universität wien

Masterarbeit

Titel der Masterarbeit

“Collective Intentionality and Coupled Systems”

Verfasser

Matthew Rachar

angestrebter akademischer Grad

Master of Arts (MA)

Wien, 2014

Studienkennzahl lt. Studienblatt:	A 066 941
Studienrichtung lt. Studienblatt:	Masterstudium Philosophie
Betreuer:	Univ.-Prof. Dr. Hans Bernhard Schmid

Acknowledgements

I owe a heavy debt of gratitude to many people for their assistance in the writing of this thesis. First among them, on the academic side, is my thesis advisor Hans Bernhard Schmid, who was not only supportive and understanding of the central ideas of the project, but also responsible in large part for whatever clarity they may have. Others who have made valuable contributions and deserve thanks here include: Michael Schmitz, Gerhard Thonhauser, Pekko Roman, Katharina Bernhard and all the members of the *Sozialphilosophie Master-und-Dissertantinseminar* and the Social Ontology Colloquium.

I would also like to express my gratitude to my parents, Lee Rachar and Irene Lanzinger, for a lifetime of unflinching support (and grammatical advice), and my girlfriend, Veronika Schreiber, for being a patient interlocutor when presented with the most crude form of the arguments and ideas contained herein.

Table of Contents

Introduction

Much of human life takes place within the context of cooperative and coordinated social interaction, which often results in the formation of a specific collection of agents who share a particular goal and take concrete steps to bring about some change in the world. In both our everyday speech and the technical language of our social sciences, we frequently represent these collections as if they were singular agents, understandable in much the same way as individuals. We regularly hear news stories or read articles about what a particular nation believes, what the Federal Reserve wants, or what a trade union intends. These ascriptions of intentional states to collections of individuals are a fundamental part of our understanding of the social world.

We use this mentalistic discourse to explain the flow of thought and action. Our understanding of the behaviour of others is loaded with this type of "folk" or "propositional attitude" psychology. This practice has shown a great deal of success in predicting, explaining and understanding everyday behaviour and has been developed into a significant literature on action and agency. What has received less critical attention is the question: How should we understand this language in relation to social groups? Recent work in experimental psychology has shown that we are more than willing to ascribe these concepts to groups in our everyday practices (Knobe & Prinz 2008; Sytsma and Machery 2009, 2010; Huebner 2008; Huebner et al. 2010), but our folk-psychological intuitions do not reveal natural kinds nor the casual-explanatory structure of the world. The more interesting questions are: What does it mean to ascribe beliefs, desires and other intentional states to entities such as teams or even nations and corporations? Should we understand these ascriptions as literal, and if so, is this an intellectually productive practice? Are these statements true, and if so, on the basis of which psychological facts? Or, should we dismiss them as a type of figurative or metaphorical, if useful, way of speaking that has no literal impact? The answers to these questions, and related questions about social groupings in general, are of foundational importance to our attempts to study social, political and economic phenomena, and they shape our conception of moral and legal notions such as rights and responsibilities in the group context. In the last few decades, a consequential body of literature concerning just this topic has been growing, comprised of philosophers who agree that our ascriptions of intentional states to groups cannot be interpreted as merely a figure of speech. Their work ranges from

relatively compact studies of small collections of people performing simple actions, such as going for a walk, painting a house, running a particular football play or playing a symphony (Bratman 1999, Searle 1990, Tuomela 2007, Gilbert 2001) to more wide-ranging attempts to analyze large, complex collective entities with enduring sets of values and organizational structures (Pettit & List 2011).

Central to the theories developed by these philosophers, and to their answers to the questions mentioned above, is the concept of collective intentionality, which they use to make sense of both the everyday interaction of individuals which lead two or more people to do something together, and as a building block in the construction of larger social phenomena.[1] 'Intentionality' is a technical term which extends beyond the common meaning of intention. Instead of referring to decisions and actions, it stands for the ability of the human mind to be aware of, or more colourfully, to reach out into, the world. You may believe that your train leaves at 4pm, desire food, and fear that you do not have enough time for both, but in each case your mind is directed towards something, and each state (belief, desire, fear) is *about* something. 'Intentionality' in this sense is simply the ability of mental states to have content.

Collective intentionality then refers to the capability of a plurality of minds to share a common directionality, whether it is toward a specific object, fact, goal, value or state of the world. A basic fact of human sociality is that we are able to share cognitive states (like belief), conative states (like intentions) and affective states (like emotions and sensations). The literature surrounding collective intentionality in the broad sense contains attempts to study this ability in a variety of forms. It therefore contains theories of joint attention, collective acceptance, shared emotion and shared belief, all topics which deserve significant attention in relation to the aims of this thesis, but the central, explicit concern here will be to develop an account of joint intention and action. The underlying conviction of this approach is that a thorough understanding of the constitution and mechanisms of small social groupings and their actions reveals some fundamental constituents of an accurate analysis of the broader social world. [2]

[1] See Searle (1995) for the central role of collective intentionality in his account of the ontology of the social world.
[2] For more on this approach see Gilbert (1990: 2; 2006: 97).

Collective intentionality is a concept which effectively works at the intersection of an analysis of social groups and mental phenomena. As a result, in addition to approaching the question from the standpoint of social groups, part of understanding our ascription of intentional states to groups is understanding what intentional states are and what types of entities can have them. The field of situated cognition, a relatively new movement in cognitive science with significant philosophical implications, has challenged many of our notions about intentional states and lead to several theses which may have a direct impact on how we conceive of many aspects of collective intentionality. Situated cognition has not developed as a unified movement and is therefore difficult to contain under one conceptual heading or set of positive views. However, as a general approach, situated cognition holds that we must pay close attention to the interaction of brain and world in which human cognition takes place. It places special emphasis on the way in which the close causal relation between neural activity in the brain, the rest of the human body and the surrounding environment supports intelligent behaviour and allows us to perform what we take to be cognitive activities.

The purpose of this thesis is to present an account of how the new ideas in philosophy of mind that have come about as a result of the development of the situated cognition movement may affect our understanding of collective intentionality and social groups. More specifically, I focus on the promise of the hypothesis of extended cognition (HEC) for an analysis of joint action. The extended mind thesis states that cognition is partly constituted by elements or processes that exist outside the boundary of the human organism. This thesis provides exciting possibilities for answering the question of group mental state ascriptions because, if it is correct, mental states, and processes, are no longer tied to individual brains, but instead, may extend into the body, artefacts in the world and potentially even other minds. The removal of this limitation to the nature of mental states uncovers the possibility of a collective entity which could be the bearer of mental states in a literal sense. In this thesis, I explore the possibility of such an account based on the existence of such an entity and such mental states.

