causes in the context of mental causation would turn out to be at odds with the received scientific world view. Therefore, since naturalism entails belief in the posits of the best scientific theories available as well as in their methods of empirical investigation, from the claim that every spatiotemporal effect must come about through purely physical causes it follows that the only way for the mental to be causally efficacious is to be physical. The premise of this argument, namely the idea that every cause of physical effects must itself be physical, is known as the 'closure (or completeness) argument'.

The argument for the completeness of physics is the clue to Searle's physicalism. His belief in naturalism, in fact, is grounded in his acceptance of the mechanisms of physical causation and the determinism of physical laws. The closure argument thus turns out to be the essential metaphysical constraint on the relation between allegedly non-physical, like mental, phenomena and their effects in the real world. Therefore physics is complete, not in the sense that there is no further progress for the discipline to be achieved in the future, but in that complete physics provides a full-blown account of the range of mechanisms governing events and states in the natural realm[42].

[42] Searle would subscribe to Tim Crane's definition of the rationale of physicalism: Tim Crane summarizes the point as follows: "Physicalism asks us to address the ontological question in this way: see what physics

In light of this, Searle describes the core thesis of his naturalistic philosophy as follows:

> Mental states are as real as any other biological phenomena, as real as lactation, photosynthesis, mitosis, or digestion. Like these other phenomena, mental states are *caused* by biological phenomena and in turn cause other biological phenomena. If one wanted a label one might call such a view "biological naturalism (Searle, 1983: 264; emphasis mine).

Searle's realist attitude towards the mind is thus inspired by a belief in the causal continuity between the biological reality of the mind and the power of mental properties to yield physical effects. Since causality is generally seen as a "natural relation between events in the world", Searle conceives of the project of "intentionalizing causality" as the crucial step toward naturalizing intentionality (Searle, 1983: 112). More in detail, the problem of articulating a naturalistic explanation of the mental is an empirical task which consists in explaining how "mental states are both *caused by* the operations of the brain and *realized in* the structure of the brain (and the rest of the central nervous system) (Searle; 1983: 265; emphasis not mine). The moral is that any acceptable program of naturalization of the mind ought to rely on intentional causation as its key working hypothesis.

We are now in a better position to analyze the origin of the argument that collective intentionality is a primitive feature of the biology of the mind. The closure argument states that any property responsible of causing effects in the real world must be realized at the physical level. The causal view of the mind holds that people act out of their thoughts when planning what to do and how to achieve their goals. Together, these two theses lead to the conclusion that mental facts are physically constituted. Namely, there must be a fact of the matter for one's being in any state of the

---

says there is, and then commit yourself to *that kind of thing* being all there is. As time develops, it may be that your commitments develop too. But this is just a reflection of the fact that you have no standard (other than physics) from which to answer the question of what there is" (Crane, 2001: 47).

mind including collective intentional mental states, if the outcome is a change in the spatiotemporal order.

In sum, existential conclusions about collective intentionality are *metaphysical* claims that derive from the belief in the completeness of physics, and *not* from the results of linguistic analysis of individual states. How to justify these conclusions is therefore an empirical matter that should be left to science. Which science? The customary practice to use the terms 'naturalism' and 'physicalism' interchangeably in current philosophy has blurred some of the differences between the two. In a nutshell, if it is correct to claim that physicalists are naturalists in spirit, the opposite is not necessarily true. Biological naturalism, in fact, is physicalist at bottom but acknowledges the explanatory status of all special sciences to a degree that hardcore physicalists would find it difficult to accept. Or, in other terms, there is more to naturalism than the idea that everything is just particles in fields of force. Searle's naturalism would be better thought as suggesting a pluralist attitude towards naturalism – as a 'global approach' so to say (De Caro and Macarthur, 2004) - based on the relevance of all scientific theories and methods. On this reading the fundamental metaphysical inquiry into what there is in reality is essentially an empirical and cross-disciplinary question to be addressed on multiple explanatory levels.

3.5.2 Ontological Reduction without Epistemological Reduction

The causal efficacy of the mind and its instantiation at the physical-biological level fully exhausts Searle's idea of scientific reduction[43]. All is necessary for a naturalistic account of intentionality is that intentional states, whether singular or plural, are (caused by) neuro-physiological processes and, in turn, have effects in the world. And the relevant explanation can be carried out by redefining the expression that denotes the reduced phenomenon in terms of its causes.

[43] Searle (1992: 112-116) lists a number of meanings for the concept of reduction, and discusses each of them with regard to the problem of consciousness.

To provide a naturalistic explanation of collective intentionality is not a straightforward and easy process, though. In principle, we don't expect a certain discipline, no matter whether in the natural or social sciences, to tell us whether a feature is natural simply by logical stipulation. Rather, conceptual analysis is integral to the empirical investigation of the world, so there is not a clear-cut line between these two methods of investigation. For these very reasons, however, it can be argued that Searle's belief in the causal efficacy of collective intentionality is rather motivated by the thesis that collective states can't be reduced to their individual parts, which is proved on purely logical grounds. If this were the case, Searle's philosophical project may be seen as an expression of the Canberra Plan. But, as I said, it is difficult to draw a line between logical and empirical questions in the overall inquiry. Searle is well-aware of the problem, and responds to those who think he is giving methodological priority to conceptual analysis as follows:

> There is now no sharp distinction between philosophy and other disciplines. In my intellectual childhood it was regarded as essential to understand that philosophy consisted in conceptual analysis and that this is quite different from any sort of empirical investigation. Now, many philosophers, and I am one, think it is not always possible to make a sharp distinction between conceptual and empirical issues, and indeed in my own work I rely heavily on all sorts of empirical results (Searle, 2007: 30-1).

Let me clarify how this passage countenances common criticisms to Searle. This challenge can be tackled by showing that 'reduction' is typically assigned distinct meanings depending on whether one is addressing logical/philosophical or biological/empirical questions. Let us start with logical analysis. According to Tim Crane, the logical structure of the mind is the set of features that must be in place for one person to *seem* to have them. These features count "as the appearance of mind, how minds seem to those who have them" (Crane, 2001: 8). Logical properties, in other terms, are those which satisfy the ordinary concepts of everyday thought and language. Searle's emphasis on the logical analysis of the mind must then be read as serving the function of describing how intentional mental states seem to us rather than how they *really* are. In fact, if it is accepted

that an intentional predicate, or kind, is part of the basic 'fabric' of the world, it also exists independently of our attitudes *qua* observers. So, the fact that mental states have a certain intentional structure by themselves is again (hypothetically, fallibly, of course) a matter of fact, independent from the meaning of the corresponding ordinary terms. No existential commitment follows from merely examining the 'appearance' of minds, because it is the job of science to unearth ontological truths by means of empirical investigation.

The logical meaning of reduction makes justice to the deep ambiguity that arises from saying that identity claims leave 'nothing over and above' the reduced phenomena (Smart, 1959). The ambiguity is that "one *thing* cannot literally be *reduced* to another thing: either the one thing *is* the 'other' thing, or it is not" (Crane, 2001: 54; emphasis not mine). That is, the identity relation between the reduced phenomenon (A) and the reducing one (B) is an ontological claim: it tells us that A and B are one and same thing. But the identity does not exhaust the sense in which we seem to know that A still 'exists' after it is successfully reduced to B. How are we to make sense of this 'mode of existence' if the identity tells us that nothing in A is not in B?

What we know when A is reduced to B is that A can now be described in terms of B *also*, precisely because the two are ontologically the same. But this is not a new fact in ontological terms but, rather, a consequence of redefining the entities so that the reduction follows from the definition. To make one thing more intelligible by showing that it actually is another thing, is in fact an explanatory (epistemic) quality of reduction. In this sense, what seems to 'survive' the reduction is the set of surface features, namely *appearance*, of the phenomenon, whether objective or subjective, which used to define it before the reductive definition is carried out. A scientific reduction, then, does not simply boil down to its identity (ontological) component but it also involves explanatory features. Or, to put it differently, if identity does not exhaust reduction, there is more to reductionism than ontological reduction alone.

One way to distinguish between the ontological component and the epistemological one is to use the notion of reduction *per se* to mean the former, and that of 'reductive explanation' to mean the latter (as proposed by Kim, 2006). As it is clear from this formulation, Searle's notion of causal

reduction may be associated with reductive physicalism as a case of reduction without reductive explanation. In other words, a feature may *seem* to be irreducible in light of the analysis of ordinary concepts, while in fact it is causally reducible on another. The central point is that the irreducibility claim stands on the level of reductive explanation, and therefore must be kept separate from the claim of reduction in ontological terms. Hence, there is no connection between the argument that collective intentionality is irreducible to the summation of individual states plus mutual beliefs and the claim that it is a biologically primitive fact of the matter. The former is the subject of philosophical scrutiny, the latter falls into the domain of scientific inquiry.

3.6 Concluding Remarks

At the heart of Searle's philosophy of mind and society lies the view that intentionality is an intrinsic property of the mind. In this chapter I have discussed this idea in the context of the debate about the naturalness of collective intentionality. Searle's realist argument moves from the claim that there is something peculiar to collective intentional phenomena, which cannot be reduced to other kinds of intentionality. The irreducibility in question is not an attribute of the content of collective intentional states, or of the subjects which they are assigned, but of the psychological mode by which people represents aspects of the world from an intrinsically 'we-perspective'. As a primitive form of mental life, collective intentionality is another kind of intentionality that lies in the heads of individuals and cannot be decomposed into more elementary units.

Individualism with regard to the ontology of society is not the only condition that Searle wants his analysis to satisfy, however. There is another condition which has gained Searle a number of criticisms inside and outside the research in collective intentionality: internalism. Internalism is the thesis that genuine intentional mental states are structurally independent of how the world is like: for individuals to refer to something in the world it is not necessary that their brains be in any relation with anything external. For most collective intentionality theorists, the idea that individuals can represent things in the world as a collective, whereas the collective must not be existent, is simply unintelligible. As I have carefully explained, this line of attack underlies a misleading

construal of the internalist lesson. Internalism is about the conditions of possibility of collective intentionality across possible worlds or counterfactual situations, rather than the truth-conditions that must be satisfied for people to have genuine collective intentional states in the head.

The most problematic aspect of Searle's theory concerns his reference to collective intentionality as a biologically primitive form of mental life. On the one side, the theory of the Background helped us situate the meaning of 'biological' in Searle's overall philosophical project: collective intentional states are underpinned by background capacities which result from processes of biological evolution rather than cultural, or linguistic, acquisition. On the other side, I have analyzed Searle's naturalism in order to clarify the motivation for the claim of biological primitiveness. In this respect, for Searle any acceptable program of naturalization of the mind ought to rely on intentional causation as its working hypothesis. Hence, his realist attitude towards collective intentionality is inspired by a belief in the causal continuity between the biological reality of the mind and the power of mental properties to yield physical effects.

The claim that collective intentionality is a natural (biological) feature of the mind, however, does not follow from the proof that it cannot be reduced to or eliminated in favour of something else. Indeed the two claims admit of different kinds of evidence in support or against them: the former is subject to scientific scrutiny whereas the latter results from the conceptual analysis of collective intentionality. At first glance, though, Searle seems to grant a central explanatory role to conceptual evidence in drawing naturalistic claims, as if the ontology of collective intentionality were postulated on the basis of linguistic analysis alone. Although some wording may suggest the opposite, I have concluded that, for Searle, existential conclusions about collective intentionality are metaphysical claims that derive from the belief in the completeness of physics, not from the results of linguistic analysis and intuition. If there is something controversial to his theory is, rather, the fact that those conclusions are premature in light of the state of art of the research in the cognitive (neurological) bases of collective intentionality.

Four

The Construction of Collective Intentionality

Social constructivists like Raimo Tuomela hold that fundamental aspects of human life, including meaning and intentionality, are contingent upon communal social-cultural habits. The view that collective intentional states are socially constructed is however consistent with a naturalistic construal of intentionality based on the philosophy of mind and language of Wilfrid Sellars. In this chapter I shall defend this view by intepreting Tuomela's theory of intentionality as a mild version of constructivism, in contrast with the radical view that collective intentionality and agency are intimately theory-dependent. Strong constructivism, I conclude, is incompatible with an empirically-based, natural-scientific approach to the ontology of collective intentionality.

4.1 Introduction

Social constructivism is a family of views in the philosophy of social science largely based on the remarks of the later Wittgenstein (1953/2001) on the social nature of meaning. The idea of an entity being socially constructed is that it depends for its constitution upon processes of socialization and enculturation. Although almost all accounts of social ontology subscribe to a very general conception of 'construction', according to which social facts are constituted and maintained through collective acceptance, Raimo Tuomela stands out among the collective intentionality theorists as the proponent of a full-blown constructivist response to the question of the existence and identity of collective intentional states.

The starting point is the view that the natural tendency of individuals to think and act *qua* members of a group derive from the social-cultural practices of the group itself. The intentionality of collective behaviour is not an intrinsic property of brains but results from what the members of a certain community take it to be. The confusion surrounding the exact meaning of this claim has led to the formulation of several variants of social constructivism. However, the source of such

confusion is likely to be found in the original remarks by Wittgenstein concerning the 'rule-following' problem – one of the most debated problems in contemporary analytic philosophy.

One way to elucidate the constructivism of Tuomela in relation to the rule-following problem is to bring in the discussion the philosophy of mind and language of Wilfrid Sellars, who has inspired Tuomela's work in profound respects. The aim of this chapter is not to assess social constructivism on its own merit, however. Instead, I shall argue that the social-constructivist theory of Tuomela is consistent with the conditions for a naturalistic account of collective intentionality. To this scope, I will defend one interpretation of the rule-following problem based on Sellars' 'verbal behaviourism', the view of the epistemological priority of language over thought.

The core insight is that we need a system of representation, notably a language, to make the contents of inner mental states intelligible to us and the others. In Sellars' vocabulary, language comes prior in 'the order of knowing', that is, in the order of how we come to *understand* what is it that our minds refer to in the world, rather than what constitutes the very capacity of entertaining thoughts and meanings in ontological terms. I shall therefore illustrate Tuomela's commitment to naturalism by defending an epistemological construal of the rule-following problem, which I shall refer to as a 'mild' version of constructivism, in opposition with the more radical constructivist approach to the ontology of collective intentionality propounded by philosophers like Antti Saaristo (2008).

The structure of the chapter goes as follows. In section 4.2 I shall present the social-constructivist basis of Tuomela's definition of we-intentions as one version of the language-over-thought stance in the philosophy of mind. In section 4.3 I shall make appeal to some of the most famous reflections of Sellars' on verbal behaviourism to discuss some aspects of the rule-following problem. I shall then provide in section 4.4 an assessment of Tuomela's theory and its compatibility with naturalism by comparing it with a stronger constructivist treatment of the ontology of collective intentionality. Despite his commitment to social constructivism, Tuomela's research program is a major pillar of the family of naturalistic views of collective intentionality, as I shall conclude in section 4.5.

4.2 The Collective Acceptance View

Tuomela is one of the towering figures of collective intentionality theory. As we saw in chapter 2, his analysis of collective intentional behaviour is the first systematization of Sellars' (1963) scattered remarks on the relation between we-intentions and norms (Tuomela, 1984). The distinguishing feature of we-intentions is that they underlie a peculiar way of reasoning at the individual level – thinking as a collective - which *entails* that there is a collective to which intentions refer to. In the 'Collective Acceptance Model', Tuomela's account of social ontology, the entailment relation consists in the fact that "people are social in the sense that they involve and tend to take into account in their thinking and acting what others think and do" (Tuomela, 2002: 10). As we have seen in §2.2.1, the remaining task is to clarify the origin of the *belief of sociality* (*i.e.* that there are other agents who also intend to act jointly) that is supposed to enter the conditions for collective intentions.

Firstly, we-intentions must be distinguished from joint intentions: the former represent the agent's willingness to do something together and can be called 'aim-intentions', whereas the latter are 'action-intentions' tied to the direct performance of a plan of action (Tuomela and Miller, 1988). Intuitively, the agent's intention to aim at a collective goal seems to be distinct from the specific belief that the other agents will also intend to do the same. So, one may be tempted to identify the belief in question with the joint intention (Tuomela and Miller, 1988: 330), except for the fact that the belief that the others will also participate in the action seems to be prior to the very plan of action (joint intention).

This confusion surrounding the characterization of we-intentions is the target of Searle's critique (1990). As you recall from the discussion of the irreducibility thesis in §2.2.3, the Tuomela-Searle controversy has contributed to establish the irreducibility thesis at the core of the collective intentionality literature. Since we-intentions presuppose the belief that there already are other agents with the same kind of intention, Tuomela's account leaves unexplained the origin of the very belief that is supposed to bring about we-intentions. This leads to the conclusion that the analysis is

circular because it resorts to the very concept in need of explication. Notice, however, that Tuomela has continued to advocate his account over the years (1995; 2002; 2007), acknowledging that it is circular though not in a vicious way (2005). So, since both Tuomela and Searle offer realist arguments for the idea that collective intentional behaviour relies on a distinctively collective mode of thinking and acting, their realism must be motivated on grounds that have only partially to do with the irreducibility thesis. In other words, Tuomela does not seem to be concerned with giving a non-circular account of collective intention as much as Searle.

At this point, one may wonder whether the belief of sociality that is so important in Tuomela's definition could not be ruled out so as to save the account from circularity. Yet, Tuomela's relentless defence of his original theory invites to ask why this additional element is necessary to construe the notion of collective intentional behaviour, if not for considerations related to the irreducibility thesis. Notice that the question is still what is to be conceptually primitive in the theory of collective intentionality, but Tuomela and Searle take different routes in response. So in order to highlight the difference in their approaches we need to call into question the very meaning of intentionality.

Recall that, in elucidating Searle's theory of intentionality in §3.2, we pointed out that philosophers tend to divide in two schools of thought as for the conception of intentionality. Searle (1983) is perhaps the most famous defender among contemporary philosophers of the view that the *locus classicus* of intentionality is thought. According to this idea, the capacity of any symbol – be it a word or a thought – to refer to something beyond itself resides first and foremost in the mind. Speech, or language, is *overt* thinking in the sense that the contents of thought come prior to their linguistic expressions (marks and sounds). On Searle's naturalistic approach to the mind, it is a fact of the matter, namely a biological fact, that brain states are intrinsically intentional and meaningful: they refer to something in the world independently of any representational system that the subject may use to understand and communicate to others what it is that they are about. Searle's (1983) way to render this idea is by saying that questions about the philosophy of language should ultimately be addressed by questions about the metaphysics of the mind.

The thought-over-language picture has dominated the philosophical scene until a *linguistic* theory of thought has gained an increasing number of advocates (Brandom, 1994). On this view, language is not just a tool for expressing thoughts by making them intelligible to the subject and to others. The importance of language for mindedness is that it makes thoughts *thinkable* first and foremost. There could not be any thought 'in mind' other than via acts of linguistic conceptualization. Another way to formulate the point that language is the site of intentionality is the claim that intentionality, whether individual or collective, only exists relative to an appropriate description, or interpretation, of it. The subject needs a representational system to conceptualize what she is thinking about, so the intentionality of thought and action is not separate from the very concepts that she - *qua* bearer of thoughts and agent – uses to conceive of them. Hence, on the language-over-thought picture, the task of providing a naturalistic account of (collective) intentionality boils down to explaining the origin and nature of conceptual acts.

Which of these pictures suits Tuomela's conception of collective intentionality? Before I address this question at length in the next section, there is one more aspect of the debate on the *locus* of intentionality that deserves attention[44]. There are various ways to specify where thought stands relative to language depending on the *kind* of priority at stake (Davies, 1998). For our discussion, it is important to distinguish between the ontological and the epistemological sense of the priority question. To say that thought is *ontologically* prior to language is to say that there cannot be any language without thought, whereas there can be thought without language. To say that thought is *epistemologically* prior to language means that the route to knowledge about language goes via knowledge about thought, in the sense that we cannot know what linguistic symbols mean independently of the content of the correspondent thoughts (Davies, 1998: 227). As it is clear, depending on whether one believes that thought comes prior to language, rather than the other way, there will be various combinations according to the kind of priority invested on thought (language).

[44] For reasons of consistency, I follow the notation introduced to discuss Searle's theory in chapter 2: I will use the notion of intentional or mental content to mean the intentionality of thought, and the notion of linguistic meaning to refer to the intentionality of linguistic symbols.

Davies (1998: *ibid*.) also introduces a third kind of priority, *analytic* or philosophical priority, which serves the purpose to set apart philosophical approaches to the priority question. For example, Searle's intentionalist picture of the mind is a clear example of a philosophical theory that postulates the ontological priority of thought over language: linguistic meaning *derives* from the content of the thoughts that language is used to express. This claim must be read in a stronger sense than that suggested by the epistemological construal: the intentionality of mental states is foundational to linguistic meanings and sets the way forward to understand what it is that linguistic symbols refer to in the world. An opposite view is Michael Dummett's anti-mentalistic view that the account of linguistic meaning does not imply reference to the intentionality of thoughts (Dummett, 1973). What language means is formative of the contents of thought and, thus, sets language as prior in the order of philosophical elucidation.

The distinction between ontological and epistemological priority is the central means to clarify Tuomela's conception of intentionality. As the brief remarks about the controversy with Searle show, in fact, there is a more fundamental difference in their views than the one concerning the circularity issue. Before I articulate this difference, it is important to reckon that Tuomela does not offer a thorough definition of his take on the priority question. It will then be helpful to analyze his broader approach to meaning and thought through the lens of his major source of inspiration, namely the philosophy of Wilfrid Sellars. Not only was the notion of we-intentions originally introduced by Sellars in the nineteen-sixties (1963), but Tuomela's own thinking was influenced by Sellars' work in ways that are not confined to questions of social ontology.

4.3 Verbal Behaviourism

Sellars is the proponent of a thesis according to which the key to understand the nature of thinking is linguistic behaviour[45]. Language is not only a medium of communication by means of

[45] 'Behaviourism' has marked a particular phase in the studies of the mind during the central decades of the twentieth century. In brief, according to behaviourism mental states are nothing but instances of observable intentional behaviour. Behaviourism has declined afterwards under the fire of a wealth of criticisms

which thoughts are expressed: it is thinking itself. Tuomela endorses this account, known in philosophy as '*verbal behaviourism*', and characterizes it as the view that "language – or rather language use - is conceptually prior to thinking, even if thoughts may be argued to cause action" (2002: 40). Notice that this formulation already gives an important answer to the above question on the position of Tuomela in the priority debate. The point is to understand the sense of the primacy of *language* over thought. Let us look at some aspects of Sellars' philosophy in detail, which have important bearings on Tuomela's view of the conceptual priority of language.

Before starting, two premises are in need of elaboration. First, it is useful to distinguish between two ambiguous senses of 'thinking': thinking about something, say '*p'*, in the sense of entertaining intentional states such as beliefs and desires that are about *p*; and the very process of conceptualizing the content of those thoughts, say the concept *that-p*. Traditionally concepts are defined as mental representations that arise from thinking processes (see Margolis and Lawrence, 2007 for an updated discussion of the literature on concepts). Thus 'having the concept *that-p*' counts as having a given thought, for example the belief *that-p*, that represents the content of the belief about *p*. Secondly, another relevant difference that belongs to the priority debate is between *having* the concept *that-p* and *expressing* it. We can distinguish between those who hold that one can believe something without having to express it – typically the position of those who believe that thought must be prior to language on some level of characterization; and those according to whom conceptualization (and thinking in general) requires one to be able to express concepts linguistically, which is the view that language is a pre-requisite for thinking. The problem is to understand to what extent conceptual activities necessarily involve some form of expression, and under which suitable conception of the term 'expression' conceptual activity can be fully explained.

Sellars' project moves from the premise that there is an explanatory leap between having a given mental state, like believing *p*, and having the concept *that-p*. What does it mean to have the concept *that-p*? We have already argued that Sellars defends the language-over-thought stance. Nonetheless,

---

concerning both the theoretical and methodological assumptions on which the theory was based. For reasons of space, I will not develop this point further but see Crane (2003) for a critical discussion.

Sellars would never deny that one is capable of entertaining 'genuine' mental states that are not observable as overt speech. Whilst he acknowledges that there are inner episodes that are not linguistic in nature (Sellars, 1981/2007: 283), he argues that intentionality is not *intelligible* unless one explains how we come to have the concept of those very states. So, it is the process of conceptualization – how we come to know to have the belief *that-p* – that makes it clear what it is to have the relevant belief. As Sellars formulates the point: "My disagreement with the classical view takes its point of departure from the fact that I construe *concepts* pertaining to the intentionality of thoughts as derivative from *concepts* pertaining to meaningful speech" (*ibid*.; emphasis in original).