The concept of a collective cognitive system which is a collection of two or more human beings that literally has mental states and processes separate from those of its individual members is an idea with many philosophical detractors and challenges. Even within the literature on collective intentionality, which is mostly comprised of arguments for the necessary ‘sharedness’ or

'togetherness' present in genuine cases of joint action, the majority of the prominent accounts maintain some type of individualism. At this point it must be made clear that I intend to defend a position which entails the conceptual and epistemological irreducibility of some collective cognitive phenomena, rather than the ontological irreducibility of a supra-individual entity. Interestingly, despite the varied nature of the theoretical projects informing this thesis, there are several strains in each body of literature that attempt to ascribe cognitive phenomena to groups based on the idea of non-aggregativity. I will attempt to build from this similarity when turning from a discussion of collective intentionality in joint action to discussions of the differing positions in extended cognition in order to link these two bodies of literature.

It should be noted at the beginning that the theses of this project are conditional. The HEC is a highly controversial proposition, so much so that a detailed account of the debate and an adequate defence of the HEC would constitute a project in its own right. Therefore, in order to address the stated aims of this thesis, many of the commitments of the HEC will necessarily go undefended. Nonetheless, if we are willing to accept a particular view of extended cognition, I will argue, it leads to an as yet unexamined area of overlap, which I believe may be fruitful to accounts of joint action.

The thesis is divided into three parts: joint action, extended cognition and plural subjects as collective cognitive systems. The first two parts serve mostly to introduce the topics and theorists that inform the discussion and arguments of the final part. Part 1 consists of two chapters. Chapter 1 introduces the concept of joint action, specifies which actions count as joint actions, locates joint action theory within the tradition of the causal theory of action, determines the constitutive question of joint action and categorizes potential joint action analyses. Chapter 2 surveys the literature on joint action. I first introduce the idea of the non-aggregativity of group mental states, which connects certain theories of collective intentionality to the theories of distributed cognition and group cognition discussed later, in relation to summative accounts of joint action. I then introduce the central tension which derives from the rejection of summative accounts and turn to three existing accounts of joint action. I focus first on theories which maintain strong individualistic and reductionist components, namely those of Michael Bratman and John Searle, and then turn to the collective, relational account put forward by Margaret Gilbert. In the course of this presentation, I consider the arguments brought by critics against

these views, and finish the chapter by attempting to generalize central components of these criticisms and clarify the aspects of each theory to which they apply. On the basis of this discussion, I maintain that an account of joint action in relational, non-reductionist terms holds the most promise for answering the constitutive question outlined in Chapter 1.

In part 2, I introduce situated cognition, outline the general conceptual terrain and follow the development of the 'first-wave' philosophical idea of extended mind into 'second-wave' concepts such as integration and complementary. I then outline the relation of these second-wave concepts to accounts of distributed and group cognition. Chapter 3 is concerned with elaborating the first-wave extended cognition ideas of active externalism, causal coupling and the parity principle and considering the entailments of these concepts. I also introduce Clark's view of cognitive agency, the Hypothesis of Organism-centered Cognition, and the motivation for this view. Chapter 4 tracks the development of second-wave extended cognition from elements of Clark's theory, deals with the concepts of complementarity, integration, manipulation and transformation, and considers that impact of this development for an account of cognitive agency. Chapter 5 consists of an introduction to the related body of literature growing around the idea of distributed cognition in cognitive anthropology, as represented by Edward Hutchins. This theoretical approach involves a radically different account of cognitive agency that is in some ways opened up by the development of second-wave extended cognition. I explore how the distributed cognition framework leads to the idea of group cognition and consider the empirical research which supports it. I also introduce the concept of non-aggregativity in the context of group cognition.

Part 3 then continues and constitutes in the main part the attempt to bring these two bodies of literature together. I begin Chapter 6 by reviewing the role of coupling in all branches of extended cognition and showing how the search for the "right kind of coupling" between individuals leads into the question of collective intentionality. I consider two attempts to apply Clark's account of extended cognition to collective intentional phenomena which diverge in two radically different directions. Tollefsen's account of collective cognitive systems applies Clark's coupling conditions to cases of extreme interconnection between individuals, while Lyre's application of Bratman's theory of joint intentions leads to a purely individualist account of interpersonal cognitive extension. I then argue that neither of these accounts adequately deals

with the origins of coupling in dynamic systems theory, and that they fail as accounts of many collectively intentional phenomena related to joint action. In the final chapter, I turn to what I take to be the most promising domain of overlap between extended cognition and collective intentionality, arguments for the irreducibility and relational nature of group cognitive phenomena based on the non-aggregativity of these phenomena. Specifically, I consider the role of various levels of coupling in the formation and maintenance of plural subjects, building on the accounts of cognitive integration and complementarity of second-wave cognition and the account of cognitive agency found in distributed cognition and group cognition. Finally, I attempt to clarify the theoretical commitments of each of these positions and the extent to which concepts drawn from each domain of literature truly overlap. I conclude that although there are some significant challenges involved in applying distributed and group cognition concepts to plural subjects, there are several aspects of the social psychological research based on dynamic coupling which support a normative, relational and non-reductionist account of joint action.

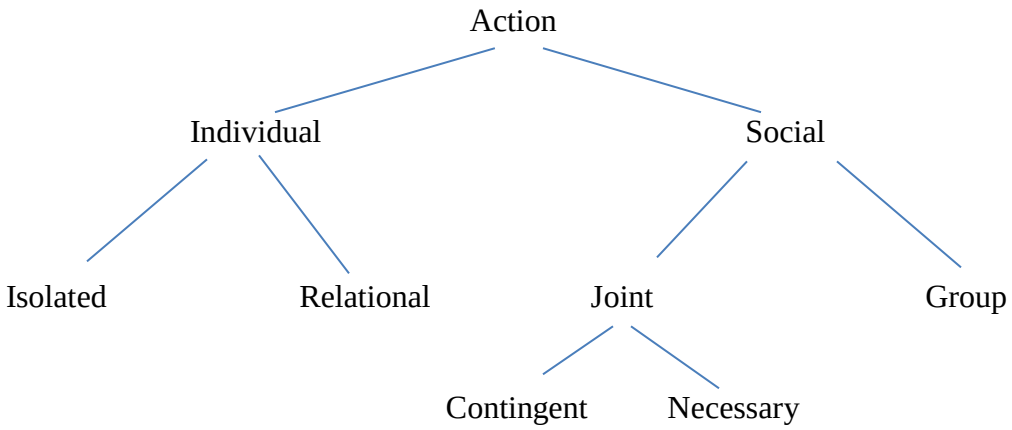
Part 1: Joint Action

Chapter 1 – Background and Taxonomy

§1 – Action Taxonomy

Some actions we can only do ourselves. Whether or not I raise my own arm does not, in the normal case, involve another person. If, however, I raise my arm in order to wave to a friend, the action does seem to involve another person, but in a derivative way. Other actions, however, require coordination with other people. I cannot sing a duet, dance the tango, carry a piano up the stairs, play in a symphony or run a football play by myself. Within this set of actions there is a further division. Some of these actions require the presence of other people because I lack the power to do them myself, although it is logically possible. I may not be able to carry a piano up the stairs, but another person may have the requisite physical strength. Other actions involve interaction by their very nature. An individual can play *in* a symphony, but he cannot *play* a symphony. Playing a symphony necessarily involves an orchestra. A final class of actions includes more than just the cooperation of a small group, but involves a large group of people with a complex institutional structure, for example, waging a war, voting on an amendment or electing a new university president. I propose that we divide these actions into two main classes, individual and social, each with their own two subclasses. Individual actions may be isolated or relational, while social actions may be joint or group actions. Joint actions and group actions are inherently social[3], but display several significant differences. In order to undertake a group action[4] there needs to be a social group with a history of practice, an organizational structure, a set of common values and procedures, and ultimately an established group that outlasts its specific members. This is not the case for joint actions[5]. People may come together spontaneously to achieve specific goals, perform some action, such as pushing a broken-down car, and then depart never to see one another again. Joint actions may also include social norms and normative practices, but they do not require the existence of that particular group after the goal has been achieved. We may then use this taxonomy to understand the division of actions underlying this thesis:

[3] This is not to say that individual actions do not have a social component, see Baier (1997), but this will not be the concern here.
[4] This definition of group action follows Pettit & List (2011).
[5] In the literature, joint actions are often referred to as shared, collective, or plural actions. To the extent that they overlap with the definition given in §1.3, I take these locutions to be equally acceptable.



This division between actions may be understood along the lines of the involvement of other individuals or institutions. Isolated actions are those actions which do not involve others, such as raising my arm or flossing my teeth. Relational actions involve others only derivatively, as when I raise my arm in order to wave to a friend. Joint actions may be contingent, as when we make a hollandaise sauce that I could have made on my own, or necessary, as when we sing a duet. What sets joint actions apart is that they involve a plurality of actors with a common goal. Finally, group actions are those actions which involve large, well-structured groups, with established institutions and social practices, and therefore involve not only many people, but also the accumulation of many actions over time. The difference between joint actions and group actions will remain one of degree in this thesis, as group actions will remain largely unanalyzed and joint actions, such as a team running a play, may contain elements also found in group actions.

My concern in this thesis is to analyze joint actions in this sense, as I hold that they are the fundamental form of social action and are required for the initial formation of the social groups required for institutional group action.[6] This chapter is concerned with further elaborating the concept of joint action and introducing the types and problems of joint action theories. I begin by explicating the difference between distributed collective outcomes and intentional collective behaviour. I then define joint action on the basis of a common goal and a collective sense of 'we'

[6] See Pettit and List (2011: 34-39). For further discussion on this point see (Schmid 2009: 22-23).

and finally I provide taxonomy of joint action theories, which introduce the problems and possibilities of analyzing collective intentional behaviour in terms of joint intentions.

§2 – Action Theory

In an analysis of joint action, such action must be separated from social behaviour that results in unintentional consequences, a concern which connects joint action theory to traditional analytic action theory. The constitutive question in action theory is what separates actions from mere events or happenings (Anscombe 2000, Davidson 1980). There is a significant difference between raising my arm due to a spasm and raising my arm to answer a question in class, despite the fact that the physical act is the same. In the second case the action reaches out into the world in order to effect some change, while the first is a purely reflexive act with no such goal. Action theory holds the conviction that the distinguishing feature between these two cases is that a genuine action is caused by the right kinds of antecedent mental events, such as beliefs, desires and intentions. The task is then to specify the mental events that serve as the proximate causes of action. We may here distinguish several features common to action theories that will help to frame our discussion of joint action. A standard idea in this debate is that events qualify as actions when they can be characterized as intentional under a particular description with relation to an agent. Intention here becomes the focal point in the attribution of the status of action to an event. Explicating the exact nature of the relationship between actions, intentions and agency in the singular case has created a vast literature. Glossing over the finer theoretical distinctions in this literature, we may hold that the standard idea requires that for there to be an action there must be four things:

1) An agent to whom that action can be attributed.
2) A complex of behavior, over which the agent has a certain degree of control.
3) A goal or state of the world that the agent is trying to bring about,
4) Some minimally rational standard connecting the chosen behavior to the desired state of the world.

Traditional action theory focuses primarily on cases of individual action, which leaves us with the question: How does this framework apply to actions involving groups of people? In the course of our interactions with others, we are able to bring about collective outcomes. In this situation, we are left with a problem, parallel to that of the individual case, about whether or not we should consider the outcome an action or merely an event.We do not intentionally cause global climate change together. It is a consequence of the sum of our individual, uncoordinated behaviour. For actions like carrying a piano we may do it together intentionally or unintentionally. I may believe that I am doing it on my own, unaware that you are helping, or alternatively, we may decide that we are going to carry the piano together. This second case implies that in addition to my doing something intentionally, *we* are doing something intentionally. This also holds for necessarily social actions. When we play a symphony it implies that there is something that we intend to do together. Necessarily joint actions require an explanation that tests the limits of individual action theory, although it may be founded on principles that adhere to the same basic tenets.

The questions that will concern us in this thesis center on the exact nature of the difference between the individual actions and the intentional states that support them, and joint actions and the intentional states that support them. If we apply the standard ideas of action theory straightforwardly to the collective case, saying that we do something together intentionally seems to entail, according to the requirements mentioned above, that there is some 'we' that can serve as the agent to which the action is attributable. This in turn implies that there is a collective entity, referred to by the use of the first-person plural that is the subject of the intention. In other words, there appears to be some group that is the bearer of a mental state, has a goal or state of the world that it has a pro-attitude towards, and displays a certain minimum standard of rationality. However, many philosophers[7] deny these claims and attempt to account for joint action based solely on concepts that reduce to individuals, or postulate some non-reductive we-states that nonetheless exist within individual minds. In chapter 2, I consider the arguments for the position that we cannot understand intentional group behaviour based solely on concepts that reduce to individuals and that we must maintain some of these implications of a straightforward application of action theory to group behaviour in a modified way.

[7] In this thesis these two positions are represented by Bratman (1999) and Searle (1990). Their theories are discussed in more detail in §2.3 and §2.4.

§3 – Joint Action Defined

A joint action can be more precisely defined as any action two or more agents purport to do together on the basis of a common goal. With its focus on a shared goal, this definition links contingently joint actions and necessarily joint actions.[8] Two people carrying a piano is a joint action if they share the same goal, just as an orchestra playing a symphony is only a joint action if all the participants share the same goal. To put this definition more formally: an action is a joint action if and only if it involves multiple participants who have a single goal. This definition allows for group actions that are also joint actions, if all the members of the group involved in the group act share the same goal, but this is not a necessary condition for all group actions.