This passage suggests that the priority of language over thought is an *epistemological* matter only. Although verbal behaviourism does not imply that there are no inner episodes of mental life, language is certainly primary in what Sellars calls 'the order of knowing', whereas thoughts remain distinct from their linguistic expressions in 'the order of being'. The conceptualization of intentional states is necessary to ground knowledge of the contents of thought. Hence verbal behaviourism relies on the claim that linguistic behaviour is actually the *bearer* of conceptual activities in the sense that it is "*already thinking in its own right*" (1969/2007: 80; emphasis in original). The language-to-though relation is structured by the very act of verbalizing (uttering *p*) the mental event: in order for a person to have the concept that-*p*, one has to be able to verbalize (express) it. As Sellars formulates the point

> We must resolutely put aside the temptation to draw the kind of distinction between *thought* and its *expression* which this formulation implies, and continue with the intriguing idea that an uttering of 'p' which is a primary expression of a belief that-p is not merely an *expression* of a thinking that-p, but is itself a *thinking*, *i.e.*, a thinking-out-loud that-p (1969/2007: 70: emphasis in original).

The notion of 'thinking-out-loud that-p' is Sellars' response to the standard critique to verbal behaviourism. Behaviourism leaves us, in fact, with the paradoxical conclusion that, lacking linguistic activity, one is not allowed to postulate mental states. But, for Sellars, the gap between

verbal acts (utterances *that-p*) and conceptual processes (thinking *p* or thinking *that-p*) is bridged precisely by "candid, spontaneous overt verbal behaviour" (1981/2007: 284). 'Thinking-out-loud' is a form of meaningful speech which does not require a context of communication or the presence of a hearer to be performed. In a way, it is an intentional action in the basic sense that the instantiation of thinking is identical with its verbalization. The analogy with actions turns out to be ambiguous in two ways, however. On the one hand, thinking-out-loud might be taken to represent a particular class of intentional actions known in philosophy of language as speech acts. Yet, although it is true that Sellars' theory emphasizes the performative character of verbalization, it actually lacks the typical structure of a speech act (see Searle, 1969 for an overview). On the other hand, one may be tempted to explain thinking processes in instrumental terms. But this is clearly problematic because there are kinds of mental states – for instance "perceptual takings, inferences, and volitions" (1974/2007: 84) - that are not performed intentionally by the agent.

In sum, verbal behaviourism is the view that verbalization is the key to the understanding of the contents of thought. That is, a person comes to know what intentional states stand for, *i.e.* to have the relevant concepts, via linguistic expression. This also means there is a remarkable symmetry between questions about the constitution and possession of concepts and the structure of language. Thus, the project of giving "a naturalistic interpretation of the intentionality of conceptual acts" (Sellars, 1969/2007: 57) will be tied to the question of what underpins the functioning and understanding of language. This is a critical aspect in the language-over-thought literature, particularly when it comes to the understanding of meaning.

4.3.1 Rule-Following

The most remarkable implication of verbal behaviourism is that, if language is a rule-governed form of behaviour, the very same rules should govern the process whereby one comes to conceptualize the contents of thought. What is it to arrive at the concept *that-p* by following the rules of language? Does understanding of a linguistic symbol imply grasp of a mental representation? Where are rules – in the mind? These questions shed light on one of the most

famous issues of modern philosophy raised by Wittgenstein: the problem of *rule-following*. In various passages of the *Philosophical Investigations* (1953/2001) and *Remarks on the Foundations of Mathematics* (1983), Wittgenstein raises a number of issues concerning the nature of rules and, more generally, the understanding of meaning and language.

The starting point is passage §185 in the *Investigations* where Wittgenstein invites us to reflect on the use of the linguistic symbol '+' to mean the sum of any pair of numbers. Although anyone familiar with '+' knows that such sign stands for the addition function, it is not immediately clear what fixes understanding of '+'. This question has given rise to a wealth of interpretation from the early 1980s, also known as 'Rule-Following Considerations' (Wright, 2007), which were boosted in particular by Kripke's *Wittgenstein on Rules and Private Language* (1982). Kripke emphasizes the sceptical nature of the paradox envisioned by Wittgenstein: if, on the one side, there is conceptual room for interpreting how to use a symbol in virtually infinite ways, on the other side, any action performed in accordance with any interpretation of the rule can be said to be correct or incorrect depending on circumstances. These two problems – the 'infinity' and the 'normativity' problem (Saaristo, 2008) – lead to the paradoxical conclusion that there is *no* fact of the matter that settles the understanding of the symbol and the correct use of it. Since the literature on rule-following cannot be summarized exhaustively in a few lines, I will focus on Sellars' considerations about the problem which set the stage for Tuomela's constructivism.

For Sellars, a simple way to think of a rule is in terms of an 'ought' statement of the form: "If one is in C, one ought to do A" (1969/2007: 58). Yet, since rules are broadly speaking linguistic constructs, in order to follow them one must be familiar with the linguistic system in which they are formulated; and this exposes us to the infinite regress of rule-following. Moreover, it does not suffice to say that the interpretative regress can be avoided by saying that the relevant rules are actually expressed in a meta-language allowing the learner to get acquainted with language – unless one knows the rules of the meta-language. Provided that rule-governed linguistic behaviour is the clue to the nature of conceptuality, *i.e.* thought, then the statement "If one is in C, one ought to do A" also implies that in order to act according to the rule one must already have the concepts of A

and C, which only comes with knowledge of language. But where does such knowledge come from? Does it imply grasp of a representation?

Intuitively, it does not seem plausible to say that rule-following presupposes that the rules of linguistic (and therefore conceptual) games have to be present to one's mind all the time. People often seem unable to give a precise answer when they are asked which rule(s) they have been following in speaking the language they do. So, it is not that conforming to rules requires that they be mentally represented in order to be grasped, apart perhaps from the very first linguistic game in which a person might have 'interiorized' the rule at stake. Faced with this difficulty, Sellars proposes two notions that help shed light on the mechanisms underlying rule-following: 'pattern-governed behaviour' and 'rules of criticism'.

The notion of *pattern-governed behaviour* captures the idea that intentional behaviour is often governed by patterns and routine and is performed without the agent being consciously aware of the underlying rule. Sellars contrasts it with the notion of rule-obeying behaviour, where the agent engages in action by representing the relevant rule of conduct; rule-obeying behaviour contains "both a game and a meta-game" (Sellars, 1967/2007: 34). Pattern-governed behaviour is rather performed unintentionally and implies no overt mental representation. In Sellars' own definition, it "exhibits a pattern, not because it is brought about by the intention that it exhibits this pattern, but because the propensity to emit behaviour of the pattern has been selectively reinforced, and the propensity to emit behaviour which does not conform to this pattern selectively extinguished" (1974/2007: 86-7).

The key insight of pattern-governed behaviour is thus that it is carried out on the basis of routine. This may create some confusion with the definition of intentional behaviour. Notice that a pattern-governed action is typically an instance of purposive behaviour in the sense of being directed at something in the world. But, for Sellars, this is a *non-action* precisely because it is carried out on a routine basis rather than in the pursuit of an overt goal. In this respect, pattern-governed behaviour differs from instances of *non*-intentional behaviour, which are characterized by the presence of reasons not to act in a certain way rather than lack of purpose, as well as from actions which are

*un*intentionally performed by mistake. For example, think about a person that is visually impaired and needs to wear glasses. The act of taking her glasses is the first thing she does every morning. For Sellars this is an example of pattern-governed behaviour in the sense of being "meaningful (functionally meaningful and meaningful in the aboutness sense of intentionality) but *necessarily* non-intentional (in the conduct sense) activities by single individuals" (2002: 46; emphasis in original).

One can think of regularities in behaviour as actions performed routinely in virtue of some background capacities. It should not surprise us that Sellars picks up the concept of 'background', which plays a relevant role in Searle's work, from Wittgenstein's reflection on "knowing how to go on" (1974/2007: 88). But, as we know from the discussion of Searle's theory of the Background in chapter 2, to stipulate a background of primitive capacities turns out to be a reasonable answer to questions of meaning-understanding if and only if the theory is coupled with an account that makes sense of the first act of understanding the meaning in question. But what is implied in understanding how to follow a rule first and foremost? Sellars adds to the debate the important distinction between 'rules of action' (ought-to-do's) and 'rules of criticism' (ought-to-be's).

Back to the initial formulation of a rule, *rules of action* require an agent to know what has to be done in context C in order to do A. *Rules of criticism*, instead, do not call for any prior knowledge of the situation in order for the action to occur: for someone to act in such a way as to have the concept *that-p* is to obey ought-to-be norms by means of pattern-governed forms of behaviour. Sellars discards, then, rules for action because they require the agent to have some prior conceptual framework, possession of which would reiterate the infinity-problem of rule-following. Conceptuality, and human thinking more in general, is underlain by behaviour conforming to rules of criticism, namely "the pattern-governed activities of perception, inference and volition, themselves essentially non-actions, which underlie and make possible the domain of actions, linguistic and non-linguistic" (Sellars, 1974/2007: 88).

If all one needs to know the contents of her thought is to routinely follow rules of criticism, how does one learn how to follow these rules? Recall that Kripke reads Wittgenstein's considerations as

suggesting that there is no fact of the matter that fixes the understanding and the correct use of any representational symbol. Kripke advances, then, a solution which has gained much support in philosophy and the social sciences[46]. The basic idea is that meaning is no longer a matter of how a person grasps a symbol, but of how her use of the symbol accords with the use of those who are already acquainted with it. Similarly, a person is said to follow a rule in the right manner insofar as she engages correctly in linguistic exchanges with others. Hence, the precondition for one to grasp a rule and use it correctly is to be part of a group of rule-followers, in the sense of being exposed to the social and cultural practices of the community of membership.

The community is seen as the place where linguistic, therefore conceptual, rules are shaped and transferred through its members. Those who participate in social practices are "*first* language *learners* and only potentially 'people', but *subsequently* language *teachers*, possessed of the rich conceptual framework this implies" (Sellars, 1969/2007: 63-4; emphasis in original). In this respect, Sellars' verbal behaviourism builds on the view that social practices are not just 'facts' in the social realm but 'forms of life' where the rules of the game are constantly taught and learnt to become routine. Social and cultural factors are the means through which human activities become meaningful, by allowing for the conceptualization of the contents of thought.

4.3.2 Social Constructivism

To say of an entity that it depends for its constitution upon processes of socialization is to claim that it is *socially constructed*. There are several meanings associated with the notions of 'social construction', 'constructivism' and 'constructionism', however, depending on how pervasive the 'construction' is and on how much one ought to commit oneself to it. At least since Peter Winch's 1958 *The Idea of a Social Science*, social constructivism has enjoyed a good deal of support in the philosophy of social science, and has also raised important questions concerning its relation with naturalism (see Mallon, 2008 for an introduction). In the remaining of this section I will propose a social-constructivist interpretation of Tuomela's realist argument of collective intentionality based

[46] But for a different interpretation of Wittgenstein's paradox see McGinn (1984).

on Sellars' analysis of the foundational role of social-cultural practices with regard to the nature of conceptuality. My main concern will be to show that Tuomela's commitment to social constructivism is compatible with the tenets of methodological naturalism.

Let us start, again, from the idea that agents share we-attitudes insofar as they realize that all intend and enact things in the same way (we-mode). How does each person form *this* concept, which is the prerequisite of sharing intentional states? Echoing Sellars, we can say that one has a certain belief *that-p*, where *p* is the fact that the others will cooperate and do their part in accomplishing a shared activity, as the result of exposition to the practices of the community. Since these practices constitute the building blocks of one's "full-blown conceptuality, *viz.* conceptual thinking and acting" (Tuomela, 2002: 7), the answer is that the capacity for sharing intentional attitudes will also be a matter of accordance with the rules of the community. So, what makes we-mode actions instances of collective intentional behaviour is participation in the processes of socialization and enculturation whereby the relevant rules are taught and learnt. It follows that, when Tuomela claims that collective intentionality presupposes the belief that all participants will act cooperatively, what is conceptually presupposed is that they be members of the same community of rule-followers. If each agent did not believe that this is the case, intentional states could not be shared, in the sense that they would not be *intelligible* from a we-perspective.

Must this concept be overtly represented? Tuomela argues that the concept of the others intending and enacting things socially, which is presupposed by collective intentional behaviour, "shows up as conformity (although it need not be based on the agents' conscious motive to conform)" (2002: 92). In fact, since rules are taught and learnt within the linguistic community through social interaction, this conceptual 'infrastructure' forms just a background of presuppositions that allow agents to perform actions on a recurrent and unintentional basis. In other words, they are "conceptually in-built" (Tuomela, 2002: 79) in social practices and do not appear in the deliberative process through which agents cooperate.

To exemplify this idea, Tuomela builds on Sellars' notion of pattern-governed behaviour and proposes the notion of *collective* pattern-governed activities. These are forms of behaviour defined

in the same way as in the single-agent case, namely as instances of *many-person* meaningful actions performed unintentionally. Rule-following circularity is therefore avoided in virtue of the fact that no representation is required for the agents to act according to the underlying rules. "Psychological circularity is blocked because *pgb* [pattern-governed behaviour] can stand on its own feet, so to speak, from a psychological and ontological point of view, *viz.*, from the point of view of what actually is going on in the agent's mind and action" (Tuomela, 2002: 50).

Can collective intentional behaviour be explained entirely in terms of collective pattern-governed behaviour? The idea seems odd in one fundamental respect. Social behaviour is No doubt largely determined by routine and repetition, and agents do not have to represent the rules of the game if they want to participate in it. But, intuitively, one can think of a time when each agent has represented what she was doing for the first time at least. What happened on that occasion? A plausible answer is that the initial representation occurs when one is a child, so by the time she will be an adult the rule will no longer be present to the mind and will have evolved in a routine. This is coherent also with the idea that the members of a community are both language learners and, later, language teachers. If this is true, we should focus on explaining what happens when rules are internalized by the members of the group in such a way that they are able to act routinely.

Tuomela formulates the point in the following terms: "Who teaches the teachers? There may be, but *need not* be, a first teacher" (2002: 128, emphasis mine). So, what should we look for in order to understand the origin of rules? We actually don't look for the person who has known first how to act in a certain way so as to instruct the others afterwards. What makes a certain activity learnable and teachable is the fact that it is collectively accepted by the community as a whole. "Thus, even if a social practice can be initiated by a single individual, it needs to be collectively accepted" (*ibid.*). In more general terms, what matters is not what is in the mind of the 'first teacher' (how one represents the rules of the game at the very beginning). Collective behaviour in the conformity sense consists, rather, in one's commitment to the belief that what is collectively accepted holds for *all* members of the community.

In sum, Tuomela's social constructivism consists in the idea that it would not be possible for one to intend and act as the member of a group independently from the texture of social conventions and rules of the linguistic community. Collective behaviour is *naturally* meaningful, *i.e.* intentional, to us because we are members of a community of rule-followers and we share the same rules of conduct. The contrast with the naturalism of Searle, among others, could not be more evident: for Tuomela the key to thinking about and enacting things in we-mode is the set of communal practices and 'forms of life' that make us part of the same group, rather than some intrinsically meaningful brain state. The *locus* of meaning resides, then, in the processes by which we construe and share social reality.

4.4 Naturalistic Constructivism

The claim that intentionality is what the members of a certain community take it to be by collective acceptance is open to several interpretations. The source of this ambiguity can be found in the Kripkean reading of Wittgenstein which informed most of the rule-following considerations. Like I said, Kripke reacted to the rule-following paradox by proposing the view that grasp of meaning comes down to one's use of that symbol in accordance with the use of the other rule-followers. Hence, it is because meanings (rules) are established within a linguistic community that people understand and act upon them in the appropriate way. This is the gist of the so-called 'community view' of rule-following – the idea that there cannot be meanings and concepts outside of the community of membership. For example, imagine a child kept in isolation from other humans who somehow manages to grow up in a desert island. On the community view, because of the lack of interaction the child would be unable to learn how to play language games and to participate in forms of life. In short, the child would not have any chance to entertain thoughts and concepts. Yet, is thinking really just a matter of social construction?

To answer this question, let us analyze very carefully the claim that meanings are social constructions. What does Tuomela mean by 'meaning' in this context? Sellars' semantics prove again a valuable source of insight for the matter. Recall from the discussion of verbal behaviourism

that Sellars does not deny that inner states can be entertained without having to be expressed in overt speech. The message of the language-over-thought priority in Sellars is that these states can only be made intelligible by means of a language, *i.e.* a system of representation. They are 'thinkable', in other words, but not yet intelligible until one conceptualizes them via linguistic expression. So there is an epistemological sense to Sellars' thesis that must be distinguished from the ontological reading of the priority of language. According to the former, one cannot get to know the content of her mental states, *i.e.* what is it that they refer to, other than by linguistic access. The fact that a mental state of mine means 'this-and-that' to me is likely to be dependent on what I have been taught by others; and here is where the practices of the community become crucial. But the *content* of that state, and not the state itself, is what results from interacting with others.

In other words, it is not the *capacity* of thinking – entertaining mental states in general – that is socially constructed, with the effect that a child raised on a desert island could never mean things and have conceptual thoughts because of her isolation from the practices of the community. What is socially constructed is the way (content) in which each of us has access to them conditional upon life experiences and cultural influences. It should be clear by now that there is a profound difference between the following questions: whether there is such a fact as the capacity to entertain intentional states; and how these states can be understood. Of course one might agree upon the view that social linguistic conventions as well as processes of enculturation give people the resources to fix on shared meanings for the sake of communication and cooperation. But these resources, not the fundamental skills that underpin understanding itself, are the outcome of public practices. Hence, the appropriate way to read Wittgenstein on rule-following is by saying, in line with some commentators (McGinn, 2002), that he points out the contrast between the inner and the outer – the impossibility of a 'private language' – instead of the individual against the community. I shall call this interpretation of social constructivism 'mild' and distinguish it from the radical claims of the community view, which inspires a strong version of social constructivism.

Therefore Tuomela is a 'mild' constructivist when it comes to the question of the ontology of collective intentionality. His theory is characterized by the view that whether the contents of

intentional states are accessed in I- or we-mode is a matter of what the members of a community take them to be by collective acceptance. But this is not to say that the irreducible we-perspective of thought and action cannot be subjected to scientific scrutiny[47]. In fact, Tuomela postulates a mind-independent reality of physical facts including intentional states, and sees his own theory as compatible with scientific realism (Tuomela, 2002: 7-8). And this is what makes his social-constructivist account an instance of naturalistic accounts of collective intentionality after all - the commitment to the view that that fundamental issues of human social behaviour should be investigated with the same tools and methods as those employed by the empirical sciences. The naturalness of collective intentionality, in sum, remains a problem of empirical social ontology.

This conclusion is of great relevance in showing that one can embrace some tenets of social constructivism while remaining a committed naturalist. In fact, it is very common in the social-constructivist literature to find arguments against the very idea that science should be treated as a successful – if not the 'ultimate' - source of knowledge about the world. The consensus has it that, if there is a core idea to 'construction', it is that research should aim at showing that socially-constructed entities are under human (social, cultural) control, rather than the control of natural factors (Mallon, 2008). Some advocates of this idea, as well as of the most radical claims of the community view, argue that strong forms of social constructivism are however compatible with a naturalistic-materialistic approach to the ontology of intentionality. Anti Saaristo (2006; 2007; 2008), in particular, argues that collective intentionality theory offers the resources to bridge the gap between social constructivism and naturalism. Yet, it is by comparing Saaristo's theory of collective intentionality to Tuomela's that it will become clear how different meanings of social construction result in distinct approaches to the subject. On this comparison, Saaristo's strong constructivism appears inconsistent with the view of naturalism proposed in this thesis.

Saaristo's 'naturalistic constructivism' moves from Tuomela's commitment to the view that the contents of intentional attitudes are social constructions. In line with our interpretation, he argues

[47] For explicit references to the scientific literature on the underpinnings of collective intentionality see Tuomela, 2007.

that Tuomela "ends up supporting something like the social solution [Kripke's solution] (...) only as a contingent claim in the sense that a social element is *not* in his view a conceptually *necessary* element of all meaningful activities" (2008: 170; emphasis mine). This is to say that the 'social element' is not a precondition for the intentionality of thought and agency, although it contributes to fix the contents of intentional states. On Saaristo's view, this solution is unsatisfactory because it leaves the rule-following dilemma unsolved. Recall that the dilemma consists in the tension between two issues: the infinity issue, namely the problem of how to explain the nature of understanding in a way that avoids a regress of interpretation; and the normativity issue, which consists in clarifying what settles the correct understanding of meaning.

The novelty of Saaristo's theory is to present the core idea of collective intentionality theory, that there is an irreducible collective mode of thinking and acting, as a *naturalistic* version of the community-view solution to the problem of rule-following (Saaristo, 2008). In brief, Saaristo claims that the irreducible sociality of collective intentionality solves the normativity issue by allowing individuals, acting *qua* group-members, to 'derive'[48] their first-person singular attitudes from a collective-level plan. The 'priority' of we-mode considerations guarantees that individual applications are normatively guided in virtue of being embedded in communal practices where they are socially sanctioned (Saaristo, 2008: 172). So, the argument goes that the individual dispositions to act in certain ways, which are at bottom causal-biological 'blind' dispositions, only become meaningful once embedded in a totality of practice. That is, intentional attitudes are not intrinsically intentional – which would reiterate the interpretative regress – but become so when individuals derive them from the attitudes of the group. As Saaristo formulates the point:

> Meanings – and intentionality in general – reside, strictly speaking, in social practices instantiating intersubjective normativity, for only within practices (...) can a biological state count as meaningful. Thus, an individual can be seen as an agent capable of intentional actions, contentful mental states and meaningful talk only to the extent she participates in

[48] An expression Saaristo (2008) borrows from Searle (1995).

such practices. In short, the psychological is *constitutively* dependent on the social, as one could formulate this claim that expresses also the core thesis of methodological holism (2008: 172; emphasis mine).

As it is clear, this interpretation of the irreducibility of sociality is no longer confined to the epistemology of intentionality. The intrinsically social nature of meaning makes the question of the constitution of intentionality a philosophical, rather than empirical, problem[49]: conceptual constructions are constitutive of the very possibility of intentional behaviour (Saaristo, 2006). Since it is we *qua* agents that describe actions to be meaningful as part of the linguistic game of giving and asking for reasons, it follows that intentionality is postulated only under some suitable description of it. This is a typical instance of the community reading of Wittgenstein - the "constructivist view of seeing the rules of rationality that constitute the very possibility of actions as grounded in social practices of treating certain inferential steps as rationally acceptable, certain states as reasons and certain behaviours as actions" (Saaristo, 2006: 56).

What is highly controversial in the move from the epistemological to the ontological interpretation of the language-over-thought priority is the idea that such move is still naturalistic in spirit. On the one hand, in fact, it is fairly plausible to assume that human behaviour can be *explained* or 'rationalized' on the basis of the linguistic game of giving reasons for action. We need a language to have access to the nature of inner mental episodes, which is to say that language may come prior to thought in the 'order of knowing'. On the other hand, things might be different when it comes to analyzing the 'order of being', namely to find a naturalistic justification for the ontological conclusion that intentionality is thoroughly created by conceptual constructions.

According to naturalism, since we *don't* know which between language and thoughts has ontological primacy so far as science has been able to show, we are not allowed to draw existential

[49] This is the subject of 'interpretationism', the view that intentionality 'exists' under an appropriate description, or intepretation, of it. Tollefsen (2002a), in particular, makes appeal to the work of Dennett on the intentional stance to provide an interpretationist construal of group intentionality. On her view, "the study of the conditions or constraints on interpretation will, according to the interpretationist, yield *metaphysical* insights. These constraints are not merely methodological but *constitutive* of the mental" (2002a: 30; emphasis mine).

conclusions about the place of either in the metaphysics of the mind. But this is exactly the central claim of the community view: from the assumption that we explain intentional behaviour 'in the logical space of reasons', radical constructivists infer that intentional states *exist* relative to some conceptual framework of explanation. If this is so, then such claim is inconsistent with at least the general principle of naturalism that research on 'what there is' falls in the domain of science. Hence, either naturalistically-leaning strong constructivists depart from the traditional conception of 'naturalism', or philosophical naturalism and social constructivism diverge inevitably in accounting for the ontology of (collective) intentionality.