Joint actions involve a type of social group which lasts as long as the action takes, and may be defined according to a particular usage of the word 'we'. For the purposes of this thesis, I label these groups 'collectives', or if the term 'groups' is used it is modified with some variation of the phrase 'according to common usage'. Let us imagine the following two scenarios. Suppose I am moving and need a couch in my new apartment. You live on the same floor and just bought a new couch. The drop-off spot for old furniture in our building is located between our apartments. You carry the couch to the drop-off area and leave it there. I see the couch before the garbage men come, pick it up and move it into my apartment. In some sense, the statement 'We moved the couch from your apartment to mine' is true. When we add up the individual actions, the result is that the couch moved from your apartment to mine. The action 'moving the couch' is predicated over the individuals. This means that the 'we' is used in a distributive sense. There is something you did and something I did, but nothing that we did *jointly*; there is nothing more than our contributing parts. What this example shows is that some actions involving more than one agent when each agent is acting on their own are only accidentally, and not intentionally, collective. As pointed out by Gilbert (2006: 145) in a different context, this sense of 'we' is often marked by the use of the terms 'We all…', 'Both of us', as in her example of an executive saying 'We were all inspired by your talk' to an inspirational speaker. She argues that this usage is equivalent to 'We were all personally' or 'We both personally' and thus refers only to a shared characteristic of separate individuals, such as when we say something along the lines of "We all have brown

[8] Starting with a concept of joint action that includes contingently joint actions has been criticized (Baier 1997), but I take the common goal condition as unproblematic and do not think that such a beginning biases the analysis.

eyes". In scenario two, we are not just neighbours, but friends. You agree to give me your old couch and we decide upon a time to move it together. You take one end, I take the other, and we move the couch into the apartment. In this case, there does not appear to be an implicit 'all' or 'both' in the statement 'We moved the couch from your apartment to mine'. The sense of 'we' used here is collective because the action is something that we accomplished *together*, and in common usage, *as a group*. This is the type of group required for the collective behaviour to be intentionally collective and therefore for the completion of a genuinely joint action.

It is also important to note that in the second case, there exists an action over and above our individual contributions that we accomplish together. As we see from the second case, in addition to my carrying the right side, and your carrying the left side, there is the action that we together accomplish, namely, the moving of the couch. Something is missing from the first case that is present in the second case. If I say that we moved the couch, in the collective 'we' sense, it implies that both of us moved the couch, but if I say that both of us moved the couch, it does not imply that we moved the couch together and intentionally. In opposition to the 'we both' or 'we all' of the distributive 'we', the collective reading of 'we' may be understood as equivalent to the common usage of 'we, *as a group*' and it is this 'as a group' that the distributive reading fails to capture. Joint action is then an event brought about by collective behaviour that is intentional under a particular description on the basis of a common goal.[9] The question joint action analyses attempt to answer then is: Exactly what does the common locution "as a group" mean in the case of action? In more technical language, what is the difference between distributive coordinated behaviour and collective intentional behaviour? Or, what distinguishes outcomes that result from a mere aggregation of individual acts from genuine joint action?

§4 – Taxonomy of Joint Action Accounts

As we saw in the discussion of traditional action theory, the determination of action as opposed to event turns on the definition of intention. If we maintain this premise, we may claim that a joint intention is simply the type of intention that we need to explain joint action, and therefore, the

[9] The expressions 'collective intentional behaviour' and 'joint action' are therefore equivalent, as defined here.

central concept in the distinction between correlated individual behaviour that produces some outcome and genuine goal-directed collective behaviour. Intentions are thought to be propositional attitudes relating a subject through a mode to a particular content. There are then three central questions surrounding the formulation of joint intention:

1. Who is the subject of a joint intention?
2. What is the mode of a joint intention?
3. What is the content of a joint intention?

With regards to the first question, there appear to be three possibilities. The first is that joint intention is had strictly by isolated individuals. The second is that the intention is had by interconnected individuals taken together and the third is that the collective itself has the intention. The difference between the first and second possibility is that the second possibility requires that the individuals stand in a particular relation to one another, while the first makes no such claim. The third possibility is the most controversial, in that it requires the ascription of intentional states to collectives.

The second question concerns the way in which joint intentions should be represented. Conceivably, joint intention may be written in the form 'I intend' or 'We intend', regardless of the subject of the intention. It is therefore possible to claim that an individual has an intention in the 'We-mode' or that a collective has an intention in the 'I-mode'.[10]

The final issue in the determination of the concept of a joint intention is the content of the intention. Traditionally in action theory it has been argued that intentions are intentions *to* do something, which is to say that intentions must be specified in action-referential terms. Conceptualizing intention in this way places restrictions on the contents of intention. They must be actions, and they must be actions which one can directly control. However, there is another strain of action theory, which is particularly relevant to joint action discussions because it holds that intentions may be conceived in propositional terms, and therefore may take the form 'intention that p', where p is a state of affairs in the world. This significantly expands the

[10] See Tuomela (2007) for an elaboration on the logical space carved out by the I-mode and the We-mode.

potential contents of an intention to include actions whose subject is not strictly the subject of the intention. Thus, it opens the possibility of intentions that take the form "I intend *that* we do X".

The possibilities that arise from different positions on these three issues may be summarized in a table as follows (Schweikaard 2008)[11]:

1. Subject	2. Mode	3. Content
a. individual	a. I-mode	a. Intention to
b. Interconnected individuals	b. We-mode	b. Intention that
c. collective		

For the sake of consistent terminology in the following, I use the term 'individualist' for theories of joint intention that take the isolated individual to be the subject of a joint intention and the term 'reductionist' for theories that conceptualize joint intention in the I-mode. Theories of joint action attempt to explain collective behaviour centered on a common goal using the concept of joint intention and are therefore developed out of some combination of these elements. In the following discussion of specific joint action theories, I will use this taxonomy to clarify the basic commitments of several positions in the literature.

§5 – Conclusion

In this chapter I have delineated the theoretical terrain by categorizing action types into two main groups, individual and social. Individual actions break down further into isolated and relational, based on the involvement of other individuals. Social actions also break down into two groups, joint and group. Group actions involve established social groups with a set of practices and norms, while joint actions involve collectives, which are defined in terms of a particular usage of the first person plural, namely, the collective 'we'. Within the category of joint actions there are both necessarily and contingently joint actions, but this distinction is ruled insignificant for the

[11] This table contains some modifications on the table present there. As I am not interested in classifying reductionist strategies in the same manner, and therefore do not need the higher-lower language, I have rearranged the content column.

purposes of this thesis by the definition of joint action as involving any action involving multiple participants with a single goal. Finally, I located the focus on joint intention in the joint action literature in its theoretical origins in traditional action theory. I then characterized several joint action issues against this theoretical background and specified the central question, with the help of several specific cases, as the following: what separates the sum of correlated individual actions from genuine joint action?

Chapter 2 – Joint Action Theories

In this chapter, I spell out the basic theoretical commitments of several important accounts of joint action, beginning with the distinction between summative and non-summative accounts and then recounting the development of the response to summative accounts. I follow Stoutland (1997) and Baier (1997) in arguing that the development of joint action theory out of traditional action theory has led to several assumptions that are harmful to an accurate account of joint action. I introduce a conception of the central problem of joint action theories put forward by Schmid and Schweikard (2013) as the tension between an irreducibility claim and an individual ownership claim, which I then use to differentiate the theories of John Searle, Michael Bratman and Margaret Gilbert. The two central aims of this chapter are to critically analyze the claim that we can consider an asocial human agent as our primary unit of investigation in the case of joint action, and to argue that in order to provide a proper account of joint action, we must recognize that certain collective concepts must be accepted.

§1 – Summative Accounts

In the introduction, we discussed the acceptability of ascribing intentional attitudes to collections of individuals. We will see that some accounts of joint action ascribe joint intentions to groups in order to explain joint action, while others reject this practice. Another type of response to the joint action problem not only denies that intentional attitudes may be literally ascribed to collectives, but also denies that there is something which separates the aggregation of individual actions under certain conditions from genuine joint intentional behaviour. These accounts are labelled summative accounts, and according to our table combine the elements 1a, 2a, and 3a in order to effectively deny that there is an interesting answer to the constitutive question of joint action laid out in §1.4. According to these views, while it may be a useful shorthand to refer to the intentional attitudes of a collective, collectives cannot literally have such attitudes, and it is just a figurative ascription for the sake of simplicity. If we desire to understand what it is for a collective to act, we should simply add up the actions or intentional states of the individual

members. Joint action then depends on individual intending subjects with direct action intentions in the I-mode. This is the view held by Anthony Quinton, who combines these two claims in the following statement,

> We do, of course, speak freely of the mental properties and acts of a group in the way we do of individual people. Groups are said to have beliefs, emotions, and attitudes and to take decisions and make promises. But these ways of speaking are plainly metaphorical. To ascribe mental predicates to a group is always an indirect way of ascribing such predicates to its members…To say that the industrial working class is determined to resist anti-trade union laws is to say that all or most industrial workers are so minded (Quinton 1976: 17).

Quinton proposes that for a group to intend *p*, in this case, for the working class to intend to oppose anti-trade union laws, most of the members must intend *p*. More formally, we may model this proposal thusly:

Collective C intends to *p* if and only if all or most of the members of C intend to *p*.

A summative account claims that ascribing attitudes such as belief or intention to a group is simply an abbreviated way of saying that the majority of the members of the group share that attitude in a distributive reading of the verb 'to share'. It is therefore the sum of the individual attitudes that constitute the group attitude.

This view and its stronger form, which includes a condition of common knowledge, is rejected by the majority of theorists working in the area of collective intentionality (Gilbert, 1987, 1989, 1994: Searle 1990: Bratman 1999). John Searle provides a simple and convincing counterexample (1990: 402). Imagine a number of individuals are sitting in a park. Suddenly, it starts to rain, and each individual runs for shelter, ending up in a gazebo located in the center. Although there may be some coordination - the people will avoid running into each other and so on - running to the shelter is not, in the collective sense, something that the people in the park do together. Now imagine another scenario with the exact same bodily movements, but in this case they are executing these movements as members of a ballet troop performing a dance. In both cases, to an outside observer there is no difference in the summation of individual behaviour. In this example, according to Searle, the dancers are engaged in joint action, whereas the collection of people running in the park is not.

On Searle's account, what distinguishes the two cases is not the outward bodily motions, but something belonging to the mental states of the participants. In the first case, each individual has an intention of the form 'I am running to the shelter' and these intentions are held independently

of one another, even though they may be held under conditions of common knowledge. Each person running may be aware that there are other people in the park who are also running and have the same intention, namely, to make it to the shelter. The majority of the members of the collective consisting of the people in the park when it begins to rain therefore meet the conditions of this type of summative account. They have an intention with the same content, and have this intention under conditions of common knowledge, which seems to fulfill the requirement laid out by Quinton. Nonetheless, it is not a case of genuine joint action. In the joint case, the outward behaviour is not a matter of coincidence. Running to the shelter is the explicit goal of the ballet troupe. According to Searle, in both cases, a participant has an intention expressed by "I am running to the shelter". But in the collective case this intention is in some sense dependent on an intention that necessarily stands in relation to the other members of the ballet troupe. Under Searle's account this intention is expressed as "We are running to the shelter." It is this 'we-intention' that distinguishes joint activity from a mere summation of individual acts.

The park example shows the inadequacy of summative accounts, in that it shows how distribution of the same intention among the separate individuals, even under conditions of common knowledge, does not amount to a collective intention or intentions in the sense of the ballet troupe. Summative accounts fail to distinguish between coincidental intentions with the same content and jointly held intentions and therefore fail to adequately answer the constitutive question in joint action as outlined in §1.3.

Another approach to rejecting summative accounts is taken by Margaret Gilbert (1989, 1994, 1996). She argues that group attitudes cannot be analyzed in terms of the sum of individual attitudes with the same content as that ascribed to the collective because collective attitudes are non-additive, that is, the collective mental state may depart radically from the individual mental states of the participants. She uses several thought experiments to clarify this point. Consider the case of a committee in which all of the members believe that eating meat is immoral, but none of the members expresses this view because they are afraid of the response they will receive from the other members. In this situation, it does not seem that we could attribute the belief that eating meat is immoral to the committee. The summative theorist may reply by introducing a condition of common knowledge, so that the members of the committee know that most of the members of the group believe that eating meat is immoral. Here, Gilbert complicates the case under

consideration in order to show that common knowledge is not sufficient. Imagine that the committee is divided in two, but that each committee has the exact same membership. The first committee is the Morality Committee, and is tasked with regulating the eating habits of the committee, while the second is the Dress Code Committee, which is tasked with developing the office dress code. Of this situation, Gilbert argues that it is possible to say, without contradiction, that (a) most members of the Morality Committee personally believe that eating meat is immoral, and that this is common knowledge among the Morality Committee; (b) the same goes, *mutatis mutandis* for the Dress Code Committee; (c) the Morality Committee believes that eating meat is immoral, whereas the Dress Code committee has no opinion on the matter.[12] The basic point of these examples is that collective attitudes are not simply the aggregation of individual attitudes and in fact, collective attitudes may be radically different from the personal attitudes of the members.[13]

§2 – The central problem

The central problem in analyzing joint action is determining to what extent collective intentional states are reducible to individual intentional states. In rejecting summative accounts, collective intentionality theorists subscribe to the idea that whatever collective intentions are, they are not simply reducible to normal individual intentions and basic interrelation. This is where the agreement on reduction ends.

Schmid and Schweikard (2013) characterize the basic problem of collective intentionality as the attempt to resolve a tension between the following two widely held claims:

1. The Irreducibility Claim: Collective intentionality is no simple aggregate, summation or distribution of individual intentionality.
2. The Individual Ownership Claim: Collective intentionality is had by the participating individuals and all the intentionality a person has is her own.

[12] For Gilbert's version of this example see Gilbert (1996: 198-200).
[13] This insight also informs, and is strengthened by, Phillip Pettit's discussion of 'social integrates' and the discursive dilemma (Pettit 2003: 185-188).