4.5 Concluding Remarks

In this chapter, I presented Raimo Tuomela's theory as a paradigmatic case of a social-constructivist theory of collective intentionality. My analysis moved from the charge of circularity against his view that collective intentional behaviour requires the agents to believe that the others will also engage in the joint action and 'do their part'. For Tuomela, to say that collective intentional behaviour is conceptually presupposed means that intentionality is what the members of the linguistic community take it to be by common acceptance. Instead of focusing on the charge of circularity, in this chapter I aimed at assessing whether this view meets the criteria set out in chapter 2 for the naturalization of collective intentionality.

In order to answer this question, I have pursued a twofold strategy. First, I individuated the rationale of Tuomela's broader approach to the metaphysics of the mind through the work of his major source of inspiration, Wilfrid Sellars. Sellars' verbal behaviourism gave us a key insight to reconstruct Tuomela's meaning of social construction. What is socially constructed is not the capacity of people to entertain 'genuine' mental states, but their contents, *i.e.* the way in which people access those very states. Language, in other words, comes prior to thought only on the epistemological level of explanation. Second, I used Sellars' considerations to distinguish Tuomela's stance from a more radical construal of social constructivism. The latter is based on a popular solution to the problem of rule-following – the problem of what fixes the understanding of

symbols and thoughts. The community-view, as this solution is often referred to, holds that meaning is no longer a matter of how one grasps a symbol, but of how the use of the symbol accords with the use of the others who are already acquainted with it. Similarly, the precondition for one to grasp a rule, and use it correctly, is to be part of a group of rule-followers in the sense of being exposed to the social and cultural practices of the community of membership.

I suggested that there is a significant difference between this view, which advocates a social-constructivist approach to the ontology of intentionality, and Tuomela's view of the epistemological priority of language over thought in epistemology. This difference explains why Tuomela's realist argument for the naturalness of collective intentionality is consistent with a broad construal of naturalism, whereas the argument that collective intentionality is intrinsically theory-dependent violates naturalism.

Five

Collective Intentionality Naturalized

The Shared Intentionality Hypothesis is proposed by psychologist Michael Tomasello to account for the uniqueness of human cognition. The gist of the hypothesis is that the complexity and variety of social-cultural phenomena depend on a species-specific cognitive and motivational 'infrastructure' for sharing mental states. In this chapter I present the Shared Intentionality Hypothesis as a naturalistic theory of collective intentionality based on comparative research in the roots of human development. By discussing the theoretical and methodological aspects of Tomasello's theory of sociality, I lay out the setting for evaluating his contribution to the naturalization of collective intentionality in the final chapter of the thesis.

5.1 Introduction

Throughout the previous chapters we have examined the philosophical approach to the naturalization of collective intentionality. Two linking themes have dominated realist approaches to collective intentionality. The first is the logical structure of collective intentional states - what it means for people to have intentional states shared with others; the second concerns the conditions of existence and identity of collective as distinct from individual mental states. The former reflects an action-theory approach to the structure of collective intentions *qua* intentions; the latter takes into account foundational issues concerning the ontology of the mind. Overall, the question whether there are good reasons to endorse realism about collective intentionality is a metaphysical question, one that social theorists and philosophers address with conceptual analysis.

In this and the following chapter of the thesis I shall turn to the natural-scientific approach to naturalization, and evaluate whether there actually is any viable scientific theory which succeeds to meet the criteria for a naturalistic account of collective intentionality. There are, in fact, relevant approaches outside of philosophy that are tangential with the collective intentionality literature, and that meet the criteria set out in §2.5. Namely, they treat collective intentionality as a problem of

empirical social ontology, invoking continuities with natural sciences to account for the phenomenon in a testable manner. For example, one strand of decision theory parallel to collective intentionality theory, the theory of 'team-reasoning', makes explicit appeal to the experimental tradition in social psychology that tests the ability of people to group-identify. Although attempts have been made to articulate the connections between the team-reasoning and the collective intentionality literature in a more systematic way, philosophers have scarcely engaged with the body of evidence on social identity and group-thinking.

For this reason, in this chapter I shall concentrate instead on a highly successful program of research which enjoys wide currency in contemporary cognitive science (Enfield and Levinson, 2006). By advocating of a twin-track (biology and culture) approach to human cognition, this program endorses some of the central tenets of collective intentionality theory like intentionalism in accounts of sociality and a drive for an integrated inquiry of social cognition. Shared intentionality features prominently in the research activity of one of its most active theorists, Michael Tomasello. A psychologist with interests that span from anthropology to philosophy, Tomasello is the proponent of a theory of sociality based on research in primate cognition, developmental social cognition and language acquisition. By drawing on the conceptual resources of collective intentionality to interpret the findings from a battery of ingenious experiments with infants and their nearest primate relatives, such as chimpanzees, Tomasello is thus the first scientist to engage constructively with the research paradigm laid out by the philosophers and social scientists' intuitions concerning group-thinking.

In this chapter I shall analyze the core element of Tomasello's theory of human cooperation and culture: the Shared Intentionality Hypothesis (Tomasello *et al*., 2005). This is the hypothesis that the social and cultural nature of humanity depends on the evolution of a set of pro-social inclinations and inferential skills for sharing mental states. There are two aspects of the hypothesis which are of great relevance for the project of naturalizing collective intentionality. The first is Tomasello's appeal to the conceptual resources of the collective intentionality literature to articulate the theory of the ontogeny and phylogeny of human social cognition. The second is his use of a

large battery of findings to help illuminate on empirical grounds issues of the debate about the naturalness of collective intentionality.

The chapter is structured in five sections. In §5.2 I shall examine the 'state of art' of analyses that bear resemblance with the collective intentionality theory outside of philosophy, and single out Michael Tomasello's program of research in psychology as the most advanced theory of collective intentionality naturalized. In §5.3 I shall discuss the conceptual and experimental framework in which the Shared Intentionality Hypothesis arises, namely the study of the phenomenon of joint attention which Tomasello has contributed to theorize and turn into an independent subject of inquiry in developmental psychology. In §5.4 I shall illustrate the most refined version of the Shared Intentionality Hypothesis to date and discuss some of the critiques to which it is exposed. As I conclude in §5.5, this discussion will form the setting for analyzing the contribution of the hypothesis to the naturalization of collective intentionality in the last chapter of the thesis.

5.2 Collective Intentionality Outside of Philosophy

Over the last two decades, collective intentionality theory has emerged as a prominent research project in the philosophy of social science. The subject, however, is by no means confined to philosophy alone. Some of the problems facing social theorists and philosophers have in fact been tackled with mixed results in several research programs that have grown parallel to the theory of collective intentionality and have engaged with it only recently. In this section I will survey those programs which have tried to provide a naturalistic account of the foundations of sociality, and prepare the ground for the analysis of one specific theory of collective intentionality naturalized.

The problem of what makes collective intentions *collective* has been interpreted by social theorists and philosophers as the question of what mental properties ground the sharing of intentional states[50] (content, type, mode, etc.). But how the sharing is actually effected - what it takes for two persons' mental states to be shared - is a question that the collective intentionality

[50] As we saw in chapter §3, this is evident in Searle's approach to the collectivity of intentional states as well as in most analyses of his account.

literature has failed to address. An articulate answer can instead be found in a body of decision-theoretic literature known as '*team-reasoning*' (Sugden, 2003; Bacharach, 2006). The theory of team-reasoning is proposed to account for situations of interaction called 'pure coordination games' which find their initial formulation in Thomas Schelling's *The Strategy of Conflict* (1960). The skeleton of a pure coordination game consists of two players faced with the challenge of choosing which plan of action to undertake to attain different rewards. Such game is a common-interest game: the interests of the players are perfectly aligned in the sense that the players get the same payoff if both choose the same action profile and zero otherwise.

A modified version of a pure coordination game is the game of 'Hi-Lo', which has attracted the attention of team-reasoning theorists. In the Hi-Lo game one of the action plans delivers a positive payoff ('High') which is strictly better than the other ('Low').

	H	L
H	2,2	0,0
L	0,0	1,1

Fig. 1. The Hi-Low Game

Although it perfectly makes sense that the players choose 'High' because it seems arguably rational and frequent in everyday interaction, mainstream decision theory is unable to explain *why* this is so. One alternative approach is to allow for the fact that agents can engage in forms of collective, and not just individual, intentionality. Focusing on the logic by which people reason in situations of strategic interaction, in the early nineteen-nineties economists like Robert Sugden and Michael Bacharach have proved that it is sufficient for agents to represent one another as members of teams in order to engage in cooperative behaviour. It takes two to tango, according to a paradigmatic example of the collective intentionality literature, so the collective action is brought about by the male and the female wanting to do it together in the sense that they 'see' it as a joint performance, something they intend and engage in as a 'we'.

Recent contributions have explored the connections between the team-thinking and the collective intentionality literature (Gold and Sugden, 2006; Bardsley, 2007). Following Bacharach in particular, it is argued that the view that cooperation presupposes a distinctive mode of thinking, thinking-as-a-team or 'we-mode' thinking, captures the strong sense of collectivity that Searle, among the others, wants his analysis of collective intention to convey. Hence, group-thinking is the carrier of collective intentionality (Bacharach, 2006: 138). The remarkable feature of these accounts is the emphasis on the *process* through which intentional states are shared, rather than on the properties of collective intentional states, which is the outcome of the particular schema of practical reasoning used by individuals as members of groups. Furthermore, in order to find empirical support for their intuitions concerning mental processes of sharing, team-reasoning theorists make explicit and constant appeal to the long history of experimentation in social psychology (see especially chapter 2 in Bacharach, 2006).

The phenomenon of group identification is at the focus of a rich experimental literature on categorization initiated by Henry Tajfel at the end of the nineteen-sixties. The first experiments were designed to test the intuitive idea that individuals behave differently as members of a group (Tajfel, 1970). Evidence, in fact, shows that group-members feel the need to discriminate by expressing a sort of 'positive distinctiveness' toward the members of the same group in contrast with 'outside' members. Tajfel and his student and collaborator, John Turner, identified a possible explanation of in-group favouritism introducing the concept of *social identity*, the idea that human beings can express who they are in terms of 'we' as well as 'I'. According to 'social identity theory', social behaviour can be analyzed as a *continuum* between interpersonal and individual characteristics, on the one side, and intergroup behaviour based upon memberships to various social groups or categories, on the other. However, Tajfel was not concerned with defining social identity in terms of a fixed structure affecting human psychology, but rather his attention was devoted to explain social identity processes in a broader context of social change.

For this very reason, Tajfel did not care much about explaining how personal and social identities co-exist. This theme was taken over and developed in the 1980s by Turner in a novel

strand of research aimed at understanding some issues that social identity theory had not developed in a satisfactory way. The novel theoretical construct, 'self-categorization theory', was thus characterized by the attempt to broaden the analysis from inter-group to intra-group processes by stressing the role played by the cognitive mechanisms responsible for making social identity salient in the formation of group dynamics (Haslam, 2004). In a passage that highlights the similarity between this area of empirical research and some of the collective intentionality theorists' intuitions concerning group-thinking, Turner claims that "a fundamental point of [self-categorization theory] is that when we perceive ourselves as 'we' and 'us' as opposed to 'I' and 'me', this is ordinary and normal self-experience in which the self is defined in terms of others who exist outside of the individual perceiver (…). It is a *shared cognitive representation* of a collective entity which exists reflexively in the minds of individual group members" (Turner and Reynolds, 2001: 135-6; emphasis mine). In sum, social identity and self-categorization theory are complementary accounts of collective intentionality, broadly conceived, in the sense that both support the view that social behaviour can be understood only in light of how people perceive and make sense of the world as a 'we'.

5.2.1 Intentionalism in the Cognitive Sciences

Despite theoretical similarities, the empirical literature in social psychology is largely ignored in the collective intentionality theory[51]. However, recent strands of thought in the cognitive sciences[52] have taken an interested and open stance towards the advances of collective intention analyses. The very concept of 'we-intention' figures in a conceptually variegated body of knowledge that a number of influential cognitive scientists – including Robert Boyd, Herbert Clark, Stephen

[51] Exceptions are Saaristo (2006) and the survey proposed by Guala (2007).

[52] I use the label 'cognitive science' in broad terms to refer, among others, to the sciences of language, developmental and comparative psychology and anthropology (biological and socio-cultural anthropology). As it will become clear in the course of the discussion, this conception includes all disciplines and sub-disciplines concerned with the study of the origins of human cognition and reflects the expertise and background of the scientists involved in this enterprise.

Levinson, Peter Richerson, Emanuel Schegloff, Dan Sperber, Michael Tomasello - have established at the core of a domain of inquiry that has emerged very recently[53].

In the introduction to one of the founding contributions in this literature, emphatically entitled "Human Sociality as a New Interdisciplinary Field", Enfield and Levinson (2006) offer a programmatic vision of the kind of paradigm-shift that they envision in the study of cognition. For too long, they contend, nativism[54] has dominated the cognitive sciences by proposing a wrong-headed and empirically ungrounded approach to cognition that does not make justice to the complexity of the biological and cultural factors at play. Although the consensus towards the twin-track perspective has gained many advocates, the question of what exactly makes the mind the unique product of co-evolutionary factors has long remained the source of speculations and the subject of isolated inquiries. By promoting a systematic dialogue among cross-disciplinary lines of inquiry, Enfield and Levinson identify the peculiarity of mankind in the distinctive character of human cooperation. In spite of its prominence in the history of ideas, then, sociality is once again brought to the front stage of a '*new*' area of research in which collective intentionality plays a significant explanatory role. Before I turn attention to this, let me illustrate what justifies this novel approach.

[53] For a visual representation of this body see the diagram in Levinson (2006: 10).

[54] Nativism has set the debate alight in the cognitive sciences since important discoveries in the studies of mind and cognition began to argue against the view that the environment is the prominent causal factor shaping human cognition. It is worth reminding two contributions in particular: Chomsky's intuition that individuals are endowed with a sort of mental faculty allowing them to naturally communicate through linguistic systems, and the hypothesis put forth by some developmental psychologists that infants show particular predispositions to attend to certain aspects of the world since their early stages of life. As Boden (2006) has suggested in her history of cognitive science, the result was some sort of agreement that inborn propensities could generate mental contents – meanings - given the right sort of environmental trigger alone. A very significant step in this direction was the appearance of a major work published in 1983, *The Modularity of Mind*, in which Jerry Fodor fully endorsed the nativist stance by suggesting an even more radical thesis. In brief, the functioning of the human mind cannot be explained except by postulating the existence of some inner fundamental mechanism. Consequently, what has to be taken as necessary for mental life is not the 'environmental triggering' that provides the mind with the basic food-for-thought in terms of sense experience. Rather, it is the underlying complex of hidden structures that make it possible to process the external data. In Fodor's computational vocabulary, these predispositions or inner mechanisms are called 'modules' and represent the building blocks of the whole cognitive architecture. They are basic, genetically specified and independent units on which the formation of mental contents ultimately depends. There is no way for the mind to have access to the kind of information that is encapsulated in the modules. On the other hand, modules can only be defined by appeal to the functional processes they attend to.

First, the uniqueness of human sociality is not predicated on the variety of its manifestations compared to the level of cooperation observed in other species. It is the *quality* of cooperative behaviour that explains the species-specific ability to realize cultural and institutional systems of astonishing complexity. The foundation of the system for interaction is thus presented as "a coherent subject for investigation constituted by intersecting principles of different orders (ethological, psychological, sociological, and cultural) that work together to produce an emergent system" (Enfield and Levinson, 2006: 1). One core insight of the shift of paradigm urged by many contemporary cognitive scientists is therefore methodological: they propose a multi-layered analysis that involves distinct levels of conceptualization, where cultural factors are interlocked with the study of the proximate and ultimate causes of social behaviour along the lines suggested by Tinbergen's 'four questions' of biological explanation[55].

Significantly, the conventional aspect that unifies all these themes and sets such approach apart in the cooperation literature is *intentionalism*. The central thought is that the world of human interaction is 'mentally mediated' by expectations about each other's behaviour, motivations and mutual beliefs (Levinson, 1995). Intentionalism thus refers to a specialized cognitive faculty – mindreading - which designates the set of skills that enable understanding of people's goal-directed (*i.e.* intentional) behaviour. Mindreading abilities are also referred to under the label of 'theory of mind'[56], one of the most widely accepted and yet critical expressions in the psychological literature. Premack and Woodruff introduced the term in their 1978 target article when they asked whether the chimpanzee – the nearest primate relative of *homo sapiens* - has a theory of mind, by which they meant a system that assigns mental states to other agents, in order to make inferences about their behavior. The huge debate in the primate and infant research that followed the publication of Premack and Woodruff (1978) was stimulated by various proposals contained in the commentaries to the paper.

[55] See §2.5 for an outline of these lines in the overall organization of the thesis.

[56] Theory of mind is an abused term in the cognitive literature: here 'theory' is used in the sense of the actor's understanding of the world rather than the analyst's theory.

Among these, some philosophers, notably Dennett (1978), argued that a convincing answer to the question posed in the title of the article would require that chimps, as well as other animals and young children, demonstrate an understanding that beliefs can be *false*. This criterion was then picked up to test the development of social understanding in infancy and led to the famous 'false belief task', the laboratory gold standard of theory of mind studies (Wimmer and Perner, 1983). Given the voluminous literature on false belies (as summarized in chapter 3 of Carpendale and Lewis, 2006), the concept is currently employed with various meanings spanning from the narrower sense of false-belief understanding, which involves the attribution of beliefs, to social understanding in its most general form. In the latter sense, "ToM [theory of mind] is a domain-specific psychologically real structure, comprising an integrated set of mental-state concepts employed to explain and predict people's actions and interactions" (Astington, 2006: 180).

A simple application of the capacity to understand intentionality is communication. One of the guiding lines of the intentionalist approach to sociality is the paradigm set out in the studies of language by Paul Grice's insights on the cooperative nature of meaning (Enfield and Levinson, 2006: 5-7). In his 1957 paper "Meaning", Grice set the stage for a novel approach to meaning which was in fact going to revolutionize the theory of communication and to become the subject of sophisticated controversies. Grice's analysis of speaker's meaning, which purports to account for meaning in the context of communicative action, takes its point of departure from the logical and psychological structure of the intention to yield a certain behavioral signal, rather than the internal structure of the tokens (utterances, gestures, etc.) issued. The fundamental insight is that the speaker's act of meaning something by a token utterance is equivalent to her intending the utterance "to produce some effect in an audience by means of the recognition of this intention" (Sperber and Wilson, 1995: 21). In ordinary situations, communication is achieved when one's reason for communicating is fulfilled by making it known to the other. This is equivalent to the claim that the simple intention to inform the audience of something, the 'informative intention', is not sufficient to achieve full communication. Gricean ('communicative') intention is fulfilled when the audience recognizes *this* informative intention as the driving motivation of the exchange. Hence, the

communicative act is realized by the publication and recognition of the informative intention with which the act is produced[57].

To sum up, what makes humans a 'cultural' species in the animal kingdom is a set of predispositions for cooperative behavior. These abilities are the building blocks of cross-cultural diversity and of the sophistication of institutional engineering. In particular, the prominent feature of all cognitivist accounts of the roots of sociality is intention attribution, which finds its most articulate formulation in the principles of Gricean pragmatics (Enfield and Levinson, 2006: 5-7). Overall, the ensemble of mindreading skills forms what Levinson (2006) has called an 'interaction engine', the uniquely human adaptation for social behavior that governs the "extraordinary shift in our thinking when we start to act intending that our actions should be coordinated with" (Levinson, 1995: 241). There evidently are aspects of this formulation that echo aspects of the team-reasoning and the collective intentionality literature. Most interestingly, all features listed above find a powerful theoretical synthesis and weighty evidence in the research program of Michael Tomasello, one of the leading proponents of the intentionalist approach to human cooperation and the first scholar to have directed scientific attention to the achievements of the collective intentionality theory in cognitive psychology.

5.2.2 Collective Intentionality in Experimental Psychology

A Co-Director of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Michael Tomasello is one of the most prominent voices in the contemporary field of cooperation studies. A psychologist with interests that span from anthropology to philosophy, since the mid-1990s Tomasello has proposed an articulate theory of sociality evolved consistently with the findings from a battery of experiments with infants and great apes. This theory largely draws on the conceptual resources of collective intentionality theory and is now formulated in a trilogy of studies that begins

[57] Yet, not all communicative situations are structured in such a way that recognition exhausts the communicator's intention. Counterexamples have been designed by philosophers showing that the conditions imposed by Grice are either too flexible (Strawson, 1964) or too restrictive for a thorough definition of communication (Searle, 1969).

with *The Cultural Origins of Human Cognition* (1999), continues with *Origins of Human Communication* (2008) and culminates in *Why We Cooperate* (2009). These wide-ranging contributions have reshaped the landscape of the sociality literature in the cognitive sciences on at least three levels of characterization: theoretical, empirical and methodological.

At the level of theory, Tomasello conducts research on the origins of human development, exploring an uncharted territory in developmental studies when he and his fellow researchers confront issues of cognitive and social development in an integrated manner. The upshot is a fresh formulation of Vygotsky's (1978) dialectic approach to mind and society (Moll and Carpenter, 2007), the view that adult-like forms of cognition develop in a niche of social-cultural exchanges, based on the *observation* that social understanding itself is grounded on a solid inferential basis of mindreading capacities. Indeed society and culture impact on the development of human cognition, but they do so because humans are endowed with species-specific nascent predispositions and motivations for cooperation. These aspects of the theory are tested in experiments on child and primate cognition which are recognized as highly original at the empirical level.

At the methodological level, Tomasello's work testifies to the value of cross-fertilization among disciplines. His method is a paradigmatic example of the recent trend in the cognitive sciences to tackle questions of human development with a threefold comparative approach – ontogenetic, phylogenetic and cultural-historical - which employs a mixed toolbox of resources from developmental and evolutionary psychology, primatology and anthropology-linguistics. In primate cognition, Tomasello and his collaborators compare the cognitive abilities of various animal species on the phylogenetic scale in search for similarities, or differences, with *homo sapiens*; in developmental social cognition, they draw attention to the ontogenetic emergence of the capacities that set the stage for the development of higher social cognitive functions; and in the field of language acquisition they run studies across distinct cultures so as to distinguish universal propensities from those which are environment-constrained. In sum, the scale and volume of the research carried out in Tomasello's lab is probably unique in the arena of contemporary programs on the foundations of sociality.

The theoretical, empirical and methodological insights of this inquiry coalesce in the central hypothesis of Tomasello's work: the 'Shared Intentionality Hypothesis'[58]. For this reason, and before I examine the proposal in much detail, it is worth stressing that Tomasello's hypothesis constitutes a theory of collective intentionality *naturalized* in accordance with the interpretation of naturalization proposed in §2.5. As you recall from the discussion of Searle's and Tuomela's theories in the first part of the thesis, there is a significant difference between the argument that collective intentionality is naturalizable and the question whether there is any viable scientific theory which actually naturalizes collective intentionality. The answer to the latter question is likely to depend on the target of naturalization and on the range of natural scientific methods available. But the purpose of distinct naturalistic programs of collective intentionality is, in general, to give an account that establishes fundamental continuities with the content and methods of 'science' - to be intended in the narrow sense of the body of most highly confirmed and reliable theories as for explanatory and predictive power. In this regard, Tomasello's treatment of the problem of sociality, and his formulation of the Shared Intentionality Hypothesis, is chiefly *scientific* for the methodological and empirical reasons listed above.

This conclusion is supported by considerations about the 'empirical' meaning of collective intentionality. How do we confirm, or disconfirm, that there is a fact of the matter for the ability of individuals to think and act as a 'we'? As we saw, philosophers have employed conceptual tools to explore the naturalness of group-thinking, which results in treating collective intentional behaviour - more or less implicitly - as a theoretical primitive. This means that we-mode thinking can only be tested empirically as part of the theoretical framework as a whole, by its power to explain and predict phenomena like human cooperation and communication. Although whether the theory can be tested is a question that naturalistic philosophers have posed in principle as a by-product of their commitment to naturalism, it has long suffered from lack of answer in practice.

[58] Tomasello is used to frame the contributions of the collective intentionality literature in terms of a 'hypothesis' which he preferably refers to as 'shared intentionality'. The terminological distinction does not stand for any substantive difference.