The tension between these two claims derives from the postulation of a separate intention for the act that we together do. During the course of a joint action, the ballet troupe in the park for example, the individuals involved have individual intentions concerning their direct bodily movement that clearly belong to them. As we saw, however, there are also intentions of the form 'we intend to *x*', which leads to the question: to whom does this intention belong? According to the irreducibility claim, this intention must involve more than simple individual intentions and common knowledge, and must therefore in some non-distributive sense belong to the group. On the other hand, according to the individual ownership claim, it must belong to each individual participant.

This framework for understanding the problem of collective intentionality locates different joint action theories within a defined boundary. The individual ownership claim obviates the possibility of a group mind over and above the individual participants, to which the collective intention could be attributed, thus demanding some degree of reduction, while the irreducibility claim limits the extent of this reduction by supplying the minimum requirement of collectivity. Most collective intentionality theorists accept some version of each of these claims. In their theories then, they attempt to develop these claims in a way which allows them to be mutually compatible. In the following, I will introduce three accounts of joint action and consider the problems that have been raised in the literature with regard to each of them.

§3 – Searle's Account

We have already briefly encountered Searle's theory of collective intentions in the section on summative accounts. He proposes a mode account of collective intentions, in that he postulates a special capacity for the case of joint action that is tied to the basic sociality of individuals. The central idea is that agents that participate in a joint action have intentions of a particular kind or form, namely, the We-mode. He therefore combines the elements 1a, 2b, and 3a from our table and his account is individualist, but not reductionist, according the usage of these terms given in §1.4. Searle strongly displays his individualist credentials from the outset of his exposition. He begins by claiming that his account is fully consistent with methodological individualism (Searle 1990: 406) and repeatedly emphasizes that "ontologically speaking, collective intentionality gives

rise to the collective, and not the other way around" (Searle 1997: 449). Nowhere is this clearer than in the conditions he lays down for his discussion of collective intentionality. He states that "anything we say about collective intentionality must meet the following conditions of adequacy:

1. It must be consistent with the fact that society consists of nothing but individuals. Since society consists entirely of individuals, there cannot be a group mind or consciousness. All consciousness is in individual minds, in individual brains.
2. It must be consistent with the fact that the structure of any individual's intentionality has to be independent of the fact of whether or not he is getting things right, whether or not he is radically mistaken about what is actually occurring. And this constraint applies as much to collective intentionality as it does to individual intentionality. One way to put this constraint is to say that the account must be consistent with the fact that all intentionality, whether collective or individual, could be had by a brain in a vat or by sets of brains in vats (Searle 1990: 407).

The first condition represents Searle's adherence to individualism because it denies that that the agency of individuals can in any way be constrained by that of a group and therefore also denies that groups themselves can be intentional agents. The second condition represents Searle's adherence to atomism[14], although of a modified form (Meijers 2003: 173). Meijers points out that while Searle may hold that our human capacities may require, in the actual world, that we exist in a social world, this is not logically necessary. This follows from his exposition because according to the second condition all intentionality, be it individual or collective, is independent of what the real world is like.

These two conditions entail that collective intentions exist in individual brains. When two people intend to do X together, they each have an intention of the form 'we intend to X'. Any individual intentions concerning the contributing actions then derive from this we-intention. Because there is nothing explicitly connecting these intentions, Searle's theory allows for the possibility that there is an isolated individual with a collective intention. In other words, Searle accepts that I may think that the intention I am acting upon is collective, but I may be wrong about the intentions of the other participants in our action and as a result mine is the only mind with a collective intention. The intention itself nonetheless remains collective. On this counterintuitive consequence, Searle states "Of course I take it in such cases that my collective intentionality is in fact shared; I take it in such cases that I am not simply acting alone. But I could have all the intentionality I do have even if I am radically mistaken, even if the apparent presence and

[14] Atomism is the position that a solitary individual, that is, an individual that is and always has been isolated from other human beings, may display all normal human characteristics (Pettit 1993)

cooperation of other people is an illusion, even if I am suffering a total hallucination, even if I am a brain in a vat" (1990: 407).

This still leaves us with the question of how it is possible for a single individual to have a collective intention. Searle argues that this capacity is biologically primitive, a feature of human (or similarly constructed) brains. We simply have an innate capacity to see others as potential candidates for cooperative activity. Searle combines this capacity with his concept of Background, which is a technical term in his theory referring to the conditions necessary for certain cognitive activities and bodily capabilities.

Searle's account then is non-summative, in that he rejects the reduction of 'we-intentions' to any combination of individual intentions and common knowledge or mutual beliefs, but remains starkly individualistic. He accepts both individualism and a form of atomism by arguing that all intentionality exists within individual minds, thus bracketing out the possibility of any structural interrelation between the we-intentions distributed among the participants of a joint action.

Searle's strict adherence to individualism and atomism has drawn several critics (Baier 1997, Stoutland 1997, Meijers 2003, Schmid 2009). In an influential critique, Meijers points out that in Searle's account no intentional state is actually shared, which makes it impossible to account for the intersubjective nature of joint action (2003: 174). Meijers states that "his account of collective intentionality is basically an account of a particular type of intentions of individuals, that is, where the sharing of these intentions is not a matter of concern" and therefore "it is questionable whether Searle has such a conception at all" (2003: 176). What Meijers emphasizes is that the successful coordination of action in pursuit of a common goal requires interrelation between both the contributing agents themselves and their intentional attitudes; otherwise, we just have a coincidentally matching, if fortuitous, configuration of action. If we imagine that each member of the dance troupe in the park is unaware of the we-intentions of the other dancers, it seems as if it is a complete accident that they acted together, and we lose the importance of the distinction between the dance troupe and the random collection of people running for shelter. It therefore seems as if isolated we-intentions are not enough to explain the direction and coordination of the individual actions which contribute to the joint action.

In addition to its individualism, Searle's account is also strictly internalist. In brief, internalism is a thesis about intentional mental states, such as beliefs and desires, and claims that having those mental states depends solely on properties intrinsic to consciousness. This means that a subject's beliefs and experiences are constituted within the mind, and not dependent on the subject's environment, culture or social context for their content. Tollefsen (2004) has criticized Searle's account on these grounds by claiming that Searle's simple statement of an internalist conception of content individuation is problematic. She points out that according to an externalist reading a brain in a vat cannot have the requisite beliefs, desires and intentions without being in the proper relation to the external environment. This seems to preclude the possibility of an isolated brain-in-a-vat having a we-intention. I will discuss externalism and the arguments for its validity in the next chapter, but for now it is worth noting the difficulties Searle's internalism may cause.