By factoring shared intentionality into a theory of human development in scientific psychology, Tomasello has made group-thinking eventually susceptible to empirical check. This is not to say that conceptual analyses haven't contributed important insights into the philosophy of collective action and social science of course. The point is that the evidence of collective intentionality is no longer the outcome of commonsense and *a priori* intuitions, but rather the result of a natural scientific approach. To understand this claim in depth, we need turn to the features of the Shared Intentionality Hypothesis.

5.3 The Ontogeny of Intentionality

We can view the history of the Shared Intentionality Hypothesis as a two-stage process stretching from the mid-1990s until today. In the first stage Tomasello defends an 'interactionist' account of human cognition that is committed to the gene-culture approach. What Tomasello adds to this literature is a full-blown account of '*joint attention*' phenomena, a series of social behaviors that correspond in his view to the first manifestation in the ontogeny of social cognition. Since his classic statement in the 1995 paper "Joint Attention as Social Cognition", the activity of Tomasello's lab has established itself as the most authoritative experimental program on joint attention in developmental psychology. Therefore, in this section I will reconstruct Tomasello's version of the dual inheritance model of the mind (Laland and Brown, 2002) in the context of the research in joint attention. Significantly, it is by doing comparative work on the cognitive 'formats' of joint attention and on their role in developmental social cognition that Tomasello has reached a very important discovery: chimps also display some rudimentary capacity for social understanding. This body of theory and evidence has prompted the fundamental shift in his own thinking that led Tomasello to postulate the Shared Intentionality Hypothesis at the foundation of the distinctiveness of human sociality.

We inhabit a world where social-cultural institutions are part of everyday life in such a way that we consider them to be no less objective than natural facts. How can it be that infants learn to see a piece of paper showing certain characteristics as money, for example, and to distinguish it from any

other piece of paper of the same size which is not money? This is the concluding act of a long process of biological and cultural co-evolution that Tomasello describes in *The Cultural Origins of Human Cognition* (1999), his first systematic account of the foundations and uniqueness of human cognition. In brief, Tomasello's view is that nature provides individuals with a basic cognitive endowment which can be later extended and refined through participation in the social-cultural practices of the community. What does this nascent cognitive ability consist in?

The study of cognitive development has historically developed on the backdrop of a broader philosophical debate based on the notions of nature and nurture. The nature-nurture dyad was coined by Francis Galton in 1874 when he claimed that nature is all that a man brings into the world whereas nurture is every influence that affects him after his birth. Since its very inception in the work of ancient philosophers, arguments in favor of either nature or nurture have been put forward to show how a certain pattern of behavior in a given organism originates and develops the way it does. The two concepts, however, have become increasingly loose as new positions emerged along the debate. So, in the current use of the terms, 'nature' can also be read as innate, native, inborn, biological, nascent; whereas 'nurture' stands for learned, culture, environment, socialization. Very generally, what we mean by saying that a property is natural is that it is biologically part of the organism without *external* factors exercising any influence on it. Such a rough definition leaves open a long list of questions, though. For example, does 'native' refer to a specific feature which already exists (or pre-exist) at birth? Is the concept of environment apt to capture the complexity of processes of socialization and enculturation, as well as their influence on the organism's development?

One of the defining features of the current nativist literature is the idea that any process of cultural learning has at its foundations basic skills that human beings share with other primates concerning space, objects, categories, quantities, and so forth (Carey, 2009). Although he argues that cognition primarily develops and flourishes in the realm of culture, a close inspection of Tomasello's writings suggests that he does not actually advocate nurture over nature, but he simply rejects innateness as a working concept in the study of cognition. This is a point which needs to be

fleshed out more carefully. In fact, the originality of Tomasello's account relies to a large extent on the ability to offer a mixed interpretation of the foundation of cognition by criticizing nature *vs*. nurture, which he takes to be a "hoary philosophical debate that has outlived its usefulness" (1999: 48), while nevertheless borrowing many of the concepts from the same debate. For Tomasello, if the target is to shed light on the origin and development of a human trait, taking it as innate does not seem to add anything to the developmental account. The point is not to question the innateness of some allegedly inner mechanism, and to presume that this accounts for the whole process through which the trait came into existence. On the contrary, even assuming that a certain predisposition is observed in human beings since their birth and therefore it is biologically inherited ought to be instrumental to the study of the *process* whereby the feature has become what it is. Nor can innate features be established only on the basis of logical considerations without paying attention to the evolutionary process that actually brought them to light.

However, to acknowledge the role of culture in the development of human cognition is not equivalent to saying that no natural capacity is biologically inherited by human beings at their birth. Tomasello (1999) gives substantive evidence in support of the claim that there is one and only one biological adaptation that human beings are uniquely endowed with at the species-level. He defines this capacity as a "single very special form of social cognition" and describes it as "the ability of individual organisms to understand co-specifics as beings *like themselves* who have *intentional and mental lives* like their own" (1999: 5; emphasis in original) [59]. Hence, Tomasello allows for the existence of features whose transmission is to be explained in biological instead of social-cultural terms, provided that we don't characterize them as innate, which would make them seem 'impermeable' to any further developmental explanation.

Social cultural processes, in other words, contribute to transform this nascent ability into more complex and higher functions which set the stage for the formation of social relations, cultural artifacts, representations and linguistic symbols. In this sense the specific capacity to understand

[59] Whereas the ability of organisms to identify with 'the ones like me' is a general biological principle common to many organisms, the emphasis in Tomasello is on the role of *mindreading* as the distinctive feature of human cognition.

others as intentional agents gives human beings access to the world of culture. "Giving access to" means that any person is able to grasp the intentional significance of a cultural artifact by understanding how it works, the function it has been assigned by previous users and perhaps even by the original creator(s). "To stand on the shoulders of giants" is a metaphorical expression that grasps the spirit behind processes of "cumulative cultural evolution" (1999: 7-8).

These processes can be grouped into two basic types of cultural learning: the ratchet effect and processes of socio-genesis. The ratchet effect is the path along which a certain primitive artifact or practice (*i.e.* considering a piece of paper as money) is brought to existence by an individual or group of people and undergoes a subsequent process of refinement. It thus stands for the idea that the social-cultural world is 'inherited' by human beings in the same way as the biological one (Tomasello, Kruger and Ratner, 1993). More in detail, every time an innovative strategy is carried out, some modification is preserved along the cultural evolutionary scale. So, infants begin to understand the world of symbols and representations that they happen to inhabit because they do not have to understand why and how a symbol (a piece of paper for money) became what it is whenever they encounter it. They simply learn how to deal with it from those who attend to them. Consequently, the artifact has a new form which can be thought of as encompassing all the 'collective wisdom' that accumulated over the cultural history of the group.

The other element is the process of socio-genesis. This notion can be deconstructed in two further components. The first refers to the ratchet process when it applies to an existing artifact or cultural element which comes to be progressively modified across time due to the interaction of several individuals. Why such a process occurs depends on contingent factors. Overall, new cultural needs may arise in such a way as to lead people in a community to devise a new strategy in order to improve the effectiveness of the artifact. The second component refers, instead, to the actual and simultaneous interaction of two or more people working on the same artifact in order to modify it by sharing ideas and further feedbacks. Therefore, culture is a unique human achievement that individuals share in virtue of biological inherited as well as learning-based cognitive mechanisms. Given that all primates are endowed with a set of fundamental cognitive abilities, what

distinguishes the specifically human adaptation for culture is the capacity to understand co-specifics as "animate beings who have goals and who make active choices among behavioral means for attaining those goals" (Tomasello, 1999: 68). In order to understand the emergence and structure of this capacity we need to broaden the discussion to encompass the notion of 'joint attention'.

5.3.1 Joint Attention

In the early months of life, eye-to-eye contact is the main form of interaction between the infant and the caregiver, usually the mother. This form of proto-communication has famously been identified in the literature with what Colwyn Trevarthen has dubbed 'primary intersubjectivty' (Trevarthen, 1979), the ability to share attention within a dyadic format of interaction (Carpendale and Lewis, 2006). Although the point is still under question, bouts of face-to-face interaction increase until they reach some constancy at about six months of life, when the child starts alternating gaze with others on a reliable basis. Interaction is generally established by the mother who introduces a third object close to the mutual line of regard, while the baby alternatively looks to either the mother or the object. The structure of the interaction, then, includes already three 'points' - the mother, the infant, and the object - but the child's engagement in the interaction scene is still oriented at either of the other two.

At around nine months of age, an important event - which Tomasello emphatically refers to as the 'nine-month social cognitive revolution' (1999) - leads young children to experience the various components of the world differently. Since the late nineteen-fifties, this relation has been known in the psychological literature as the triangle of 'joint attention': infants now engage themselves in a *triadic* relation both with inanimate and animate beings (Bruner, 1995). In an extensive longitudinal study of 24 infants aged nine to fifteen months, Carpenter *et al.* (1998) have proved that there is a remarkable synchronic emergence in the appearance of several triadic episodes of mother-infant interaction. This speculation is justified by two observations. The first is that there clearly is an increase of complexity in the ability of the child to understand the set of causes and mechanisms lying behind intentional behavior and phenomena in the world. When the infant engages with the

interactant dyadically before nine months she is likely to do it in a ritualized manner, whereas the behavioral patterns that accompany the nine-month cognitive revolution suggest a form of engagement that outgrows ritualization (Tomasello, 1999). The second observation is that the change in behavior also results in a broader range of attention-coordination abilities[60]. Children now hold up to objects for others to share attention to, and they check back and forth between the other's facial reaction and the focus of attention.

What kind of phenomenon is joint attention? Let us first consider the structure of attentive behavior. Attention is a state of *intentional* behavior corresponding to awareness of something either external or internal to the subject (Brinck, 2001). One view that enjoys currency in the psychological and the philosophical literature is that attention is an occurrence of perceptual intentionality: the attender is an 'intentional perceiver' (Gibson and Rader, 1979). Yet, what is the object of attentive behavior, and how is it distinguished within the cognitive architecture of perception? After the demise of behaviorism in the nineteen-fifties, the psychology of attention has been dominated by research programs that concentrate on attention as a selective process of information-processing (Moll, 2008). Perceptual awareness becomes attentive when the flow of information in the subject's environment is filtered out in a way that leads the attentional focus to be selected.

[60] In the literature, these abilities are generally classified within three classes of joint attention engagement: gaze (and point) following, pointing gesture and social referencing (Carpendale and Lewis, 2006: 82-6). *Gaze following* is the ability of the child to look reliably in the same direction of the adult's gaze. In the early days of empirical research in joint attention, Scaife and Bruner (1975) have devised a procedure which has long been replicated to test the development of the infant's competence to follow another's gaze over the first months of life. Research has since then shown that gaze-following develops over a range of time that extends from three to eighteen months depending on the strictness of criteria used in the experimental setting. *Pointing* behavior refers to gestures orienting somebody else's attention toward some event or object in the surrounding environment (often called 'deictic' gestures). A topic of intense empirical research, pointing is likely to be a form of communication present in all societies (Kita, 2003). Interestingly, the debate about the emergence of pointing behavior in infancy is concerned with the question whether the production of points precedes their comprehension - which is a case of point-following anyway. Bates *et al*.'s (1975) predominant distinction between the function of pointing of either obtaining objects from adults (imperative) or sharing information and experiences with them (declaratives), has recently come under attack after a host of new findings from naturalistic observations as well as lab experiments have suggested novel ways to deal with pointing gestures (Liszkowski and Tomasello, 2007). Finally, *social referencing* refers to uncertain situations where infants look at parents for getting a clue at what they jointly attend to. Hence, the adult represents a point of social reference for the child.

Perception of the intentional object is not just the result of sensory stimulation, though. We can understand the intentionality of attention in a twofold sense. First, attention is a piece of intentional behavior in the sense that for a subject to be aware of something 'as such-and-such' she must be able to single out the aspect of the state that triggers her attention. In other words, perceptual categorization is an essential attribute of attention in enabling the perceiver to identify the specific aspect under which the intentional object is phenomenally presented to her[61]. Second, the intentionality of attention also consists in the subject's motivation to engage in attentive behavior. People embark in an active and purposeful search for information in the service of goal-directed behavior. And this calls for the kind of abilities for understanding intentions and goals that are central in episodes where attention becomes *joint*.

By sharing the attentional focus, the parties to a joint attention exchange are attributed some level of understanding of each other's intentional behavior that, given the age, is supposed to play a grounding role for later forms of social cognition. Yet, what kind of understanding is involved in joint attention exactly? Analyses of joint attention often take off the ground from an example provided by Stephen Schiffer in his study of meaning (Schiffer, 1972). The story is meant to illustrate "a very common, ordinary feature of our everyday life, one which has to do with interpersonal knowledge" (Schiffer, 1972: 30).

> Suppose that you and I are dining together and that we are seated across from one another and that on the table between us is a rather conspicuous candle. We would therefore be in a situation in which I am facing the candle and you, and you are facing the candle and me. (…) I submit that were this situation to be realized, you and I would mutually know that there is a candle on the table. (…) I also know that you know that there is a candle on the table. How do I know this? (Schiffer, 1972: 31).

[61] It has been extensively debated in psychology whether attention entails some active *versus* passive mechanism of information-processing. In fact, attention does not involve only active behavior bestowed upon some clearly identified object. Otherwise it would be impossible to make justice to the fact that infantile perception *seems* to be attentive in ways that resemble adults' intentional behavior, although infants' attention is mostly caught on an involuntary basis. This phenomenon is referred to as 'passive attention' (James, 1890).

The first element to highlight is the format of the scene. Two people stare at an object, the candle sparkling on the table between them, and to each other staring at it. The points of the 'triangle' are: the attender, the first-person subject attending to the object; the co-attender, namely the subject who attends to the object along with the attender; and the object or state of affairs on which the subjects' focuses jointly converge (Campbell, 2002). On the widely accepted interpretation that has emerged from the intense debate of the mid-1990s (Moore and Dunham, 1995), the triangulation of joint attention is not just a "geometrical" or a "psychological" phenomenon of common visual orientation or attention (Tomasello, 1995: 106). The two subjects could each be attending to the candle in the presence of the other doing the same, yet the referential scene would not be one of joint attention (Peacocke, 2005).

The reason is that joint attention designates a suite of phenomena of triangular interaction based on a *perceptual* relation between two subjects and the attended-to entity (Striano and Tomasello, 2001). We must be careful to distinguish the perceptual relation of the joint attention triangle from the perception that characterizes attention as an individual act of intentional behavior. In Schiffer's example, the fact that the two subjects attend to the same object simultaneously is a perceptual phenomenon, but it does not amount to joint attention until both realize that they are attending to the object *together*. The 'jointness' of the joint attention situation, in other words, consists in the special kind of bond by which the attender and the co-attender attune into one another's mind in order to grasp each other's focus of attention. Attention to the object must be mutually experienced in the sense that each subject perceives the other as attending to it *and* display awareness of this very fact, for the relation to be one of joint attention. For this reason it has become customary in the literature to render the full sense of shared attention with the concept of 'perceptual co-presence', where the element of 'co-consciousness' expresses the mutuality of awareness established between *ego*, *alter* and the focus of attention (Clark, 1996).

Joint attention is thus the subjects' mutual understanding that they share attention to an outside entity under the same aspect. What does this state of mutual understanding consist in? And what makes it possible? Before we consider how Tomasello answers these questions, note that our

characterization has pointed to some widely accepted aspects of joint attention, without committing itself to any specific interpretation, or account of the functioning, of it. It is worth reminding us of this non-committal approach insofar as some key-words like 'mutuality', 'awareness', 'mutual knowledge', 'sharing', are used to describe joint attention. Controversies arise, in fact, with regard to the exact meaning associated with these concepts in the relevant literature. For the time being, however, the task is not one of giving an analysis of the meanings of these terms, but rather of illustrating Tomasello's theory of joint attention and how it constitutes the background of the Shared Intentionality Hypothesis.

5.3.2 Joint Attention as Shared Intentionality

The first step of the process leading to the Shared Intentionality Hypothesis is the claim that joint attention behaviors are the first systematic manifestation of social cognition, or understanding, in ontogeny (Tomasello, 1995; 1999). One-year olds are capable of sharing attention with their caregivers because they 'see' the others as subjects of intentional action: they recognize the thoughts and motivations that drive the behavior of others towards the achievement of certain goals. Tomasello then concludes that the capacity of two persons to establish joint attention calls for some articulate form of inferential processing of the kind observed in adult-like patterns of communicative exchange. Yet, despite the increasing amount of evidence in support of this conclusion, the ascription of psychological understanding to children around their first birthday is one of a number of highly debated topics in infancy research, notably the controversy between so-called cognitively 'rich' and 'lean' interpretations of joint attention (Eilan *et al*., 2005), which will be discussed in more depth in chapter 6. In the remaining of this section I shall discuss the issues concerning the theory and the evidence of joint attention that urged Tomasello to revise the initial proposal and move on to formulate the latest version of the hypothesis.

One general difficulty with intention attribution is that it seems at odds with the claim that the jointness of joint attention consists in a perceptual state. To recognize the perceptual nature of sharing attention, however, is *not* to explain the specific mechanisms that bring joint attention

about. In fact, provided that the final state of mutuality is one of occurrent perceptual awareness rather than personal-level inference, the problem remains of how to account for those experiences in which one entertains a state of conscious perception that results, at least in part, from some inferential iteration at the sub-personal, *i.e.* computational, level. This is what happens, for example, with our tacit knowledge of the rules of a grammar for a natural language (Peacocke, 2005). Tacit knowledge of a grammar is usually processed in a way that leads people to *perceive* a sentence uttered in their native language as being ungrammatical, for instance, or as having a certain syntactic or semantic structure – recognition of which requires underlying computation.

There is, however, another difficulty that plagues Tomasello's theory and has important bearings on our discussion. Tomasello makes appeal to the concepts of Gricean pragmatics (Grice, 1957; 1969; 1975) to construe the infant's understanding of sharing attention. As we said, according to Grice, meaning in communication is conditional on the communicator's intentions to produce the intended effect on the audience[62]. Clearly, the communicator succeeds in conveying a certain communicative message to the receiver depending on the fact that they share the same focus of attention. Communication therefore begins with recognition of the *referential* intention, namely the object in the environment to which attention is directed. Having the referential scene set up, however, is not sufficient for the message to get across. By looking in the direction of an ostensive finger, for instance, the recipient might be in the position to discriminate the 'objective' referent among possible candidates without understanding why she should pay attention to that in particular, namely what specific aspect of the object is to be jointly attended. In order for the message to be fully conveyed, the communicator's motive or *social* intention must be grasped by the recipient. Grasp of the social intention is the clue to the subject's understanding of the particular aspect of the referred-to object that the communicator wants her to co-attend.

The claim that two persons have a natural tendency to grasp each other's mental states as a pre-requisite for sharing attention, especially in the context of 'ambiguous' communicative gestures, is

[62] The logic of this construal can also be formulated as "You intend for [me to share attention to (X)]" (Tomasello, 1999: 102).

reasonable. But would we say that this is true of communication at one year as well? The problem is that we tend to attribute psychological understanding to infants based on the intuition that early joint attention gestures are *meaningful* episodes of collective intentional behavior, despite the fact that cognitive and conceptual abilities at this age are still limited in fundamental respects[63]. It is at this point that the question about the mechanisms of joint attention, *i.e.* where the psychological understanding ascribed to infants originate from, becomes urgent.

We can make a start on this task by reminding that joint attention arises when attention is shared in full awareness of the attended-to entity and of each subject's focus of attention. This way of characterizing the mutuality of joint attention often slips into the natural description of the subjects as knowing that they are jointly attending to the same entity. On this description, as the previous passage from Schiffer made clear, joint attention entails *mutual knowledge*[64], expressed in the paradigmatic form of the 'I-know-that-you-know-that-I-know-that-you-…' iteration of propositional clauses. Mutual knowledge of a fact is an information state of a set of people that arises from a situation in which each of them knows about the fact, and each knows that all agents in the group know about the fact, and so on *ad infinitum* (Barwise and Moss, 1996; Sillari, 2008).

However, if it is accepted that joint attention consists in a state of mutual understanding characterized in the terms of the common-knowledge literature, the subjects may need *more* than the inferential abilities of mindreading envisioned by Tomasello. After all – let aside the sophistication of Gricean pragmatics - if all that is needed for the relation of joint attention to obtain were a rudimentary theory of mind, then it would *not* be clear why Tomasello is so explicit in claiming that there is more to the 'mental attunement' of joint attention than simply reading into one

[63] Useful critiques of Tomasello's stance can be found in Campbell (2002), Roessler (2005) and Seemann (2007).

[64] There are several ways to conceptualize the relation between mutual knowledge and joint attention. Unfortunately, the point is hardly appreciated in interpretations of the evidence of joint attention. Most of the disagreement among empirical scientists arises from uncertainty about where joint attention stands relative to cognate phenomena, like mutual knowledge, and *not* from confusion about what kind of psychological phenomenon joint attention is (*contra* Carpendale and Lewis, 2006). For a detailed discussion of the problem see Peacocke (2005). Mutual knowledge turns out to be a controversial concept in the joint attention literature as much as it is in the collective intentionality theory.

another's mind. This additional component would consist in the kind of mechanism that puts the subjects' mindreading abilities at work in achieving the full sense of mutuality of joint attention. One important consequence would be to reformulate the hypothesis by allowing for *this* mechanism, rather than mindreading alone, to account for the uniqueness of human cognition and sociality. What is left out of the initial picture, then?

The response lies in a number of important studies concerning the evolution of human sociality that Tomasello and his collaborators present in the target article appeared in *Behavioral and Brain Sciences* in 2005 (also summarized in Tomasello, 2008a: 44-49). A paradigmatic example of Tomasello's comparative methodology – whereby aspects concerning the ontogeny of social understanding are enlightened by research on the phylogeny of human cognition, and vice versa – this article contains a thorough discussion of the commonalities and differences in the structure of joint attention phenomena resulting from tests on children and chimps (as representative of great apes). The lesson of these studies, most of which conducted in Tomasello's lab, is that chimps are highly social creatures in the specific sense that they display social cognitive skills involving the understanding of their co-specifics' goals and perceptions. But such basic understanding of intentional action is purely individualistic.

In a 'role-reversal' task, for example, one that tests the ability of the players to reverse their roles for the sake of achieving a specified goal in coordination, it was proved that, unlike human infants, chimps do not reverse roles and perform their action without reference to the others (Tomasello and Carpenter, 2005). In another set of tests administered to fourteen to twenty-four-month-old children and three human-raised juvenile chimps, the focus was behavior in instrumental *versus* purely collaborative tasks (where the former involves the pursuit of concrete goals whereas the latter does not). While chimps succeed to coordinate and bring about the desired result in problem-solving tasks, they show no motivation to participate in social games and to engage their partners in the common activity. On the contrary, human infants seem highly skilled not just in carrying out coordination-tasks of instrumental nature, but also in re-engaging their partners in collaborative activities just for the sake of doing things together (Warneken, Chen and Tomasello, 2006).

Based on the large amount of evidence now available, Tomasello and his fellow researchers conclude that the chimps' theory of mind is individualistic in that it lacks grasp of the goal of the collective action as a joint goal. By this Tomasello means that chimps are incapable of understanding the intentionality of collective action from a 'bird's-eye view' (2008: 179; 2009: 68). Namely, as the findings from role-reversal experiments show, chimps cannot capture the roles of their partners from a third-person perspective, one that represents the others as engaging with each other to act cooperatively. Humans, instead, frame the interaction with co-specifics in a single format, which represents the multi-person action as resulting not from the sum of individual efforts but from truly collaborative behavior. Another way to express the idea is to say that chimps do not understand themselves and the others as members of the same group, so their rudimentary intention-attribution is always performed from a first-person perspective. In addition to missing representational abilities of a certain kind, they also lack the fundamental motivations to act in a manner that does not serve individualistic purposes only (see Tomasello, 2009 for an overview).

Based on these interpretations, Tomasello concludes that:

> Human infants create with others joint goals and complementary roles in collaborative activities in a way that our nearest primate relatives do not. The *sine qua non* of collaborative action is a joint goal and a joint commitment among participants to pursue it together, with a *mutual understanding* among all that they share this joint goal and commitment (Bratman, 1992; Gilbert, 1989). Joint goals also structure joint attention, since acting with a partner toward a joint goal, with mutual understanding that we are doing this, quite naturally leads to mutual attention monitoring. And so, one important reason that nonhuman primates do not participate in collaborative activities in human-like ways, or participate in joint attentional interactions in human-like ways, is that although they have human-like skills for understanding individual intentionality, they do not have *human-like skills and motivations for shared intentionality* (Tomasello, 2008: 180-1; emphasis mine).

Two aspects of this passage are of crucial importance for our discussion. First, what is left out of the picture that mindreading abilities are sufficient to establish joint attention is the participants' mutual understanding of the goal of their actions as joint. By assuming a bird's-eye point of view on the interaction scene, humans not only understand each other's goals and intentions behind individual actions, but they *share* them in the pursuit of a collective outcome. Second, Tomasello appeals to the philosophers' concept of shared intentionality to articulate his view of social phenomena like joint attention and, more generally, collaborative activities involving mutual understanding of a joint goal. This formulation of the Shared Intentionality Hypothesis marks the evolution from the first to the second stage of Tomasello's own thinking on the problem of the foundations of human sociality.

5.4 The Shared Intentionality Hypothesis

The Shared Intentionality Hypothesis is the theory of the ontogeny and phylogeny of social cognition proposed by Tomasello in *Behavioral and Brain Sciences* (Tomasello *et al*., 2005), and later refined in *Origins of Human Cooperation* (2008a) and *Why We Cooperate* (2009). The gist of the hypothesis is that human society and culture are underpinned by the species-specific cognitive and motivational 'infrastructure' for understanding and *sharing* mental states. Emphasis on the latter makes this version of the Shared Intentionality Hypothesis differ from the previous one: no longer are intentional states to be understood, they also must be shared. 'Mutualism' – as Tomasello calls the state of mutual understanding achieved by sharing pro-social motives and intention-attribution skills – is a kind of mindset that manifests itself in collaborative forms of interaction unknown in the animal kingdom, where individuals helping others are simultaneously advantaging themselves. This mindset - the "uniquely human sense of 'we', a sense of shared intentionality" (Tomasello, 2009: 57) – is therefore the key to the complexity and variety of cooperative and cultural phenomena. In this section I will discuss the twofold, motivational and cognitive, structure of mutualism and how it illuminates the biological foundations of collective intentionality.

On the motivational side, mutualism could not have evolved but in a scenario where social cooperation prevailed over constant competition, a scenario that Tomasello represents as the result of a 'stag hunt' (Tomasello, 2009: 54). A stag hunt is a common-interest type of strategic interaction in which the best plan for the players is to collaborate ('S' in Fig. 3), because it yields a payoff bigger than the payoffs that the players can get on their own ('H') (Skyrms, 2004).

	S	H
S	2,2	0,1
H	1,0	1,1

Fig. 2. The Stag Hunt

Tomasello speculates that such scenario might have offered the kind of phylogenetic niche required for the emergence of a series of predispositions for acting cooperatively. And when some of these predispositions are also detected in episodes of primate behavior having profound evolutionary roots in great apes – this gives decisive evidential back-up to the claim that humans come into life biologically prepared for altruism.

Consider *helping*, the first social proclivity that Tomasello identifies along with *sharing* resources, like food, and *informing* as a special instance of offering help (2009). Instrumental helping manifests in plenty of real life situations, simulated in laboratory settings, where children typically give assistance to adults in achieving something that falls out of reach. Tomasello proves that helping is emphatically not a form of altruism that depends on parental training or cultural transmission. Interestingly, the best proof of independence from processes of socialization comes from the evidence that external, material rewards decrease the amount of helping expected in a second round of cooperation, in contrast with the commonsense idea that they would rather make children somewhat keener on cooperation (Tomasello, 2009: 13). Notice that, far from being a homogenous and general trait of human behavior, cooperative behavior clusters a complex of tendencies with specific characteristics depending on the domain of activity.

The proposal of a stag hunt as the scenario that best represents the evolution of social cooperation has recently been criticized in the evolution-of-altruism debate (see Silk's commentary in Tomasello, 2009). In the stag hunt the players converge on the socially superior profile because this is what they expect as the best strategy for *each* of them. That is, a stag hunt scenario obtains when individual and group interests are perfectly aligned. In reverse, when the interests of the players diverge from the welfare of the group, like in the typical prisoner's dilemma situation (Fig. 4), the preferred strategy is not one in which all benefit from working collaboratively with each other.

	C	D
C	2,2	0,3
D	3,0	1,1

Fig. 3. The Prisoner's Dilemma

Reasonable as it is, this critique misses the gist of the Shared Intentionality Hypothesis. The point of mutualism is not only to speculate on the kind of evolutionary scenario that might have led to the emergence of distinctively human altruistic behavior. In showing that there are several forms of mutualistic behavior, as we have seen, Tomasello does not direct the scientific attention to the question of whether individual agents are 'generous' or 'nice' towards each other by nature. So, the problem is not what altruism is, in general, and how it became a special feature of human social behavior. To read Tomasello within the framework of the classic question of the evolution-of-altruism debate is simply mistaken (2009: 52). The point of mutualism is mostly to show that without the appropriate cognitive skills, human-unique social proclivities would have never evolved the way they did. So, the question is what sort of mechanism might have enabled humans to start picking the best strategy to everybody's benefit *if* they had not first known how to discern the group's from their own interests. To know how to achieve gains that benefit everybody, single agents must be in the position to discriminate the strategy that favors the group from their own.

What is special about altruism is, in sum, the mindset responsible for the subjects' ability to engage in collaborative activities, namely shared intentionality.

Let us analyze this mindset in detail. A collaborative, or mutualistic, activity is individuated by people acting in the pursuit of a joint goal (Tomasello *et al*., 2008a: 193-4). In the case of joint attentional activities, for example, it is by realizing that they are attending to a common focus that the subjects attune into one another's mind in the full sense of joint attention. What does it mean for two persons to realize that they have the same goal? For a goal to be shared, the agents must represent it as such. As we said, Tomasello uses the metaphor of the 'bird's-eye view' to describe the representational process taking place in the minds of the subjects involved in a joint action. Such process consists in the individuals' ability to grasp the goal of the actions of others, including themselves', in a single format where all are represented as thinking and doing things *together*. At first glance, it might seem that this account is not immune to the charge of circularity that characterizes the debate on the irreducibility of collective to individual intentional behavior. In fact, if a mutualistic activity is enabled by the subjects' understanding of what is relevant for achieving a joint goal, then the problem is to explain what makes them *know* that they have a joint goal to start with.

As it stands, the confusion arises from the meaning of 'knowledge'[65]. What makes the participants in a joint activity know that they have the same goal, which is the pre-requisite for them to attune into one another's mind and establish the jointness of joint attention, is not a distinct representation of 'togetherness' but the type of intentional attitude[66] that brings about shared states in the minds of individuals. For the sake of clarity, single agents realize that they share the same understanding of the referred-to object when they see-together, perceive-together, want-together, etc. (Gross, 2010: 239). Once again, it is the type of psychological mode – intending *together*, or in we-mode – that allows people to share mental states. This characterization of the Shared

[65] This passage is emblematic in this respect: "If we both *know* that we have the joint goal of making this tool together, then it is relatively easy for each of us to know where the other's attention is focused because the locus of attention is the same for both of us: we are focused on that which is relevant to our goal (Tomasello, 2009: 69; emphasis mine).

[66] See §3.2.1 for a detailed discussion of Searle's view.

Intentionality Hypothesis makes it plain that Tomasello follows closely in the steps of Searle's theory of collective intentionality (Tomasello, 2009: 57-9).

In spite of treating shared intentionality as a theoretical primitive, however, when it comes to the exact mechanisms that articulate the sharing of mind, Tomasello often describes the mutualism of collaborative activities in contradictory terms and with reference to the common-knowledge literature. Based on the evidence that humans are relentless mindreaders and that chimps, too, show some rudimentary skill for understanding intentionality, the most recent account of the Shared Intentionality Hypothesis is that mindreading is *recursive* (2008a: 94-6). There is currently much debate on the meaning of recursive mindreading. In fairness, Tomasello acknowledges that there is significant uncertainty on how best to characterize the mutuality of awareness achieved by the subjects in a collaborative activity (2009: 69). On the one hand, what seems problematic with his formulations is the fact that most collective intentionality philosophers, especially Searle, have made it clear that there is more to the 'sense' of collective intentionality than a succession of epistemic states of the kind 'I know that you know that I know that…' issuing in a state of mutual knowledge. In line with what I have shown, in what seems to be his considered view, Tomasello discards this construal in favor of a primitivist account of shared intentionality that bears fundamental resemblance with Searle's approach (2008a: 336). At present, the question of how to interpret the Shared Intentionality Hypothesis is a matter of controversy. In chapter 6, I shall consider various interpretations of the hypothesis and put forward an alternative reading.

5.5 Concluding Remarks

A theory of collective intentionality naturalized treats the problem of collective intentionality as a problem of empirical social ontology, invoking fundamental continuities with the content and methods of science. Although there are currently various research programs outside of philosophy that bear substantial similarities with collective intentionality theory, in this chapter I have singled out Tomasello's program of research as the most advanced theory of collective intentionality naturalized. Three aspects motivate the choice to focus on Tomasello's theory of sociality:

intentionalism, that is, the view that human interaction is mentally mediated by expectations about each other's behavior and underpinned by mind-reading abilities; the integrated nature of Tomasello's inquiry, dealing with phenomena of sociality in developmental social cognition, primate cognition and language acquisition; finally, a novel formulation of the gene-culture approach to the foundations of human cognition.

These elements coalesce in the Shared Intentionality Hypothesis, which I have illustrated in the context of the broader debate on human uniqueness – what sets human cognition apart in the animal kingdom. The current version of the Hypothesis was preceded by the view that *homo sapiens* differs from our nearest primate relatives, such as chimpanzees, in the ability to understand intentional behavior. Since the mid-1990s Tomasello has provided a body of evidence in support of this claim based on the study of joint attention in infancy. Joint attention is the phenomenon by which one-year olds are capable of sharing attention with their caregivers because they 'see' the others as subjects of intentional action. It is by doing comparative work on the underpinnings of joint attention, however, that Tomasello has reached an important discovery: chimps, too, display some rudimentary capacity for social understanding.

What distinguishes humans, according to the latest version of the Shared Intentionality Hypothesis, is the capacity not only to understand but, most importantly, to share mental states. The sharing is made possible by species-specific social inclinations, such as helping, informing and sharing, which could not have evolved but in a scenario characterized by a common-interest type of strategic interaction. To engage in collaborative activities of this kind, certain cognitive skills must be in place including the capacity to read into other minds in a recursive way. There is currently much debate on how best to characterize the notion of 'recursivity', which has suggested problematic interpretations of Tomasello's own thinking on the subject. In order to clear the field from possible misunderstandings, I have indicated some issues of dispute to which we will turn our attention in the next chapter.

Six

Mental Attunement

The Shared Intentionality Hypothesis marks a significant step forward in the naturalization of collective intentionality. In this chapter I shall argue that the SIH provides a strong, evidence-based argument for the irreducibility of collective to individual intentional states. In order to develop my argument, I frame the SIH as an externalist theory of the nature and acquisition of reference in the context of joint attention behaviors. Since its inception in developmental social cognition, commentators have interpreted and criticized the externalism of the SIH in purely semantic terms. In contrast, I argue that pragmatist, rather than semantic, externalism captures the rationale of the SIH. I shall conclude discussing the meaning of reduction, and suggest that the pragmatist lesson sheds important new light on the irreducibility of collective intentionality.

6.1 Introduction

The Shared Intentionality Hypothesis (SIH) has recently been proposed by experimental psychologist Michael Tomasello to account for the uniqueness of human cognition. In a number of contributions, Tomasello and his fellow researchers have argued that shared intentionality is the distinctive trait that sets humans apart from our nearest primate relatives (Tomasello and Rakoczy, 2003; Tomasello *et al.*, 2005; Rakoczy and Tomasello, 2007; Tomasello, 2008a; Tomasello 2009). Over the last two decades, the concept of collective intentionality has featured prominently in debates of the nature of sociality across a number of sub-disciplines in philosophy and the social sciences. Only recently, however, has the SIH made its way into discussions of the development and evolution of the human mind (Tomasello, 2008b).

What does it mean to share intentional mental states? In the first part of this thesis, I have shown that, since its initial formulation, research in the nature of collective intentionality has mirrored the more general concern of philosophers to identify the place of the mind in the natural realm. Naturalists like Searle, among others, have interpreted the central question of collective

intentionality as a demand for the conditions of reduction of collective to individual mental states. Based on the impossibility to individuate such conditions by means of linguistic analysis and intuition, Searle concludes that collective intentionality is a 'biological primitive phenomenon that cannot be reduced to or eliminated in favor of something else" (1995: 24). Claims like this raise various important questions: What can justify talk of collective intentionality as a biological phenomenon? And among those who hold a reductivist view of collective intentionality, how is reduction effected? Clearly the argument for the irreducibility of collective intentionality belongs to a family of issues of broader scope which concern the meaning of naturalization and the role of conceptual analysis in philosophy.

In this chapter I shall confront the issue of the irreducibility of collective intentionality through the lens of Michael Tomasello's research project in social cognition. Tomasello uses shared intentionality to interpret studies from his laboratory suggesting that the capacity to engage in meaningful episodes of communication emerges in children towards the end of their first year of life. In this regard, 'joint attention' behaviors refer to a set of cases in which infants are motivated to, and indeed capable of, understanding their caregivers as subjects of intentional action (Tomasello, 1995; 1999; Tomasello *et al.*, 2005). This characterization suggests a 'rich' interpretation of the cognitive abilities of infants, one that has been criticized for implying a 'mentalist metaphysics' (Racine and Carpendale, 2007). But, several scientists and philosophers now believe that 'leaner' interpretations of the evidence on infant social understanding should be preferred on various conceptual and empirical grounds to rich theories like Tomasello's (see Eilan and Roessler, 2005 for an overview).

However, the issue between Tomasello and his critics cannot be settled by producing novel and more robust evidence in favor or against the SIH. To appreciate the explanatory role of the SIH in accounts of the origin of cooperation and communication, as well as to settle the question of the irreducibility of collective intentionality in naturalistic terms, a shift in the framework of analysis is needed. This shift can be illustrated as a sequence of two steps. First, I shall present the SIH as an externalist theory of the nature and acquisition of reference. The fact that young children around

their first birthday understand the structure of intentions and goals behind the communicative behavior of adults is evidence that reference is not merely a linguistic phenomenon. In its most general formulation, the 'problem of reference' is the problem of how any two persons can know that they *mean* the same thing in communication, be it linguistic or pre-linguistic. However, to describe the SIH as an externalist theory of reference leaves open the question of what specific construal of externalism Tomasello subscribes to. Since the motivation for proposing the SIH is to identify the actual psychological factors that ground reference in the context of interaction, I shall criticize the tendency of most commentators to interpret the SIH as a semantic externalist theory of reference, and I will propose pragmatism as an alternative construal.

The chapter is organized as follows. In §6.2 I shall outline the SIH as a theory of the nature and acquisition of referring abilities in pre-linguistic communication. In §6.3 I shall criticize the approach of semantic externalists who draw conclusions about the nature of reference from commonsense intuitions, which therefore fall short of explaining how reference is settled in early episodes of real-life interaction. Pragmatist externalism, which I shall discuss in §6.4, is best suited to clarify why shared intentionality is a necessary mechanism of reference-fixation in joint attention situations. Another reason for advocating the pragmatist construal of the SIH, which will be the subject matter of §6.5, is that it promises a fresher perspective on the issue of the irreducibility of collective intentionality.

6.2 Joint Attention, Reference and Shared Intentionality

The formulation of the SIH was preceded by the observation that, around their first birthday, infants seem proficient communicators when they interact with adults. What does it mean to be 'proficient communicators' for one-year olds? At this age infants don't make utterances, as language is still far from being acquired in earnest, yet they display basic skills for engaging meaningfully in referential acts. For example, infants show some understanding of what is said by their caregiver before they can 'respond' by speaking, as revealed by their pointing to an intended object in the surrounding (Carpendale and Lewis, 2006). Much empirical research on pre-linguistic

communication is driven by the conviction that infants are *social* in more primitive ways than a purely linguistic approach may suggest. Although this observation is interpreted in a variety of different and often competing ways in the literature, there is wide consensus in describing the 'sociality' of primitive forms of communication as the result of an active 'negotiation' of attention between infants and caretakers.

The key phenomenon is the observation, approximately by the age of nine to twelve months, of a suite of new behaviors grouped under the rubric of 'joint attention' (Moore and Dunham, 1995; Eilan and Roessler, 2005; Racine and Carpendale, 2007). For developmental scientists, it is relatively uncontroversial that joint attention behaviors are episodes of inter-subjective engagement where the subjects are no longer involved in dyadic interaction with either others or outside entities separately. Infants now form a 'perceptual triangle' with adults by holding up to objects for them to share attention to, and checking back and forth the others' body reaction and their focus of attention (Carpenter, Nagell and Tomasello, 1998). The remarkable feature of such interaction is that, by sharing attention to a third object, both parties seem able to discriminate the referential target which they are jointly attending to. The question, then, is how complex the mechanisms are which bring about the mutuality of attention. Great apes, for instance, interact for reasons and in forms that do not suggest the sharing of psychological states achieved by humans in joint attention (Tomasello *et al*., 2005; Tomasello, 2008). Whether this is because humans deploy sophisticated inferential abilities in pre-linguistic communication is the subject of a heated controversy between two classes of explanation of joint attention (Eilan and Roessler, 2005). Before I present the SIH as the paradigmatic example of one class, let me clarify the angle of the debate from which I intend to analyze joint attention.

Questions about joint attention entered the agenda of psychologists in the guise of questions about language development and the evolution of intentional communication, like the transition from pre-verbal to verbal reference (Scaife and Bruner, 1975; Bates, Camaioni and Volterra, 1975). Mainly through the research paradigm established by Jerome Bruner in the nineteen-seventies, the experience of joint attention has come to be seen as foundational to the understanding of reference

in the context of communication (Bruner, 1983). The intuition of Bruner and his followers is that, given its relevance in the ontogeny of cognition, joint attention must provide the kind of format that allows individuals to 'attune' into one another's mind for the sake of sharing the reference of thoughts. In brief, reference is "a form of social interaction having to do with the management of *joint attention*" (Bruner, 1983: 68; emphasis in original). Joint attention formats set the stage for perceiving each other as selectively attending to the same target, leading to the discrimination of reference that makes communication effective (Brinck, 2001).

This characterization can be read in two ways. In one respect, it is highly advantageous to frame questions regarding the emergence of joint attention behaviors in terms of the problem of reference. While joint attention theory is a relatively recent, though rapidly growing, field of research, the voluminous literature on the nature and acquisition of reference offers a deep-rooted and sound basis for exploring the meaning of sharing attention. Theories of joint attention, including the SIH, can thus be understood as theories of reference broadly conceived. In another respect, the joint attention debate has mainly grown out of experimental research in psychology. The evidence on joint attention behaviors has the potential to reveal aspects of the process of reference-fixation that may give a decisive twist to the philosophers' discussions of reference. For example, if the results are robust enough to show that joint attention is achieved by sharing intentional states, it will follow that collective intentionality is a pre-condition of reference, with important consequences for the thesis of the irreducibility of collective to individual intentional states.

What is it to share attitudes in the context of joint attention phenomena? The SIH is a fairly recent approach to joint attention, but the relation between attention and reference has a much longer philosophical history (Eilan, 1998). The most explicit statement of the causal role of joint attention in settling the problem of reference can be found in the writings of John Campbell (2002; 2004; 2005). For Campbell, the 'feel' of mutual awareness experienced by the subjects in joint attention consists in a state of consciousness, which only comes about with co-attendance to the same target (Campbell, 2002: 163). All that is needed for grasp of reference, particularly when young children begin selecting among would-be referents in the outside world, is that the subjects

be aware of jointly highlighting the thing in perception. I call the accounts on which discrimination of reference is rooted at the causal-behavioral level along the lines suggested by Campbell, *lean*[67].

The SIH instead belongs to the class of cognitively *rich* explanations of the nature and acquisition of reference. Tomasello formulates the question of reference as "how children might come to identify more precisely the specific aspect of the world adults intend for them to attend to when using a linguistic symbol" (Tomasello, 1998: 237). Comprehension and production of referential acts like pointing gestures depend then on the ability of the child to discern among possible 'layers of intentionality' (Tomasello, Carpenter and Liszkowski; 2007), by making sense of adults' behavior in terms of the intentional states informing it. The 'richness' consists in the claim that it takes some complex processing at the cognitive level, or active interpretive effort (Roessler, 2005), to discriminate among those layers. Precisely because one-year olds are so proficient in following and producing acts of reference, they must be acting on "a shared space of common psychological ground" (Tomasello and Carpenter, 2007: 121). Tomasello calls 'shared intentionality' this *space*, and defines it as "what is necessary for engaging in uniquely human forms of collaborative activity in which a plural subject 'we' is involved: joint goals, joint intentions, mutual knowledge, shared beliefs – all in the context of various cooperative motives" (Tomasello, 2008a: 6-7).

In contrast with Campbell, what makes Tomasello's account rich is the sophistication of the psychological infrastructure of shared intentionality. As I said, for Campbell reference is established by the subjects' attentive behavior to each other's focus of attention. Joint attention is thus "a primitive phenomenon of consciousness" (Campbell, 2002: 170). If 'primitive' here means that the 'feel' of jointness requires nothing other than the individuals' mutual awareness of the referential scene, joint attention is *not* a primitive state of consciousness for Tomasello. Although he does not clearly pull the mechanisms of joint attention and collective intentionality apart, he claims that for the subjects to think and act as plural subjects they must attune into one another's mind, which however calls for some exercise of mindreading.

[67] I follow the literature and, especially, Eilan and Roessler (2005).

In fact, the SIH lies at the heart of an articulate theory of the foundations of human sociality, which has been refined and expanded in accordance with the findings of a large battery of experiments with primates and children (see Tomasello, 2008a for a survey)[68]. However, there is one major aspect which unifies all subsequent formulations of the SIH, namely the ability of human beings to represent their co-specifics as subjects of intentional (attentional) behavior. In this respect, the joint attention experience represents the first manifestation in ontogeny of the capacity for intention attribution or mindreading. So, the kind of pre-linguistic 'mental attunement' whereby infants understand the reference of communication is essentially achieved by reading into their caretakers' minds.

The SIH is cognitively sophisticated in many respects. One aspect is the problem of how complex, and/or implicit, the child's theory of mind must be for intention attribution to take place. Since this problem has polarized the debate about shared intentionality in psychology, I take it as the departure point for elucidating the conception of reference entailed by the SIH. Let us start by assuming that the child possesses some rudimentary understanding of psychological concepts. The question is what enables pre-linguistic children to recognize the specific aspect of the referential object that the adult wants them to attend to jointly.

[68] The original formulation (Tomasello, 1999) of the hypothesis lends itself to both developmental and evolutionary critiques. Firstly, the ascription of intentional states to others is a pervasive feature of everyday interaction, so it is unclear how mindreading, taken on its own, can make sense of the feel of 'jointness' characteristic of joint attention episodes (Peacocke, 2005). Secondly, if understanding of intentionality is all humans need to perform meaningful acts of communication, which forms the building block of the social-institutional reality, the question is what does prevent primates from creating phenomena of the same complexity since they also exhibit rudimentary abilities for intention-reading (Tomasello *et al*., 2005). Such considerations have urged Tomasello and his collaborators to refine the SIH by distinguishing between socio-cognitive skills and the background of species-unique motivations for cooperation. On the side of cognitive development, "the central unifying concept is something like recursive mindreading (…) between two or more human beings who each know that the other knows, and so forth, back and forth indefinitely –at least in one way of looking at it (Tomasello, 2008: 335). On the evolutionary side, the latest version of the SIH postulates the existence of some species-unique pro-social motives for cooperation (as described in Tomasello, 2009).

Scientists and philosophers tend to interpret this question as a demand for the *background* capacities that allows for recognition of meaning[69]. In the philosophical literature, in particular, it is customary to distinguish the sub-personal level of brute, causally-defined mechanisms internal to minds (brains) from the complex of socio-cultural, external rules and practices that enable mutual understanding of reference. The central thought is that both biology and culture underpin mental attunement, but a dichotomy, rather than the conjunction, between the two categories has long affected interpretations of the origins and development of cognition inside and outside of philosophy. These categories apply to theories of reference as well, including the SIH.