§4 – Bratman's Account

The second theory of joint action I will analyze belongs to Michael Bratman.[15] Bratman aims to develop a reductionist account that is "broadly individualistic in spirit" (1992: 341), but not strictly according to our classification system because he does not propose isolated individuals as the subjects of joint intention. He claims that joint intentions can be understood in terms of the intentions of the individual participants and their interrelations. The key move he makes is to hold that the participants not only intend their individual contribution to an act, but they also intend the joint act. In other words, the content of each individual's joint intention is the joint act and this joint intention is primary, in the sense that all individual intentions concerning the individual contributions to the joint act derive from it. When each individual intention of the joint act is interrelated in the proper way - in Bratman's language, when the subplans supporting the action mesh - we have a case of a genuine joint action. Formally, he presents his theory, which takes the 1b, 2a, 3b form according to table 1, thusly:

[15] It should be noted that Bratman frames his analysis as one of "shared intention". I here take the concept of "shared intention" to be equivalent in the relevant sense to "joint intention" as it is used here.

We intend to *J* if and only if

1. (a) I intend that we *J* and (b) you intend that we *J*.
2. I intend that we *J* in accordance with and because of 1a, 1b, and meshing subplans of 1a and 1b; you intend that we *J* in accordance with and because of 1a, 1b, and meshing subplans of 1a and 1b.
3. 1 and 2 are common knowledge between us. (Bratman 1999: 121)

Bratman further outlines several aspects of joint intention. He holds that joint intentions help to coordinate our intentional actions. If I intend that we perform a ballet and you intend that we perform a ballet, then we have a concurrence of mental states that serves to guide us to executing the bodily motions of the performance and achieving our goal. One way in which joint intentions do this is that they give us a reason to create meshing subplans. If I plan to perform some dance move, and you plan to perform some dance move, we will check to see if there is room for both moves, if that combination is aesthetically pleasing and so on. 'Meshing' is a technical term for Bratman, in that the subplans do not need to match exactly, they just need to avoid direct conflict and function properly together. For example, if Jim and Janice intend that they paint a house together, but Janice intends that they paint it red, and Jim that they paint it blue, then their respective subplans concerning the color of the paint don't mesh. If, on the other hand, Jim intends that they paint the house blue all over but has no preference as to where they should buy the paint and Janice does not care about the color but intends that they buy the paint at a particular store, then their respective subplans mesh without completely matching. This meshing of subplans in turn implies that the intentions of the participating agents must be interlocking, which means that each individual must have the efficacy of the other participants' intentions in his own intentions, in addition to the efficacy of his own intentions. It may be noted that by requiring that individuals stand in particular structural relations to one another in order to have a joint intention, Bratman's account conflicts with Searle's second condition.

An important aspect of this account is that it holds that the primary intention concerns the action that we perform together, not our individual part in the action. I intend to perform my part in the action *because* I intend that we together perform that action. The converse, which is rejected in this account, would be that a joint intention arises in virtue of our each intending to do our part in

combination with our common belief that the other participants will act in a complementary manner. This would make talk of joint intention figurative, a helpful metaphor to explain what is happening when individual intentions align properly along similar lines as a summative account, but there would be no literal manifestation of a genuinely joint intention.

This account faces a problem in that to intend the joint act requires the cooperation of at least one additional person. It is not clear how I can intend to perform a ballet or symphony or basketball play that requires a group of people. I may be able to intend to perform my part in the action but it is difficult to imagine that I can determine what the other participants intend. I do not have the authority or control over the other participants to influence their intentions in this manner, and if I did the action would not be shared, but coerced. With this in mind the problem this theory must answer is to what extent or under what conditions can our J-ing be the content of my intention.

Bratman here points to his use of the locution 'intending that', instead of 'intending to'. He argues that intentions of the 'that' form are not subject to the same conditions as those of the 'to' form and this is what allows participants to intend the joint act, in addition to their individual contributions. Strictly speaking then, what an individual intends can extend beyond what is under that individual's control, if she can predict that the other parties will act appropriately.[16] To take an example from the individual case, I may intend to work on my tan this afternoon, as long as I can predict that it will be sunny. Similarly in the joint case, I may intend that we J if I can predict that all the other members of the group will act in a complementary manner. This predictive quality, it is claimed, accounts for my ability to intend that we J. In other words, both agents are properly positioned to predict whether or not the other will develop the requisite intention and because they are then both involved in the origination of the shared intention, neither may claim to control the other's formation of the intention or their action. This account is therefore reductionist to the extent that first-person-plural extensions are analyzed in terms of individual intentions with a collective content and common knowledge.

Several critics have taken issue with Bratman's technical creativity (Baier 1997, Stoutland 1997, Velleman 1997, Schmid 2009). Bratman puts forward a view in which the intentionality of one

[16] For a detailed discussion of this issue see Bratman (1999) 142-161.

individual stretches into other individual's intentional realm. In order to avoid this criticism Bratman introduces his technical conception of 'intention that' and it is precisely this move that many of his critics focus on. For instance, Stoutland claims that the primary question of collective intentionality is how intentions may have a collective content and to simply introduce a technical notion of intention that allows for collective content begs the question. He states, "The analysis rests on *postulating* a technical notion of intention whose point is just to permit common content, which begs the question at issue, namely, whether intentions can have common content" (1997: 58).

Schmid makes a related point in his charge that Bratman's notion of joint intention is redundant (2009: 36). He claims that "Bratman's account presupposes the element of sharedness it aims to explain" because I can only intend that we *J* if we already intend to *J*. Schmid's point is that if we have some joint intention to do J, I do not need to form another intention *that we J*. Instead, I will form some derivative or participatory intention concerning my contribution to our J-ing and intentions of the form 'I intend that we *J*' seem to require that there already be an intention of the form 'We intend to *J*'. In other words, if the content of an intention contains a concept of collectivity, the sense of collectivity cannot itself then derive from this intention.

Baier also points out that Bratman's account requires a degree of cooperation that we would not attribute to all joint action (1997: 23). In order for us to make our subplans mesh, we must be willing to assist one another. Bratman acknowledges this point when he states that our intentions must not be coerced and must be "minimally cooperatively stable" (1992: 338). Bratman distinguishes between shared cooperative activity, in which the cooperation is explicit, such as in the case of painting a house together, and joint intentional action, in which there may be some competitiveness between participants, such as in the case of two hostile singers singing a duet together. Baier argues that the range of competitive joint activity is much larger than what Bratman imagines, and includes a wide range in the level of mutual assistance from the cooperative activities Bratman analyzes to thoroughly competitive activities such as tennis matches to hostile activities such as quarrelling and barroom brawling. She concludes that "both hostility and competition limit what Bratman calls 'cooperation'".

Another criticism of Bratman that focuses on the inability of his theory to account for many cases which we generally consider to involve joint action is put forward by Tollefsen (2005) and reinforced by Pacherie and Dokic (2006). They point out that Bratman's theory places high cognitive requirements on the participating agents of a joint action. Not only must they have concepts of mental states, because each participant must represent that the other participants have the appropriate intentions and other relevant attitudes, such as the beliefs and desires which allow the subplans to mesh. Furthermore, in order for the contents of the intentions of each participant to refer to both their own intentions and the intentions of the other participants in a such a thoroughly interrelated fashion, the must have robust meta-representational capabilities. This requirement on the cognitive abilities of participating agents contradicts a growing body of empirical research that suggests certain animals and small children, who do not have or have not yet developed such abilities, are nonetheless able to partake in genuine joint action (Rakoczy 2006, 2007; Tomasello and Carpenter 2007; Tomasello et al. 2005).