In order to understand this, let us look at Tomasello's conception of the 'common ground'. On his construal, the common ground is the background of concepts and experiences against which communicator and recipient understand each other's referential acts. Various sets of studies from Tomasello's lab have investigated whether infants rely on their shared experience, *i.e.* skills and/or practices, with adults to determine the meaning of otherwise ambiguous communicative acts[70]. But 'experience' is so broad a concept, at least in the way in which Tomasello and his colleagues use it in their studies, that the SIH turns out to be consistent with both biological and socio-cultural background factors. Moreover, to assume that mutual understanding requires *shared* experience takes us to conceive of the problem from a novel, though substantially unchanged, perspective – what makes experience shared in the first instance?

Of course infants around their first birthday do not display adult-like cognitive and communicative abilities. From the opposite perspective, however, it could be replied that they have

[69] The concept of 'background' or 'common ground' has fostered a lively debate across several areas of study, notably pragmatic linguistics (Levinson, 1983; Sperber and Wilson, 1995; Clark, 1996) and the philosophy of language and mind (Searle, 1983; 1992; 1995).

[70] For example, Moll *et al*. have proved that one-year olds know the target object that the adult is referring to while pointing to a range of 'distractor' objects; and recognition of reference is based on the kind of experiences the two of them had shared before with each of the objects in various experimental conditions (Moll, Richter, Carpenter and Tomasello; 2008). The same conclusion is also valid in cases of communication where the referent of the point is unambiguously determined by the features of the context. In these cases, Liebal *et al*. (2009) have provided evidence that infants fix the particular aspect of the referred-to object depending on what they 'know together' with adults (Liebal, Behne, Carpenter and Tomasello; 2009). Shared experience is thus the key aspect in determining not only what object the communicator is directing attention to, but also the reason for co-attending to it (Tomasello, 2008a: 75).

already shared the common ground of rules and conventions of their community for one year; a lapse of time that gives them sufficient resources for working out unambiguously the reference of the other's communicative acts. This is not the line of reasoning pursued by Tomasello, though. For Tomasello, the common ground is necessary to allow people, and especially young children, to make inferences about one another's knowledge so as to anchor it in something they know together. The question, then, is not what grounds shared experience in the first instance, but what the background is ultimately *for*. Since it enables one-year olds to successfully engage in referential acts so early in development, the common ground is evidence of the richness of infant cognition. Further support comes from research on pre-linguistic communication showing that shared information and experience are processed in a way that is not merely tied to the perceptual and behavioral present (Carpenter, 2009).

As a consequence, most commentators interpret the SIH as implying an *internalist* conception of reference. The idea is that, given the early stage in cognitive development, the socio-cultural background that pre-linguistic children share with adults is not 'wide' enough to bring about the mutual understanding of agency, attentional states and goals that constitute shared intentionality. Therefore, infants represent "intentions as internal causal mental entities" grasped via some form of reflective understanding (Racine and Carpendale, 2007: 14; emphasis mine). Notice that, if this reading of the richness of infant cognition is correct, it could easily be taken as backing the view that concepts of agency are part of the core cognition with which humans come into existence (see chapter 5 in Carey, 2009). In addition, such an interpretation of the SIH would be consistent with a broader 'internalist' definition of the nature of mindreading capacities as grounded in some innate or acquired theory of human psychology stored in the brains of people (Stich and Ravenscroft, 1994). The SIH would then appear to be based upon a "mentalistic metaphysics" (Susswein and Racine, 2008: 146), one that accounts for discrimination of reference as the result of access to some internal, *i.e.* psychological, state of mind.

There are, however, stronger reasons for discarding internalist interpretations of the SIH. A more careful reading of Tomasello's writings suggests that he leans philosophically towards an

externalism that mixes evolutionary and causal-historical considerations (Tomasello, 2008a: 10). A disciple of Vygotsky's dialectic approach to the mind (Vygotsky, 1978; Moll and Tomasello, 2007), and a fierce critic of nativism in philosophy and the cognitive sciences (Tomasello, 1999: 48-51), Tomasello (2009) argues at length that biology and culture go hand in hand in determining the evolution and development of human cognition. Whereas the understanding of language and meaning calls for some nascent capacity for mindreading, mental contents (*i.e.* meanings) are not accessed by people through some inborn, modular faculty of the mind. They are embedded in the kind of social environment that "we call culture, and it is simply the species-typical and species-unique 'ontogenetic niche' for human development" (Tomasello, 1999: 78-9). Hence, it would be impossible for people in interaction to grasp the meaning of what is 'said' (broadly conceived) outside of the practices of the community of membership.

The fact that infants display powerful cognitive capacities for referring *in spite of* their sharing little experience of the real world at one year of age is no evidence in support of the thesis that they must have an inborn body of conceptual knowledge. It is an empirical question - Tomasello contends[71] - how infants can ever have developed the resources for unveiling the intentional structure of attentional states by their first birthday; a question that cannot be settled by ignoring the fact that they may have already acquired decisive resources from the outside context of interaction. Thus, while episodes of shared intentionality are enabled by a biologically specified, inherited psychological infrastructure of mindreading skills and motivations, referents can only be individuated against a background of shared meanings.

In conclusion, the SIH is an externalist theory of mind and language in philosophy, coupled with cognitivism in developmental social cognition. Why, then, has the SIH mainly be read as an internalist theory of reference? What has prevented critics from realizing that the internalist-externalist schema is inadequate to capture the complexity of the SIH?

[71] In private conversation.

6.3 Does Semantic Externalism Tell the Full Story about the Shared Intentionality Hypothesis?

In the literature, not only does the problem of reference lend itself to a number of possible approaches, but also the externalist conception of reference contemplates numerous characterizations. In this section I will show that critiques of the SIH that suggest an internalist interpretation hang on the implicit assumption of much philosophical thinking that reference is the privileged notion in the order of *semantic* explanation (see Brandom, 2000 for a critique). This is an unfortunate move which results in a misleading reconstruction of the externalist tenets of the SIH. In fact, whilst a semantic construal of reference is a theory of the meaning of words and thoughts, Tomasello does not construe shared intentionality as a meaningful or contentful attribute of minds, but of their relation with the context of interaction broadly construed.

Externalist theories of reference in semantics constitute a family of views for the idea that an expression refers to whatever it is causally connected to in the appropriate way. Reference is thus determined by causal-historical chains established and stretched in communicative contexts by mechanisms external to the mind. These mechanisms are classified in the two-stage process of reference-grounding and borrowing envisioned by the Kripke-Putnam account of the reference of proper names and natural kind terms (Putnam, 1975; Kripke, 1980). What is crucial of both stages is that perception of the object has a causal impact on the subjects, by triggering acquisition of the semantic *competence* to designate the object with the same name.

Competence in using a name to designate an entity is the ability, causally grounded in the processes of reference-fixation, to use the name without identifying its reference by way of some internal knowledge (Sterelny and Devitt, 1999). By 'internal knowledge' externalists mean the psychological state accessed by one person when she recognizes the reference of thoughts in the outside reality. As Putnam (1975) famously argued, the justification for rejecting the internalist claim that reference is settled by access to some internal descriptive content (Searle, 1983) is that this mental content hardly amounts to knowledge. For it to be knowledge, namely a justified true belief, it needs to be empirically tested. Moreover, the theory of understanding of the causal

theorists suggests that semantic competence is causally acquired also in another sense. Not only is the recognition of reference independent from allegedly internal knowledge, it is causal in the sense that it involves *no* intentional process altogether, no 'thinking about' the designated object – let aside true belief – independent from the process whereby one becomes competent with names. The contention is that "since the name's sense is its property of designating by [a certain] type of [causal] chain, we could say that, in a *psychologically austere way*, competence with a name involves 'grasping its sense'" (Devitt and Sterelny, 1999: 67; emphasis mine).

Does this 'psychologically austere' conception suit Tomasello's conception of reference-understanding in the context of joint attention phenomena? Tomasello describes the relevant mechanisms as follows:

> In joint attention the child coordinates her attention to the object and the adult at the same time that the adult coordinates her attention to the same object and the child. And in both cases this coordination is of a very special nature (…). This implies an understanding of the other participant (…) as a *person* who intentionally perceives a certain aspect of the environment that is *the same* as one's own, or could be made to be *the same* (Tomasello, 1995: 107; emphasis in original).

This characterization clearly lends no support to the interpretative route pursued by semantic externalists in confronting the problem of reference. Psychological understanding is indeed vital on Tomasello's reading for the subjects to establish the 'jointness' of joint attention that help settle reference. In contrast, the causal theorists' view of reference is consistent with a lean conception of joint attention, one that identifies the rationale of joint attention in some behaviorally-based convergence of each other's attention, rather than internal access, to the referential object. Yet, as we said, Tomasello emphatically belongs to the opposite camp which appeals to complex inferential abilities to explain why children are so good at understanding the reference of communicative exchanges so early in childhood. Therefore, the SIH falls outside of the scope of the causal-historical views in semantics.

It might be argued that this failure is due to reasons intrinsic to externalism in semantics. In fact, questions about reference in the order of semantic explanation are traditionally couched in the form of questions about the reference of linguistic expressions[72]. Since the problem at hand is how reference is established between two subjects in *pre-linguistic* contexts of interaction, the semantic approach –the argument goes- cannot address the question of the SIH in a satisfactory way. While this objection sounds correct at first sight, it is worth pointing out that externalist discussions are largely animated by the question whether 'meanings are in the head' or not (Putnam, 1975; Searle, 1983). In other words, it is at the level of the reference of thoughts, and not just words, that the debate between internalist and externalist theories takes place. Therefore, we must look at this side of the controversy to understand why externalism in semantics fails to capture the thrust of the SIH and, more broadly, of Tomasello's externalist stance in social cognition.

The question is: What becomes of a theory of shared mental states on the picture sketched by semantic externalists? Consider the *locus classicus* of externalist discussions, Putnam's Twin Earth thought-experiment (Putnam, 1975). Putnam construes his argument in such a way as to demonstrate that physical duplicates – subjects populating Twin planets who exhibit the same history and body structure and, therefore, psychological experiences - can nonetheless mean something different when they utter the same word, say 'water'[73]. Since the duplicates perceive water in the same way - as a transparent, odorless, etc., entity - the linguistic intuition of many commentators (but not all[74]) dictates that a difference in understanding of the reference of 'water' will reflect a difference in chemical properties alone. Externalists then conclude that the conditions for the existence and identity of thoughts are to be found in the existence and identity of the entities thought about, namely their actual reference.

[72] It is worth recalling that the Kripke-Putnam line of argumentation takes the problem of meaning in the context of natural languages as its starting point, focusing then on the reference of proper names and natural kind terms.

[73] By construction, 'water' designates substances with identical external features but distinct chemical compounds on the Twin planets.

[74] See Crane (2003) for a critical discussion of this intuition.

It thus becomes clear that, if two subjects knowingly understand each other as referring to the same object in the world, there must be a fact of the matter that identifies the state of shared intentionality. So, for two subjects to 'share' mental states, their minds need only be causally connected with the referred-to entity in the appropriate way. Causation, along with the socio-linguistic practices of the community, guarantees the worldly 'anchorage' that identifies the conditions for individuating mental contents. Understanding of reference, in sum, is a metaphysical fact, a fact about how things *are* in the world rather than how any two subjects sharing the same contents perceive, describe or know about them (Rey, 1983).

Consider now what follows from the view that reference is the clue to the metaphysics of meaning. The possibility that two subjects associate their communicative gestures with the same referent, like in the joint attention triangle, is indeed part of the overall picture but as a *logical* possibility, one that semantically-leaning philosophers mainly explore in counterfactual terms by means of modal reasoning and thought-experiments. Questions about what makes it possible for two subjects to associate their communicative gestures with the same referent are a metaphysical matter, one that can be explored in abstraction from the psychology of people. As we said, semantic externalists aim at rejecting the classic (internalist) claim that knowledge of reference is needed for understanding of language, which is only possible –they contend- because of the causal-historical conditions that tie the subjects' mind to the outside world.

However, it is unclear what these conditions are other than Putnam's generic appeal to the 'linguistic division of labor' within the community of membership. *This* question simply falls out of the philosophical agenda of semantic externalists, and demands an alternative approach to reference that sheds light on the the externalist presuppositions of the SIH left out of the semantic picture. When the problem of reference is treated in the lab as a psychological rather than a logical problem, in other words, it becomes evident that the picture is more complex than, and partly contradictory with, the one described by semantic externalists. Tomasello shows some awareness of this problem when he claims that his analysis

> does *not* of course touch on the *logical* problem of reference with which Wittgenstein (1953/2001) and Quine (1960), in particular, were so deeply concerned (…). Empirical observations cannot solve this logical problem, but they can demonstrate the surprising fact that in the real world young children do not very often seem to have enormous difficulties in determining specific referents – provided that they are in certain kinds of communicative situations. This is a very interesting psychological fact that itself requires explanation (Tomasello, 1998: 237-8; emphasis mine).

To sum up, there appears to be at least one important reason to discard a semantic construal of externalist interpretations of the SIH. The hypothesis of shared intentionality implies that the understanding of reference is not just a *logical* possibility but a psychologically real situation. People come to identify the reference of their communicative acts way before language is acquired in earnest, not just because their individual minds are causally connected with the outside reality in the 'right' way, but most importantly because they have an ability to attune into one another's mind in the context of interaction. Therefore, it is in the triadic relation between the subjects and the object of reference, rather than in the dyadic relation between each person's mind and the world, that we ought to find the mechanism of shared intentionality that grounds understanding of reference.

6.4 The Pragmatist Roots of the Shared Intentionality Hypothesis

Reference is an act of communication, but communication is not exhausted by language. I have already mentioned that communication is possible because people can have the same understanding of reference, provided that their minds are causally connected with the world in the appropriate way. But once this connection is secured, how is communication actually effected?

Most semantically-oriented critics neglect that, for Tomasello, communication requires that the field of reference be shared. It is by construing the space of interaction as 'a shared space of common psychological ground' that two subjects realize that they have the same reference 'in

mind'. Thus, if there are reasons to believe that Tomasello leans towards externalism in general, this is because he confronts the problem of reference from a *pragmatist*, instead of semantic, standpoint. In this section I will discuss the pragmatist approach to the cooperative nature of pre-linguistic communication, and demonstrate that it best enlightens the conceptual background of the SIH.

Externalist conceptions of reference in semantics and pragmatics trace back to the same ancestry, so pragmatists are externalist in spirit (Wittgenstein, 1953/2001; Quine, 1960). Both families of theories have been inspired by the rejection of the 'classic' view that knowledge of reference comes prior to, and 'mediates', active processes of referring. However, whereas causal-historical views in semantics explore the conditions for the possibility of reference in theory, pragmatists focus on the contextualization of language and thought in practice. The idea is that acts of reference are not performed *in vacuum*; on the contrary, they are highly context-sensitive. So, the problem of reference is to understand what gives the subjects clues to the targets of referential acts in the actual context of interaction, rather than in abstraction of it (Quine, 1974).

A typical example of a pragmatist conception of reference is Jerome Bruner's, whose empirical studies of the ontogenesis of reference in the context of joint attention phenomena have deeply influenced Tomasello's own thinking and have paved the ground for the externalism of the SIH. Whereas in conceptualizing the role of the context he evidently endorses Putnam's causal-historical theory of reference (Bruner, 1983: 67-8), Bruner's contribution represents a relevant step forward in amending the flaws of the semantic construal of reference, as it helps isolate the specific bits of the causal chain that links reference grounding to reference borrowing[75].

For Bruner the psychological problem of reference is one of disentangling the 'standing for', *i.e.* the relation of the mind with the world in the context where referring takes place, rather than identifying the "isolated bit of mental furniture produced by the linking of a sign, a thought, and a referent" (Bruner, 1977: 275). By 'isolated' Bruner alludes to the fact that reference must be contextualized: "It is obscure what any utterance refers to and means independently of the contexts and conditions in which it is uttered" (Bruner, 1983: 17-8). However, it is precisely when Bruner

[75] For an introduction to the limits of causal theories of reference see Devitt and Sterelny (1999).

articulates the causal-historical chain in terms of the psychological notion of context that his self-proclaimed externalism appears less radical than previously supposed.

In pragmatics 'context' is a psychological construal that refers to *all* features of the perceptual and cognitive environment which establish mutual understanding of reference (Sperber and Wilson, 1995). This is to say that the context is not just the physical ambience of play but also, more broadly, the set of *intentional* facts about the subjects performing referential acts, including people's "beliefs, intentions and expectations" (Stalnaker, 1973: 447). Thus, by acknowledging the role of psychological understanding in the process of reference-fixation, pragmatists take a significant departure from the purely causal-historical approach advocated by semantic externalists. All in all, while Bruner generally favors semantically-driven explanations of reference in philosophy along externalist lines, his psychological approach to reference constitutes a problem space of its own.

Let us look at the features of this approach more in detail. As I said, Bruner is credited with having laid out the mainstream approach to the study of reference and attention in experimental psychology (Moore and Dunham, 1995). Reference is conceived as the process whereby "somebody communicates to another person that there is something particular at the focus of his attention the he wishes to bring to the attention of that other person, in return for which he wants some indication that the other has, as it were, 'got the message'" (Bruner, 1998: 209-10). The inter-subjective agreement essential to establish reference depends then on "developing *procedures for constructing and using a limited taxonomy* for distinguishing among limited arrays of extralinguistic objects" (Bruner, 1977: 275; emphasis in original). What kind of interpersonal and psychological procedures do establish joint attention then?

Communication in pre-linguistic children, and presumably in early humans before language, draws on a vast repertoire of gestures to direct and follow someone´s attention to a given target. Among these, deictic *i.e.* context-directed gestures are responsible for 'transferring' knowledge of reference from the caretaker to the child along the causal route. The typical scene is one in which "infant and caregiver look jointly at a common object, then look back to each other eye-to-eye with

evident enjoyment" (Bruner, 1998: 212). But the context for Bruner is not just responsible for the intelligibility of the reference of thoughts. The sense in which the environment of action 'anchors' the reference of the subjects' thoughts and actions is not the sense envisioned by semantic externalists. The point is that the recipient of a message cannot *practically* identify the reference of a gesture in isolation from the real context of interaction.

Joint attention procedures thus shape the field of reference by imposing the relevant constraints for what is to be attended jointly. According to Bruner, they form a 'scaffold' or format which contributes "to get the infant started in the business of figuring out what was being *meant* by what was being *said* – what interpretants were needed to form a bridge between a sign and its significate(s)" (Bruner, 1998: 220-1; emphasis in original). Similarly, for Tomasello the context is not the terminus of thoughts in the reality, that is, the worldly anchorage that causes thoughts to be accessed as 'thus-and-so'. In fact, the parties to a communicative exchange cannot recognize what each refers to unless they share the 'right' meanings and experiences. And the problem of 'what-stands-for-what' can only be solved by engaging in the socio-cultural practices of the community where reference is grounded and borrowed.

This first divide between the pragmatist meaning of 'context' and the semantic one only tells part of the story about the conceptual roots of the SIH. As it stands, in fact, the above conception would suit any cognitively lean account of joint attention that strengthens the role of merely causal-behavior mechanisms in reference-fixation. In order to reconcile the cognitivist character of the SIH with externalism in pragmatics, we need to take into consideration one more issue that Bruner only formulates in general terms, and that Tomasello has later developed in more systematic ways. This aspect has to do with the fact that the acquisition of meanings by any one person, although socially and culturally variable, requires significant discriminating knowledge on the side of the subjects involved in communication; something that semantic externalists would hardly accept (Evans, 1982).

Thus, Bruner often feels urged to question whether pre-linguistic communication requires some form of understanding on the side of the subjects *over and above* the causal link that grounds the

reference of their thoughts in the actual reality. "Does a young child naturally 'understand' (…) that she and her mother are looking at something together, sharing a common experience?" (Bruner, 1995: 2). In other terms how do people direct each other's attention toward the specified object of reference, if they don't *know* that they share enough in their attentional focus? These questions suggest that reference needs more than some contextualized clue to be fixed. First and foremost, young children and their caretakers need to realize that they are acting within a joint field of reference. And to make the context joint is for each subject to grasp the common target of reference and understand that the other grasps it too. This is evidently not an internalist conception of reference – minds refer in virtue of their causal connection with the objects of reference. But it does not fall either into the province of semantic externalism insofar as some mechanism of intentional understanding is considered vital for establishing a comprehensive conceptual ground for reciprocal exchanges. Tomasello brings this line of reasoning to completion by introducing the concept of 'recursive mindreading' (Tomasello, 2008a). Shared intentionality is not only underpinned by pro-social motives and inferential skills, but it also, prominently, issues in a state where the subjects are mutually aware of referring to the same object in the world. And he concludes: "In explaining how contemporary humans operate in real time, it is possible that no notion of recursivity is actually operative, but rather humans simply possess a *primitive* notion of we-intentionality. Indeed, I think this is exactly what young infants do" (Tomasello, 2008a: 336; emphasis mine).

Based on intense scrutiny of the evidence available, Tomasello arrives at the conclusion that collective intentionality is an irreducible feature of human psychology. The common space of interaction, which for pragmatists makes communication possible, must also be construed as irreducibly 'collective' by the subjects for them to discriminate among possible referents. And this calls for the capacity of individuals to think of themselves and others primarily in the first-personal-plural perspective.

6.5 Irreducible Collective Intentionality

Shared intentionality is therefore a pre-condition for knowledge of reference. Developmental scientists have not fully appreciated the importance of such conclusion. But neither have collective intentionality philosophers worked out the consequences of the SIH enough to realize that it provides an answer to the fundamental question of the irreducibility of collective to individual intentional states. It would then be helpful to specify the relation between shared intentionality and reference in more detail. However, to develop this point into a novel version of the irreducibility thesis would engage us in a different project from that pursued here. In the remaining part of this paper, then, I will only concentrate on setting the stage for this future direction of research, by explaining why the experimental program in cognitive psychology can fruitfully meet the irreducibility problem and contribute a step forward in the naturalization of collective intentionality.

At the outset I defined collective, or shared, intentionality as the capacity of people to think of individuals – both themselves and the others - as members of groups. Whether this capacity is irreducible has been a highly debated matter in the last two decades. The irreducibility-question is commonly framed by asking whether the concept of collective intentional behavior can be decomposed into the concepts that we already deploy in understanding individual action and thought. And the answers, by and large, fall in two camps. For non-reductivists like Searle there are plenty of arguments supporting the idea that group-thinking requires some primitive 'sense' of sociality. Reductivists like Bratman, in the other camp, hold that all is needed to share intentions is that the mental states of the individual agents be properly connected and supplemented with mutual knowledge.

This debate presupposes a conception of irreducibility such that a cognitive trait either is irreducible or is not. But why should one subscribe to this dichotomic construal of 'irreducibility'? Let us examine the status of the claims made by collective intentionality philosophers when they advocate, or reject, the irreducibility thesis. The argument is that, if no account succeeds in reducing the *concept* of collective to individual intentions in non-circular fashion, we will have reasons for being realist about collective intentionality. If any account succeeds, the opposite conclusion is warranted. But this is equivalent to saying that both reductivist and non-reductivist philosophers

reach conclusions concerning the alleged 'naturalness' of collective intentionality by means of commonsense intuitions. Conceptual analysis is now implicitly assumed to be an adequate method for settling ontological questions concerning the place of collective intentionality in the natural realm. So, questions like 'Is there a fact of the matter that justifies realism about collective intentionality?' are addressed by exploring the folk-psychological attributes of collective intentionality in ordinary language. And the result of this analysis is taken as 'evidence' for or against existential commitments.

As the history of science shows[76], however, the claim that a feature is a theoretical primitive means that it cannot be reduced to more basic constituents in the same *definitional* domain, and not that the feature is a brute fact, or that it cannot be given a reductive explanation in another theoretical framework. For experimental scientists like Tomasello, the only evidence that backs up or refutes claims about the ontology of collective intentionality is empirical. Tomasello is often critical of those analyses which aim at providing the conditions of possibility of intentional phenomena on *a priori* grounds – analyses that he refers to as 'logical'. It is up to science, notably scientific psychology, to discover whether there are reasons for being realist about collective intentionality. In this sense, to be sure, the SIH is a naturalistic account of the mind, or at least of part of it, because it exploits the tools and techniques of science for the solution of philosophical problems of social ontology. To put it in pragmatist terms, it is in the context rather than in abstraction of it that shared intentionality finds its place in the order of things, whether the 'context' is real-life situations or the psychologist's lab.

Therefore, the fact that collective intentionality is irreducible in the intentionalist, *i.e.* folk-psychological, framework of agency is no proof that it is a basic intentional phenomenon. *Prima*

[76] Recall the example of fitness in evolutionary biology that I have illustrated in §2.4.1. Defined as the capacity of biological organisms to adapt to their environment, fitness is the key concept of the theory of natural selection and, therefore, it is essential to evaluate the explanatory power and testability of Darwin's theory. For philosophers like Alexander Rosenberg, the only way not to trivialize the theory of natural selection is to treat fitness as "a primitive or undefined term *with respect to* the theory of natural selection" (Rosenberg: 1983: 463-4; emphasis in original). The 'primitiveness' of fitness is postulated with regard to its definitional domain; indeed it might be the case that fitness contemplates a reductive explanation outside of the theory of natural selection.

facie, also Tomasello formulates his conclusion regarding the irreducibility of shared intentionality in the language of intentional agency. But there is a crucial difference between the philosophical and the scientific understanding of collective intentionality, which a closer look at the meaning of reduction will help us bring to light. As I explained in §3.5.2, what we know when one entity is reduced to another is that the former can now be explained in terms of the latter *also*, precisely because the two are ontologically the same. But to make one thing more intelligible by showing that it actually is another thing is an explanatory (epistemic) quality of reduction, not a new fact in ontological terms. We still refer to the two entities as if they were distinct, when in fact they are one and the same. Along the same lines, when philosophers dispute whether collective intentional states can be reduced to their individual components, what they mean is that perhaps we don't need the *concept* of collective intentionality because we can understand collective intentional behaviour in the terms of individual intentional behaviour. And, once again, no ontological conclusion follows from there.

But this is not the sense in which Tomasello concludes that reference-fixation is underpinned by the kind of motivational and inferential 'machinery' that articulate the SIH. In the context of a natural-scientific explanation of human development, Tomasello does not test the concept, but the *capacity* of shared intentionality. That is, to say that reference requires people to construe the action space as one of shared intentionality, *i.e.* to reason as a collective, is not to say that the agents must have the same, irreducible understanding (concept) of collective behaviour. It is, rather, to make the empirically-grounded causal claim that certain cognitive and motivational abilities must be in place for establishing the reference of communicative exchanges. Thus, it is because individuals are motivated and indeed capable of reading into their minds that they grasp what it is for each of them to refer to something together. This argument closely follows in the steps of Searle's (1995) view that, in collective intentional behaviour, the contents of first-person singular thoughts are *derivative* from group-thinking - but with an important proviso. The claim is now made against a background of scientific causation, rather than folk-psychological explanation.

In general, the lesson of the naturalistic conception of irreducibility is that conclusions about the ontology of collective intentionality are to be drawn relative to the conceptual background against which the issue is framed, as well as the best empirical evidence available. So, the argument that shared intentionality is a pre-condition for understanding reference, which is the most original outcome of Tomasello's pragmatist approach to the origins of communication, is particularly significant for the naturalization of collective intentionality because it backs up the irreducibility thesis on sound scientific grounds.

6.6 Concluding Remarks

The work of Tomasello and colleagues has attracted a great deal of interest among those who want to give a naturalistic account of the roots of sociality based on the capacity for collective intentionality. But the hype surrounding this research program seems premature; more needs to be done to elucidate the conceptual bases of what I called the Shared Intentionality Hypothesis (SIH), before we evaluate the implications of experimental psychology for the naturalization of collective intentionality. As I have emphasized, the debate about the acquisition of reference in communication offers a solid standpoint to distinguish among interpretations of the SIH.

More specifically, I have shown that there is a profound conceptual difference between Tomasello's pragmatist theory of reference and the semantic interpretation his critics have attributed to him. First, reference is not merely a linguistic phenomenon, as is proved by the ability of infants to successfully engage in referential exchanges before language is acquired in earnest. It might be replied that semantic discussions mostly revolve around intuitions and thought experiments about the reference of mental states rather than words. But such a reply makes the contrast between Tomasello and the semantic externalists all the more evident. The latter are interested in the conditions that make the reference of thoughts logically possible, while Tomasello and his fellow researchers explore the psychological features of the problem of reference in practice.

The alternative formulation of the SIH that I favor takes into consideration that Tomasello comes from a different background from most of semantic externalists. His approach is rooted in the pragmatist theory of reference of Jerome Bruner. The field of interaction, according to pragmatists, offers the objective anchorage that the parties to a communicative exchange require to recognize the reference of their acts. Distinguishing between semantic and pragmatist theories would then enlighten Tomasello's particular view of shared intentionality as a common psychological 'space' of mutual understanding. As we have seen, knowledge of reference relies on shared intentionality to the extent in which it requires that the subjects attune into one another's mind via a joint construal of the referential field.

The argument that shared intentionality is a precondition of reference-acquisition exploits only part of the potential of the SIH in tackling philosophical issues. There are still plenty of questions about the nature of sociality that the SIH could fruitfully address which nonetheless prove recalcitrant to the methods of purely conceptual analysis. The current trend towards an interdisciplinary approach to the philosophy of society goes precisely in the direction of strengthening the role of empirical evidence in philosophical quarrels, to be coupled with an increased attention to the conceptual foundations of experimental research.

Seven

Conclusions

I have started this investigation about the problem of collective intentionality with two platitudinous observations. One is that the clue to everyday interaction is the capacity of people to share mental states. Philosophers capture this intuition by noticing that, when you and I come together to achieve a common goal intentionally, the fact that we do something together implies that neither one of us does it on her own. We can in principle understand what we do in distributive terms, as the result of individual actions; but there also is a collective reading based on recognition that the goal of our actions is achieved by the two of us intending and enacting things as a group. This capacity of individuals to represent themselves and others as thinking and doing things together is collective intentionality.

The other observation is that, despite its intuitive strength, social theorists and philosophers have encountered significant difficulties in giving an account of the nature of collective intentionality. In the classic framework, this problem is formulated in terms of the question whether there are conditions for reducing collective intentional states to the concepts that we already deploy in understanding individual action. In the first part of this thesis, I have analyzed various answers to the irreducibility question, which however exhibit the same methodological feature. Whether one person entertains irreducible collective intentional states is a matter of how decisive the conceptual evidence is in support or against either solution. By 'conceptual' I mean the evidence that derive from the analysis of collectivity concepts by way of thought experiments and counterexamples.

A serious problem arises at this stage. There is a sense of 'irreducibility' by which the reduction of collective to individual intentionality is an epistemological issue. It is the question whether our understanding of collective intentional behavior goes via understanding of individual behavior. However, in discussions of the nature and structure of collective intentional states, the epistemological question is mostly interpreted as to whether there is a 'fact of the matter' for the

capacity of group-thinking. Evidently, this formulation calls into question ontological, not only epistemological, considerations regarding the existence and identity of collective states. As a consequence, it is not possible to confront issues concerning the ontology of collective intentionality on the basis of intuition alone. For naturalists of all stripes, including collective intentionality theorists, the only way to ascertain the naturalness of collective intentionality is by doing the appropriate sort of science.

The naturalization of the mind is one of the most prominent and wide-ranging research projects in contemporary analytic philosophy. In very general terms, naturalistic theories of mind purport to show that one or more fundamental mental properties have their foundation in the natural order of things. When theorists debate the naturalness of collective intentionality, they usually question whether first-person plural intentional predicates can be reduced, at least in principle, to states of the brain, the study of which undoubtedly falls in the domain of science. But, as I have argued at length in this thesis, there is more to the naturalization of collective intentional states than arguments that give reasons for *believing* that they are natural attributes of reality. This suggests the following, crucial question: Can we give a naturalistic account of collective intentionality based on scientific theory and practice? In this conclusion I shall try to answer this question by summarizing the results of the previous analyses, and by pointing to future directions of research in the naturalization of collective intentionality.

Studies of collective intentional behavior tend to exhibit high eclecticism in their methodology, due to the inter-disciplinary nature of the subject and the availability of various methods of inquiry. Since 'naturalism' is an il-defined concept with various meanings in philosophical and scientific circles, in chapter 1 I have argued for the distinction between naturaliz*ability* arguments of collective intentionality and theories of collective intentionality naturaliz*ed*. The former put forward a realist account of collective intentionality in philosophy: to be realist about collective intentional states is to assume that they are real, *i.e.* true of the reality that we inhabit. Hence, for collective intentionality to be naturalizable is for us to believe that the difference between the intentionality of the first-person plural and that of the first-person singular is a natural attribute of the world.

There are two strategies that pursue a realist approach to collective intentionality. One has been first offered by Searle in the context of biological naturalism, the view that mental states are as real as any other biological phenomenon. Searle argues that there is something irreducible to our capacity to intend and enact things as a collective, which cannot be captured in terms of first-person singular states plus mutual beliefs. Collective intentionality consists in a peculiar kind of intentionality which lies in the brains of individuals and cannot be decomposed into more elementary units. The problem of naturalization, Searle concludes, is a particular case of the problem of treating intentional predicates in a way that makes them consistent with the physical facts, broadly construed. Since any acceptable program of naturalization of the mind ought to rely on intentional causation as its working hypothesis, Searle's realist attitude is inspired by a belief in the causal continuity between the biological reality of the mind and the power of mental properties to yield physical effects.

This conclusion concerning the naturalness of collective intentionality, however, does not follow from the claim that collective intentional states cannot be reduced to or eliminated in favour of something else. Indeed the two claims admit of different kinds of evidence in support or against them: the former is subject to scientific scrutiny whereas the latter results from the conceptual analysis of collective intentions. In this respect, it is very important not to misunderstand Searle's words, notably the fact that he seems to grant a central explanatory role to conceptual evidence in drawing naturalistic claims, as if the ontology of collective intentionality were postulated on the basis of linguistic analysis alone. In chapter 3 I have contrasted the general dissatisfaction towards aspects of Searle's philosophical approach with arguments that show that, for him, existential conclusions about collective intentionality are metaphysical claims that derive from the belief in the completeness of physics, and not from the analysis of collectivity concepts. If there is something controversial to his theory is, rather, the fact that those conclusions are premature in light of the state of research in neuroscience and the cognitive sciences more generally.

An alternative defence of realism is offered by Tuomela in the context of a social-constructivist account of intentionality. The contention is that collective intentional behaviour must be

conceptually presupposed, in the sense that intentionality is what the members of the linguistic community take it to be by common acceptance. This characterization brings to the fore the main difficulty that we have encountered in presenting social constructivism as one possible naturalizability argument. If there is a core idea to 'construction' in the literature, it is that certain aspects of reality including mental properties (meaning, intentionality, etc.) are under human, *i.e.* social-cultural, control rather than the control of nature. In chapter 4 I have argued that this general definition of social constructivism is open to many readings, and a mild version is to be preferred in interpreting Tuomela's realist strategy if only to make justice to his self-proclaimed commitment to naturalism.

For Tuomela, what is socially constructed is not the capacity of people to entertain 'genuine' mental states, but their contents, *i.e.* the way in which people access those very states. In fact, through the prism of Sellars' verbal behaviourism, I have argued that Tuomela makes a point that concerns the epistemology of collective intentional states, based on the view that language comes prior to thought only at the epistemological level of explanation. There is, then, a significant difference between this interpretation and a family of more radical arguments which accounts for the ontology of collective intentionality as the result of processes of socialization and enculturation. Therefore, Tuomela's argument for the naturalness of collective intentionality is consistent with a broad construal of naturalism in contrast with the anti-realism of the argument that collective intentionality is intrinsically theory-dependent.

In spite of tackling the problem of collective intentionality from different perspectives, the underlying theme of Searle's and Tuomela's strategies adopt a common method of investigation. Their analyses proceed from the intuition that there is more to the understanding of collective intentional behavior in we-mode than the understanding in purely I-mode terms would suggest. There might be a problem, then, in trying to settle the question of the naturalness of collective intentionality with the traditional tools of conceptual analysis. In the case under consideration, I have brought this problem to the fore by contrasting two kinds of evidence and their implications on the issue at stake. The first is the conceptual evidence for the irreducibility of collective to

individual intentional states, whereas the second is the factual evidence of their instantiation at some biological level. Since for a naturalist only science – the body of most highly confirmed and reliable theories as for explanatory and predictive power at a given time – provides a reliable answer to what there is in nature, the question of the naturalness of collective intentionality asks for an empirically-grounded response. A scientific inquiry into the foundations of collective intentionality must be able to address proximate and ultimate questions concerning the biology of collective intentional behavior. The pursuit of proximate causes purports to answer 'how-questions', concerning the way a trait operates in an organism; in contrast, ultimate causes can be succinctly described as those concerned with 'why-questions', that is, why the trait came to be in the organism.

These considerations find a systematic and influential synthesis in Tomasello's work on the roots of sociality in the cognitive sciences, which I have presented in the second part of the thesis as the most advanced theory of collective intentionality naturalized that is currently available. A theory of collective intentionality naturalized treats the problem of collective intentionality as a problem of empirical social ontology, invoking fundamental continuities with the content and methods of science. Tomasello's theory is the ideal candidate for three reasons. The first is that he endorses an intentionalist view of the structure of social reality – a recurrent theme of the analyses of collective intentionality in social theory and philosophy, according to which human interaction is mentally mediated by expectations about each other's behavior and underpinned by mind-reading abilities. The second is the breadth of Tomasello's highly integrated line of inquiry, which deals with phenomena of sociality in developmental social cognition, primate cognition and language acquisition. The third reason consists in a novel formulation of the nature-nurture debate on the origins of the mind.

All these elements coalesce in the Shared Intentionality Hypothesis. My concern in chapter 5 was to elucidate the evolution of Tomasello's own thinking on the mechanisms of shared intentionality over the last fifteen years. The Hypothesis has gone through various phases of elaboration, summarized in a two-stage sequence, which was suggested by the results of a large

battery of experiments on the ontogeny and phylogeny of social cognition. It is by studying joint attention in particular, the phenomenon by which one-year olds are capable of sharing attention with their caregivers because they 'see' the others as subjects of intentional action, that Tomasello and his collaborators have discovered that chimps, too, are capable of mind-reading though on a lesser scale than humans. This body of empirical evidence gave important new insights on the nature of joint attention, which are partially at odds with the extant version of the Shared Intentionality Hypothesis. So a novel version of the Hypothesis has followed up, according to which humans are capable not only to understand but, most significantly, to share mental states with others based on species-unique social proclivities and intention-attribution skills.

The Shared Intentionality Hypothesis has applications in a number of fields dealing with phenomena of sociality, including the debate over the origins of communication. There are very interesting connections between Tomasello's view of the irreducibility of shared intentionality and the 'problem of reference', the problem of how any two persons can know that they mean the same thing in communication. Psychologists have not fully appreciated the potential of these connections, but neither have collective intentionality philosophers worked out the consequences of the Hypothesis enough to realize that it can provide an answer to the fundamental irreducibility question. This failure is partly due to the fact that, since its inception in developmental social cognition, notably in discussions of the nature of joint attention, the Shared Intentionality Hypothesis has been attacked for a number of reasons which are largely unmotivated. The main goal that I have pursued in chapter 6 is to clear the field from possible misinterpretations of Tomasello's philosophical position, so as to strengthen the similarity between his approach and that of most collective intentionality theorists, in particular Searle.

More specifically, I have shown that there is a profound conceptual difference between Tomasello's pragmatist theory of reference and the semantic interpretation his externalist critics have attributed to him. For Tomasello the problem of reference is not a logical problem, but one that can only be settled in practice by exploring its contextual, psychological features. The formulation of the Hypothesis that I favor is based on Bruner's pioneer work on the pragmatics of

reference, and it is justified by the fact that Tomasello comes from a different background from most semantic externalists. The field of interaction, according to pragmatists, offers the objective anchorage that the parties to a communicative exchange require to establish reference. The difference between semantic and pragmatist theories would then enlighten Tomasello's particular view of shared intentionality as a common psychological 'space' of mutual understanding.

The work of Tomasello and colleagues fills a significant gap in the current literature on collective intentionality: it turns the irreducibility thesis into a scientific theory of human development, making it subject to empirical check eventually. For this reason, it is unsurprising that this program has attracted a great deal of interest among philosophers of social science interested in the naturalization of collective intentionality. However, in spite of its potential in solving issues of social ontology, Tomasello's account faces some challenges as well. To explain why, let us consider again the claim that shared intentionality is a precondition of reference-acquisition. At the end of chapter 6 I have insisted on one feature of the Shared Intentionality Hypothesis: the assertion that mutual understanding of reference requires individuals to construe the action scene as one of shared intentionality, with the effect that collective intentionality figures among the causal conditions of first-person singular mental states, is backed up by a large body of data. For the time being, this is the only consideration that can be drawn from Tomasello's theory.

In fact, the Shared Intentionality Hypothesis does not provide any structured, or even partially worked-out, theory of the irreducibility of collective to individual intentionality. Put in different terms, any conclusion about the causal influence of group-thinking in setting the conditions of possibility of individual intentional states – or, as Searle would put it, the causal 'priority' of collective over individual intentionality – remains a speculation in need of further elaboration. The next task is thus to spell out the naturalization process in detail – how is it that individual minds are causally influenced by collective intentionality in achieving the full sense of mutual understanding observed in episodes of joint attention. Such account is lacking at present, and it therefore suggests a line of inquiry for future research in the process of sharing that brings about the state of collective intentionality.

Another challenge concerns the status of the claims based on the research in scientific psychology. Tomasello and his fellow researchers administer behavioral tests to humans and non-humans in laboratory settings. This line of inquiry does not contemplate any reference to the work of neuroscientists on the biological underpinnings of group-thinking. However, as I have specified in chapter 1, such approach might be seen as unsatisfactory in certain philosophical circles, notably among physicalists who identify the 'ideal' candidate for scientific reduction in a theory that reduces the mental to its most elementary brain processes. To some, the status of Tomasello's claims might look 'less' naturalistic than that of claims about neurons and cellular activities in the brain. Although the most appropriate way for setting out the conditions for the naturalization of the mind is an interesting problem on its own, I have treated it as tangential to the main concern of this thesis.

On a pluralist conception of naturalism, the problem is not where scientific psychology stands relative to the mind-body problem, but what kind of conceptual framework informs the design of experimental settings as well as the interpretation of the relevant findings. In fact, there appears to be remarkable similarities in the way cognitive scientists and philosophers talk about the same subject. In describing the cognitive and motivational underpinnings of shared attention, Tomasello resorts to the widespread construal of attention as a phenomenon of perceptual intentionality. This is an 'intentionalist' characterization, which is certainly closer to philosophers' talk about the nature of intentionality than scientists' construal of the neural basis of attention. If we want to preserve the significance of Tomasello's results to the problem of giving a scientific theory of collective intentionality, we ought to cash out the implications of his research in ways that increase our understanding of the conditions for the naturalization of collective intentionality. In brief, we should look at the data more as scientists than philosophers. And this brings us back to the question of how to make a step forward in exploiting science-philosophy continuities, for example by working out the aspects of causation that promise to illuminate the irreducibility problem.

A cautious approach would suggest considering also the other naturalistic approaches to collective intentionality that my analysis touched upon only in a cursory way. The interest that

scientists have shown in the past few years for the foundations of collective intentionality testifies to the interdisciplinary nature of the subject, and it reminds us of the continuity between philosophy and science in treating issues of social ontology. In fact, although significant results have been achieved in social neuroscience and in social psychology, philosophers have devoted little if no attention to scrutinize them in current discussions of collective intentional phenomena. As a consequence, the debate remains open on how to integrate novel results into a more systematic theory. In this thesis I have contributed to this goal by mapping some of the issues facing the naturalization of collective intentionality, as well as by indicating future directions of research in empirical social ontology.

Bibliography

Adenzato, M., Becchio, C., Bertone C. and Tuomela, R. (2005). 'Neural correlates underlying action-intention and aim-intention', in Bara, B.G., Barsalou, L. and Bucciarelli, M. (eds.), *Proceedings of the Twenty-Seventh Conference of the Cognitive Science Society*, Mahwah, NJ, Lawrence Erlbaum Associates.

Alexander, J.M. (2008), Evolutionary Game Theory, *Stanford Encyclopedia of Philosophy* [Online], Available: http://plato.stanford.edu/entries/game-evolutionary/ [July 19, 2009].

Armstrong, D.M. (1968), *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.

Astington, J. W. (2006), 'The developmental interdependence of theory of mind and language', in Enfield, N.J. and Levinson, S. (eds.) *The roots of human sociality: Culture, cognition, and human interaction*, Oxford, UK: Berg.

Bacharach, M. (2006), *Beyond the Individual Choice*, Princeton: Princeton University Press.

Bardsley, N. (2007), 'On Collective Intentions: Collective Action in Economics and Philosophy', *Synthese*, 157, pp. 141-159.

Barwise, K. J. and Moss, L. (1996), *Vicious Circles. On the Mathematics of Non-Wellfounded Phenomena*, Stanford: CSLI publications.

Bates, E., Camaioni. L. and Volterra, V. (1975), 'The Acquisition of Performatives Prior to Speech', *Merrill-Palmer Quarterly*, 21, pp. 205-226.

Bird, A. and Tobin, E. (2008), Natural Kinds, *Stanford Encyclopedia of Philosophy* [Online], Available: http://plato.stanford.edu/entries/natural-kinds/ [September 17, 2008].

Boden, M. (2006), *Mind As Machine*, Oxford: Oxford University Press.

Botterill, G. and Carruthers, P. (1999), *The Philosophy of Psychology*, Cambridge: Cambridge University Press.

Brandom, R.B. (1994), *Making It Explicit: Reasoning, Representing, and Discursive Commitment*, Cambridge, MA: Harvard University Press.

Brandom, R.B. (2000), *Articulating Reasons*, Cambridge, MA: Harvard University Press.

Bratman, M. (1992), 'Shared Cooperative Activity', *Philosophical Review*, 101, pp. 327-341.

Bratman, M. (1993), 'Shared Intention', *Ethics*, 104, pp. 97-113.

Bratman, M. (2009), 'Shared Agency', in Mantzavinos, C. (ed.) *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice*, Cambridge: Cambridge University Press.

Brentano, F. (1874/1973), *Psychology from an Empirical Standpoint*, London: Routledge.

Brinck, I. (2001), 'Attention and the Evolution of Intentional Communication', *Pragmatics and Cognition*, 9, pp. 255-272.

Brinck, I. (2004), 'Joint Attention, Triangulation and Radical Interpretation: A Problem and Its Solution', *Dialectica*, 58, pp. 179-205.

Bruner, J. (1977), 'Early Social Interaction and Language Acquisition', in Schaffer H.R. (ed.) *Studies in Mother-Infant Interaction*, London: Academic Press.

Bruner, J. (1983), *Child's Talk: Learning to Use Language*, New York: Norton.

Bruner, J. (1995), 'From Joint Attention to the Meeting of Minds: An Introduction', in Moore, C. and Dunham, P. (eds.) *Joint Attention: Its Origins and Role in Development*, Hillsdale, NJ: Erlbaum.

Bruner, J. (1998), 'Routes to Reference', *Pragmatics and Cognition*, 6, pp. 209-227.

Burkhardt, R.W. (2007), 'Niko Tinbergen: The Ethologist as Field Naturalist', *Biological Theory*, 2, pp. 87-90.

Byrne, A. (2001), 'Intentionalism Defended', *Philosophical Review*, 110, pp. 199-240.

Campbell, J. (2002), *Reference and Consciousness*, Oxford: Oxford University Press.

Campbell J. (2004), 'Reference as Attention', *Philosophical Studies*, 120, pp. 265-276.

Campbell, J. (2005), 'Joint Attention and Common Knowledge', in Eilan, N., Hoerl, C., McCormack, T. and Roessler, J. (eds.) *Joint Attention: Communication and Other Minds, Problems in Philosophy and Psychology*, Oxford: Oxford University Press.

Carey, S. (2009), *The Origins of Concepts*, Oxford: Oxford University Press.

Carpendale, J.I.M. and Lewis, C. (2006), *How Children Develop Social Understanding*, Oxford: Blackwell.

Carpenter, M. (2009), 'Just How Joint is Joint Action in Infancy?', *Topics in Cognitive Science*, 1, pp. 380-392.

Carpenter, M., Nagell, K and Tomasello, M. (1998), 'Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age', *Monographs of the Society for Research in Child Development*, 63, no. 255.

Carruthers, P. and Botterill, G. (1998), *The Philosophy of Psychology*, Cambridge: Cambridge University Press.

Churchland, P. (1981), 'Eliminative Materialism and the Propositional Attitudes', *Journal of Philosophy*, 78, pp. 67-90.

Clark, H. (1996), *Uses of Language*, Cambridge: Cambridge University Press.

Crane, T. (2001), *The Mechanical Mind*, 2nd Edition, London: Routledge.

Crane, T. (2003), *Elements of Mind*, Oxford: Oxford University Press.

Crane, T. (2007), 'Intentionalism', in Beckermann, A. and McLaughlin, B. (eds.) The *Oxford Handbook to the Philosophy of Mind*, Oxford: OUP.

Davidson, D. (1970), 'Events and Particulars.' *Nous*, 4, pp. 25-32.

Davies, M. (1998), 'Language, thought, and the language of thought (Aunty's own argument revisited)', in Carruthers, P. and Boucher, J. (eds.), *Language and Thought*, Cambridge: Cambridge University Press.

De Caro, M. and Macarthur, D. (eds.) (2004), *Naturalism in Question*, Cambridge, MA: Harvard University Press.

Dennett, D. (1978), 'Beliefs about Beliefs', *Behavioral and Brain Sciences*, 1, pp. 568-70.

Dennett, D. and Haugeland, J. (1987), 'Intentionality', in Gregory R. L. (ed.), *The Oxford Companion to the Mind*, Oxford: Oxford University Press.

Devitt, M. and Sterelny, K. (1999), *Language and Reality*, 2nd edition, Oxford: Blackwell.

Dummett, M. (1973), *Frege. Philosophy of Language*, Cambridge, MA: Harvard University Press.

Dupre', J. (2003), *Darwin's Legacy. What Evolution Means Today*, Oxford: Oxford University Press.

Durkheim, E. (1963), *Sociology and Philosophy*, Glencoe, IL: Free Press.

Eilan, N. (1998), 'Perceptual Intentionality, Attention and Consciousness', in O'Hear A. (ed.) *Current Issues in Philosophy of Mind*, Royal Institute of Philosophy Supplement 43, Cambridge: Cambridge University Press.

Eilan, N. (2005), 'Joint Attention, Communication, and Mind', in Eilan, N., Hoerl, C., McCormack, T. and Roessler, J. (eds.) *Joint Attention: Communication and Other Minds, Problems in Philosophy and Psychology*, Oxford: Oxford University Press.

Eilan, N., Hoerl, C., McCormack, T. and Roessler, J. (eds.) (2005), *Joint Attention: Communication and Other Minds, Problems in Philosophy and Psychology*, Oxford: Oxford University Press.

Enfield, N. J. and Levinson, S. (eds.) (2006), *Roots of human sociality: Culture, cognition and interaction*, Oxford: Berg

Evans, G. (1982), *The Varieties of Reference*, Oxford: Oxford University Press.

Fodor, J. (1974), 'Special Sciences: Or the Disunity of Science as a Working Hypothesis', *Synthese*, 28, pp. 97-115.

Fodor**,** J. (1983), *The Modularity of Mind*, Cambridge, MA: MIT Press.

Fodor, J. (2008), *LOT2. The Language of Thought Revisited*, New York: Oxford University Press.

Frege, G. (1892/1980), 'Über Sinn und Bedeutung', in *Zeitschrift für Philosophie und philosophische Kritik*, 100, pp: 25-50. Translated as 'On Sense and Reference' in by Geach, P. and Black, M. (eds. and trans.) in *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell, third edition.

Gardner, H. (1985), *The Mind's New Science. A History of the Cognitive Revolution*, New York: Basic Books.

Garzon, F.C. (2008), 'Towards a General Theory of Antirepresentationalism', *British Journal for the Philosophy of Science*, 59, pp. 259-292.

Gibson, E.J. and Rader, N. (1979), 'Attention: The perceiver as Performer', in Hale, G. and Lewis, M. (eds.), *Attention and cognitive development*, New York: Plenum Press.

Gilbert, M. (1989), *On Social Facts*, New York: Routledge.

Gilbert, M. (1990), 'Walking Together. A Paradigmatic Social Phenomenon', *Midwest Studies in Philosophy*, 15, pp. 1-14.

Gilbert, M. (2006), 'Rationality in Collective Action', *Philosophy of the Social Sciences*, 36, pp. 3-17.

Gold, N. and Sugden, R. (2007), 'Collective Intentions and Team Agency', *Journal of Philosophy*, 104, pp. 109-137.

Goldman, A.I. (2006), *Simulating Minds*, New York: Oxford University Press.

Grayling, A.C. (1997), *An Introduction to Philosophical Logic*, 3rd edition, Oxford: Blackwell.

Grice, P. (1957), 'Meaning', *The Philosophical Review*, 66, pp. 377-88.

Grice, P. (1969), 'Utterer's Meaning and Intentions', *The Philosophical Review*, 68, pp. 147-77.

Grice, P. (1975), 'Logic and Conversation', in Davidson, D. and Harman, G. (eds.) *The Logic of Grammar*, Encino, CA: Dickenson.

Griffiths, P. (2008), 'Ethology, Sociobiology, Evolutionary Psychology', in Sarkar, S. and Plutyinski, A. (eds.) *Blackwell's Companion to Philosophy of Biology*, Oxford: Blackwell.

Gross, S. (2010), Review of *Origins of Human Communication* (by M. Tomasello), *Mind & Language*, 25, pp. 237-46.

Guala, F. (2007), 'The Philosophy of Social Science: Metaphysical *and* Empirical', *Philosophy Compass*, 2, pp. 954-980.

Guala, F. (2007), Review of The Grammar of Society (by C. Bicchieri), *British Journal for the Philosophy of Science*, 58, pp. 613-618.

Haslam, S. A. (2004), *Psychology in Organizations: The Social Identity Approach*, 2nd edition, Thousand Oaks, CA: Sage Publications.

Hornsby, J. (1997), 'Collectives and Intentionality', *Philosophy and Phenomenological Research*, 57, pp. 429–434.

Jackson, F. (1998), *From Metaphysics to Ethics. A Defense of Conceptual Analysis*, Oxford: Oxford University Press.

James, W. (1890), *The Principles of Psychology*, Cambridge, MA: Harvard University Press.

Kim, J. (2006), *Philosophy of Mind*, 2nd edition, New York: Westview Press.

Kita, S (2003) (ed.), *Pointing. Where Language, Culture, and Cognition Meet*, Hillsdale NJ: Erlbaum.

Kripke, S. (1980), *Naming and Necessity*, Cambridge, MA: Harvard University Press.

Kripke, S. (1982), *Wittgenstein on Rules and Private Language: an Elementary Exposition*, Cambridge, MA: Harvard University Press.

Laland, K. and Brown, G. (2002), *Sense and Nonsense: Evolutionary Perspectives on Human Behavior*, Oxford: Oxford University Press.

Levinson, S. (1983), *Pragmatics*, Cambridge: Cambridge University Press.

Levinson, S. (1995), 'Interactional biases in human thinking', in Goody E. N. (ed.) *Social intelligence and interaction*, Cambridge: Cambridge University Press.

Levinson, S. (2006), 'On the human 'interaction engine'', in Enfield, N. J. and Levinson, S. (eds.) *Roots of human sociality: Culture, cognition and interaction*, Oxford: Berg.

Levinson, S. (2006), 'Cognition at the heart of human interaction', *Discourse Studies*, 8, pp. 85-93.

Lewis, D. (1966), 'An Argument for the Identity Theory, *Journal of Philosophy*, 63, pp 17-25.

Lewis, D. (1969), *Convention*, Cambridge, MA: Harvard University Press.

Liebal, K., Behne, T., Carpenter, M. and Tomasello, M. (2009), 'Infants Use Shared Experience to Interpret Pointing Gestures', *Developmental Science*, 12, pp. 264-271.

List, C. and Pettit, P. (2006), 'Group Agency and Supervenience', *Southern Journal of Philosophy*, XLIV (Spindel Supplement), pp. 85-105.

Ludwig, K. (2007), 'Foundations of Social Reality in Collective Intentional Behavior', in Tsohatzidis, S. (ed.) *Intentional Acts and Institutional Facts: Essays on John Searle's Social Ontology*, Dordrecht: Springer, 2007.

Lukes, S. (1973), *Individualism*, New York: Harper & Row.

McGinn, C. (1984), *Wittgenstein on Meaning*, Oxford: Blackwell.

McGinn, C. (2002), *The Making of a Philosopher: My Journey Through Twentieth-Century Philosophy*, London: Simon&Schuster.

Mallon, R. (2008), Naturalistic Approaches to Social Construction, *Stanford Encyclopedia of Philosophy* [Online], Available: http://plato.stanford.edu/entries/social-construction-naturalistic/ [November 10, 2008].

Margolis, E. and Lawrence, S. (eds.) (1999), *Concepts: Core Readings*, Cambridge, MA: MIT Press.

Margolis, E. and Lawrence, S. (2007), 'The Ontology of Concepts: Abstract Objects or Mental Representations?', *Nous*, 41, pp. 561-593.

Mayr, E. (1961), 'Cause and effect in biology: Kinds of causes, predictability, and teleology are viewed by a practicing biologist', *Science*, 134, pp. 1501–1506.

Meijers, A.W.M. (2003), 'Beyond Searle's Individualism', *The American Journal of Economics and Sociology*, pp. 167-183.

Mill, J. S. (1867), *A System of Logic*, London: Longmans.

Mole, C. (2010), 'Attention', in Margolis, E., Samuels, R. and Stich, S. (eds.) *Oxford Handbook of Philosophy and Cognitive Science*, Oxford: Oxford University Press.

Moll, H., Carpenter, M., and Tomasello, M. (2007), 'Fourteen-month-olds Know What Others Experience Only in Joint Engagement', *Developmental Science*, 10, pp. 826-835.

Moll, H., Richter, N., Carpenter, M. and Tomasello, M. (2008), 'Fourteen-month-olds Know What 'We' have Shared in a Special Way', *Infancy*, 13, pp. 90-101.

Moore, C., and Dunham, P. J. (eds.) (1995), *Joint Attention: Its Origins and Role in Development*, Hillsdale, NJ: Erlbaum.

Nagel, E. (1961), *The Structure of Science: Problems in the Logic of Scientific Discovery*, London: Routledge and Kegan Paul.

O'Neill, J. (1973), *Modes of Individualism and Collectivism*, London: Heinemann.

Pacherie, E. and Dokic, J. (2006), 'From mirror neurons to joint actions', *Journal of Cognitive Systems Research*, 7, pp. 101-112.

Pacherie, E. (2007), 'Is collective intentionality really primitive?', in Beaney, M., Penco, C. and Vignolo, M. (eds.) *Mental processes: representing and inferring*, Cambridge: Cambridge Scholars Press.

Papineau, D. (2007), Naturalism, *Stanford Encyclopedia of Philosophy* [Online], Available: http://plato.stanford.edu/entries/naturalism/ [February 22, 2007].

Peacocke, C. (2005), 'Join Attention, Its Nature, Reflexivity and Relation to Common Knowledge', in Eilan, N., Hoerl, C., McCormack, T. and Roessler, J. (eds.) *Joint Attention: Communication and Other Minds, Problems in Philosophy and Psychology*, Oxford: Oxford University Press.

Pettit, P. (1993), *The Common Mind*, Oxford: Oxford University Press.

Premack, D. and Woodruff, G. (1978), 'Does the chimpanzee have a theory of mind?', *Behavioral and Brain Sciences*, 4, pp. 515-526.

Putnam, H. (1975), 'The Meaning of Meaning', in K. Gunderson (ed.), *Language, Mind, and Knowledge*, Minneapolis: University of Minnesota Press.

Quine, W.V.O. (1960), *Word and Object*, Cambridge, MA: MIT Press.

Quine, W.V.O. (1974), *Roots of Reference*, New York: Columbia University Press.

Racine, T.P. and Carpendale, J.I.M. (2007), 'The Role of Shared Practice in Joint Attention', *British Journal of Developmental Psychology*, 25, pp. 3-25.

Rakoczy, H. and Tomasello, M. (2007), 'The Ontogeny of Social Ontology: Steps Towards Intentionality and Status Functions', in Tsohatzidis S.L. (ed.), *Intentional Acts and Institutional Facts*, Dordrecht: Springer.

Reimer, M. (2009), Reference, *Stanford Encyclopedia of Philosophy* [Online], Available: http://plato.stanford.edu/entries/reference/ [May 20, 2009].

Rey, G. (1983), 'Concepts and Stereotypes', *Cognition*, 15, pp. 237-262.

Rilling, J.K. (2008a), 'The Neurobiology of Social Decision-Making', *Current Opinion in Neurobiology*, 18, pp. 159-65.

Rilling, J.K. (2008b), 'Social cognitive neural networks during in-group and out-group interactions', *Neuroimage*, 41, pp. 1447-1461.

Roessler, J. (2005), 'Joint Attention and the Problem of Other Minds', in Eilan, N., Hoerl, C., McCormack, T. and Roessler, J. (eds.) *Joint Attention: Communication and Other Minds, Problems in Philosophy and Psychology*, Oxford: Oxford University Press.

Rosenberg, A. (1983), 'Fitness', *Journal of Philosophy*, 80, pp. 457-474.

Rosenberg, A. (1988), 'Is the Theory of Natural Selection Really a Statistical Theory?', *Canadian Journal of Philosophy*, 14, pp. 187-206.

Rosenberg, A. and Bouchard, F. (2008), Fitness, *Stanford Encyclopedia of Philosophy* [Online], Available: http://plato.stanford.edu/entries/fitness/ [April 17, 2008].

Saaristo, A. (2006), 'There is No Escape from Philosophy: Collective Intentionality and Empirical Social Science', *Philosophy of the Social Sciences*, 36, pp. 40-66.

Saaristo, A. (2007), *Social Ontology and Agency: Methodological Holism Naturalized*, unpublished Ph.D. thesis, London School of Economics and Political Science (University of London).

Saaristo, A. (2008), 'On the Ontology of Collective Intentionality: A Constructivist Perspective', in Psarros, N., Schmid, H.B. and Schulte-Ostermann, K. (eds.) *Concepts of Sharedness*, Frankfurt: Ontos Verlag.

Samuels, R. (2002), 'Nativism in Cognitive Science', *Mind and Language*, 17, pp. 233-265.

Sawyer, K. (2001), 'Emergence in Sociology: Contemporary Philosophy of Mind and Some Implications for Sociological Theory', *American Journal of Sociology*, 107, pp. 551-585.

Scaife, M. and Bruner, J. (1975), 'The Capacity for Joint Visual Attention in the Infant', *Nature*, 253, pp. 265-266.

Schelling, T. (1960), *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.

Schiffer, S. (1972), *Meaning*. Oxford: Clarendon Press.

Schmid, H.B. (2003), 'Can Brains in Vats Think as a Team?', *Philosophical Explorations* 6, pp. 201-218.

Schmid, H.B. (2009), *Plural Action. Essays in Philosophy and Social Science*, Dordrecht: Springer.

Searle, J.R. (1969), *Speech Acts*, Cambridge: Cambridge University Press.

Searle, J.R. (1978), 'Literal Meaning', *Erkenntniss*, 13, pp. 207-224.

Searle, J.R. (1979), *Expression and Meaning*, Cambridge: Cambridge University Press.

Searle, J.R. (1983), *Intentionality: An Essay on the Philosophy of Mind*, Cambridge: Cambridge University Press.

Searle, J.R. (1990), ‘Collective Intentions and Actions’, in Searle J.R. (2002) *Consciousness and Language*.

Searle, J.R. (1992), *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

Searle, J.R. (1995), *The Construction of Social Reality*. New York: Free Press.

Searle, J.R. (2002), *Consciousness and Language*, Cambridge: Cambridge University Press.

Searle, J.R. (2004), *Mind. A Brief Introduction*, New York: Oxford University Press.

Searle, J.R. (2007), *Freedom and Neurobiology*, New York: Columbia University Press.

Searle, J.R. (2010), *Making the Social World*, Oxford: Oxford University Press.

Seemann, A. (2007), ‘Collective Knowledge and the ‘'We'-Perspective’, *Social Epistemology*, 21, pp. 217 - 230.

Sellars, W. (1954/2007), ‘Some Reflections on Language Games’, in Sharp, K. and Brandom, R.B. (eds.) *In the Space of Reasons*, Cambridge, MA: Harvard University Press.

Sellars, W. (1956), ‘Empiricism and The Philosophy of Mind’, in Feigl, H. and Scriven, M. (eds.) *Minnesota Studies in the Philosophy of Science*, 1, Minneapolis: Minnesota University Press.

Sellars, W. (1963), ‘Imperatives, Intentions, and the Logic of 'Ought'’, in Castañeda, H.N and Nakhnikian, G. (eds.) *Morality and the Language of Conduct*, Detroit: Wayne State University Press.

Sellars, W. (1967/2007), ‘Some Reflections on Thoughts and Things’, in Sharp, K. and Brandom, R.B. (eds.) *In the Space of Reasons*, Cambridge, MA: Harvard University Press.

Sellars, W. (1969/2007), ‘Language as Thought and as Communication’, in Sharp, K. and Brandom, R.B. (eds.) *In the Space of Reasons*, Cambridge, MA: Harvard University Press.

Sellars, W. (1974/2007), ‘Meaning as Functional Classification’, in Sharp, K. and Brandom, R.B. (eds.) *In the Space of Reasons*, Cambridge, MA: Harvard University Press.

Sellars, W. (1981/2007), ‘Mental Events’, in Sharp, K. and Brandom, R.B. (eds.) *In the Space of Reasons*, Cambridge, MA: Harvard University Press.

Sillari, G. (2008), ‘Common Knowledge and Convention’, *Topoi*, 27, pp. 29-40.

Skyrms, B. (2004), *The Stag Hunt and the Evolution of Social Structure*, Cambridge: Cambridge University Press.

Smart, J.J.C. (1959), 'Sensations and Brain Processes', *Philosophical Review*, 68, pp. 141-156.

Smith, B. (ed.) (2003), *John Searle*, Cambridge: Cambridge University Press.

Sperber, D. and Wilson, D. (1995), *Relevance. Communication and Cognition*, 2nd edition, Oxford: Blackwell.

Sperber, D. (1996), *Explaining culture: A naturalistic approach*, Oxford: Blackwell.

Stalnaker, R. (1973), 'Presuppositions', *Journal of Philosophical Logic*, 2, pp. 447-457.

Sterelny, K and Griffiths, P. (1999), *Sex and Death: An Introduction to the Philosophy of Biology*. Chicago, University of Chicago Press.

Stich, S. and Laurence, S. (1994), 'Intentionality and Naturalism', *Midwest Studies in Philosophy*, 19, pp. 159-182.

Stich, S. and Ravenscroft, I. (1994), 'What is Folk Psychology?', *Cognition*, 50, pp. 447-468.

Strawson, P. (1964), 'Intention and Convention in Speech Acts', *Philosophical Review*, 73, pp.439-460.

Striano, T., and Tomasello, M. (2001), 'Infant physical and social cognition', in Baltes P. (ed.), *International Encyclopedia of the Social and Behavioral Sciences: Human Developments*, Oxford: Elsevier.

Stroud, B. (1991), 'The Background of Thought', in Lepore, E. and van Gulick, R. (eds.) *John Searle and His Critics*, Oxford: Oxford University Press.

Sugden, R. (2003), 'The Logic of Team Reasoning', *Philosophical Explorations*, 6, pp. 165-181.

Susswein, N. and Racine, T.P. (2008), 'Sharing mental states: Causal and definitional issues in intersubjectivity', in Zlatev, J., Racine, T. P., Sinha C. and Itkonen, E. (eds.) *The shared mind: Perspectives on intersubjectivity*, Amsterdam: Benjamins.

Tajfel, H. (1970), 'Experiments in intergroup discrimination', *Scientific American*, 223, pp. 96-102.

Tinbergen, N. (1963), 'On Aims and Methods in Ethology', *Zeitschrift für Tierpsychologie*, 20, pp. 410-433.

Tollefsen, D. (2002a), 'Collective Intentionality and the Social Sciences', *Philosophy of the Social Sciences*, 32, pp. 25-50.

Tollefsen, D. (2002b), 'Organizations as True Believers', *Journal of Social Philosophy*, 33, pp. 395-411.

Tollefsen, D. (2004), Collective Intentionality, *Internet Encyclopedia of Philosophy* [Online], Available: http://www.iep.utm.edu/coll-int/ [August 4, 2004]

Tomasello, M. (1995), 'Joint Attention as Social Cognition', in Moore, C. and Dunham, P. (eds.), *Joint Attention: Its Origins and Role in Development*, Hillsdale, NJ: Erlbaum.

Tomasello, M. (1998), 'Reference: Intending that Others Jointly Attend', *Pragmatics and Cognition*, 6, pp. 229-243.

Tomasello, M. (1999), *The Cultural Origins of Human Cognition*, Cambridge, MA: Harvard University Press.

Tomasello, M. (2008a), *Origins of Human Communication*, Cambridge, MA: MIT Press.

Tomasello, M. (2008b), 'How are humans unique?', New York Times Magazine, May 25, 2008.

Tomasello, M. (2009), *Why We Cooperate*, Cambridge, MA: MIT Press.

Tomasello, M. and Carpenter, M. (2005), 'The emergence of social cognition in three young Chimpanzees, *Monographs of the Society for Research in Child Development*, 70, no. 279.

Tomasello, M., Carpenter, M., Call, J., Behne, T. and Moll, H. (2005), 'Understanding and Sharing Intentions: The Origins of Cultural Cognition', *Behavioral and Brain Sciences*, 28, pp. 675-735.

Tomasello, M., Carpenter, M., and Lizskowski, U. (2007), 'A new look at infant pointing', *Child Development*, 78, pp. 705-22.

Tomasello, M., Kruger, A., and Ratner, H. (1993), 'Cultural Learning', *Behavioral and Brain Sciences*, 16, pp. 495-552.

Tomasello, M. and Rakoczy, H. (2003), 'What Makes Human Cognition Unique? From Individual to Shared to Collective Intentionality', *Mind and Language*, 18, pp. 121-147.

Trevarthen, C. (1979), 'Communication and Cooperation in Early Infancy: A Description of Primary Intersubjectivity', in Bullowa M. (ed.) *Before Speech: The Beginning of Interpersonal Communication*, Cambridge: Cambridge University Press.

Tuomela, R. (1984), *A Theory of Social Action*, Dordrecht: Reidel.

Tuomela, R. (1995), *The Importance of Us. A Study of Basic Social Notions*, Stanford, CA: Stanford University Press.

Tuomela, R. (2002), *The Philosophy of Social Practices: A Collective Acceptance View*, Cambridge: Cambridge University Press.

Tuomela, R. (2005), 'We-Intentions Revisited', *Philosophical Studies*, 125, pp. 327-369.

Tuomela, R. (2007), *The Philosophy of Sociality: The Shared Point of View*, Oxford: Oxford University Press.

Tuomela, R. and Miller, K. (1988), 'We-Intentions', *Philosophical Studies*, 53, pp. 367-389.

Turner, J. C. and Reynolds, K. J. (2001), 'The social identity perspective in intergroup relations: Theories, themes and controversies', in Brown, R. J. and Gaertner, S. (eds.) *Handbook of social psychology. Intergroup processes*, Oxford: Blackwell, pp. 133–152.

Velleman, J.D. (1997), 'How to Share an Intention', *Philosophy and Phenomenological Research*, 57, pp. 29 – 50.

Vromen, J.J. (2003), 'Collective Intentionality, Evolutionary Biology, and Social Reality, *Philosophical Explorations*, 6, pp. 251-264.

Vygotsky, L.S. (1978), Mind in Society: The Development of Higher Psychological Processes, Cambridge, MA: Harvard University Press.

Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L. and Bara, B.G. (2004), 'Understanding intentions in social interactions: The role of the anterior paracingulate cortex', *Journal of Cognitive Neuroscience*, 16, pp. 1854-1863.

Warneken, F., Chen. F., and Tomasello, M. (2006), 'Cooperative activities in young children and chimpanzees', *Child Development*, 77, pp. 640-663.

Williams, M. (1970), 'Deducing the Consequences of Evolution', *Journal of Theoretical Biology*, 29, pp. 343-385.

Wimmer, H. and Perner, J. (1983), 'Beliefs about Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception', *Cognition*, 13, pp. 103-128.

Winch, P. (1958), *The Idea of a Social Science*, London: Routledge.

Wittgenstein, L. (1953/2001), *Philosophical Investigations*, 3rd edition, Oxford: Blackwell.

Wittgenstein, L. (1983), *Remarks on the Foundations of Mathematics*, revised edition, Cambridge, MA: MIT Press.

Wright, C. (2007), 'Rule-Following without Reasons: Wittgenstein's Quietism and the Constitutive Question', *Ratio*, 20, pp. 481–502.