§5 – Gilbert's Account

Margaret Gilbert develops an account of social groups and the actions they perform centered on the concept of a plural subject, to which intentional attitudes can be ascribed (Gilbert 1989, 1990, 1996, 1997, 2006?, 2009). In this section, I introduce Gilbert's theory.

Gilbert begins her account by considering the simple example of two people walking in close physical proximity. She then asks the question: what are the minimum conditions we need to add to this situation to say that these two people are going for a walk together? She rejects an analysis that holds that they are walking together if each individually hope that they continue walking in this manner and they each know that the other feels the same way because it lacks an important normative dimension (1990: 4-6). According to Gilbert, when people are genuinely walking together, there are *obligations* and *entitlements* between them, such that, when one participant fails to perform the necessary contributory actions, or perform them in an appropriate way, the other participant has the *right* to rebuke her. Joint actions involve a special standing to make demands, which is itself a function of the joint activity (Gilbert 2006: 104).

This normative dimension is explained by the existence of a *joint commitment* between the participants. Gilbert characterizes this type of commitment as, "a kind of commitment of the will. In this case, the wills of two or more people create it, and two or more people are committed by it" (2006: 134). A joint commitment may only be brought about, or rescinded, jointly. Joint commitments come about when each participant expresses willingness to partake in the action, making this readiness in some way manifest for the others. Gilbert holds that this expression of readiness may take many forms, which range from explicit agreement to a type of "falling into it" (Gilbert 2006: 120). Nonetheless, it is present in all cases of joint activity.

Joint commitments also involve a special type of normative relationship between the participants. She holds that if *A* and *B* are jointly committed, then each of them is obligated to act accordingly and each of them is entitled to demand the other's conforming actions. This normativity applies only between the participants and is separate from any external moral concept of normativity which may call their goals into question. In other words, this relationship exists even between thieves planning to rob a bank. Further, the joint commitment is entirely separate from the individual commitments of the participants; none of the individuals involved must be personally committed to the goal or attitude in question.

When two or more people are jointly committed they form a *plural subject.* Plural subjects and joint commitments are correlated concepts; there can be no plural subject without joint commitments and all instances of joint commitments involve the formation of a plural subject. According to Gilbert, a plural subject is an entity or, "a special kind of thing, a 'synthesis *sui generis*'" (1996, 268) formed when individuals bond or unite in this particular way. She emphasizes that the concept of a plural subject does not require a "single centre of consciousness" nor a "distinctive form of 'subjectivity'". In any case, this "special kind of thing" can be the subject to which intentional action and psychological attributes are attributed. The conceptually necessary and sufficient conditions for the existence of plural subjects are formulated as follows (Gilbert 2006: 144-145):

> A and B (and…)…constitute a *plural subject* (by definition) if and only if they are jointly committed to doing something as a body – in a broad sense of 'do'.

The notion of committing to some goal 'as a body' is a technical term in her theory. It involves the commitment to constitute as far as possible a single body in service of that goal. She states that "the concept of a 'single body' is neutral with respect to the question whether the body in question is in some sense composed of individual humans beings" (2006: 137).

Gilbert then extends her analysis to intentional states. She does so simply by replacing the commitment to a goal with commitment to a belief, intention, acceptance of a state of affairs and so on. In the general case, the schema becomes:

> Individuals $A_1...A_n$ form a plural subject of X-ing that *p* if and only if $A_1 \ldots A_n$ form a joint commitment to X-ing as a body

The different joint commitments then involve different substitutions for X. The interesting substitution for X in our case is for intention. Gilbert holds that jointly committing to a goal and an intention to bring about some outcome in the world gives each of the participants in a joint action sufficient reason to coordinate her behaviour with the other participants in the pursuit of the joint goal. She explicates the necessary and sufficient conditions for joint action as follows (2006: 146):

> Two or more people are acting together if and only if:
>
> 1. They are jointly committed to espousing as a body the appropriate goal
> 2. They are fulfilling the behavioural conditions associated with the achievement of that goal
> 3. Their satisfaction of these conditions is motivated in each case by the existence of the joint commitment.

Gilbert's theory has received significant critical attention, which has resulted in several lines of objections. I will here highlight two. The first involves a concern that Gilbert's theory contains a fundamental circularity (Tuomela 1992; Tollefsen 2002; Schmid 2005, 2009). The charge of circularity takes a different, but related, form in each critique. The basic idea is that the bringing about of a joint commitment already involves collective intentionality. It seems as if Gilbert's

explanation of the origin of a joint commitment in the "expression of readiness" of two or more people requires that there be some form of communication, whether verbal or otherwise, and is possibly a joint action in itself. If all forms of communication are themselves joint actions, Gilbert's story for how joint commitments come about seems to require another joint commitment, which appears to lead to an infinite regress. Schmid (2005) has noted that this is only a problem if Gilbert's claim is that *all* collective intentionality is in the form of a joint commitment. It remains possible for Gilbert to claim that joint commitments presuppose more basic forms of collective intentionality, such as joint attention, which are themselves not joint commitments. Nonetheless, when focusing on the normative nature[17] of Gilbert's account, Schmid claims that (2009: 53),

> Gilbert seems to hold that obligations and entitlements are essential to shared intentional activity because any shared intentional activity ultimately originates in some form of (perhaps tacit) *agreement*. However, for something to count as an (however tacit) agreement, some form of shared intentional activity has already to be in place, for 'agreeing' is not anything single individuals can do, but something we have to do *together*.

The discussion of circularity in Gilbert's account is related to the vague nature of her account of an "expression of readiness". Another critique comes from Pacherie (2011), who claims that the metaphor of "acting as a single body" is equally vague. She claims that Gilbert's statement that intending as a body means to emulate as far as possible, by virtue of the actions of each participant, a single body that intends to do X, "is not very illuminating" and requires interpretation (Pacherie 2011: 181). She also holds that fleshing out this idea will clarify the role of individual and social normativity in Gilbert's account of joint action. She suggests that one way to flesh out the idea of acting as a body is in terms of satisfying the type of rationality constraints that apply to individual agency. According to Pacherie, "to intend as a body would then be a matter of acting in such a way that the actions each together satisfy norms of consistency, agglomeration and means-end coherence" (2011: 181). This would in turn include commitments to making the relevant subplans mesh and mutual responsiveness.

[17] There is a significant debate on the normative nature of collective intentionality between Gilbert and Meijers, who claim that collective intentional behaviour is always normative, and Searle, Bratman and Tuomela, who claim that while collective intentional behaviour may often involve a normative component, it does not necessarily do so. In supporting a Gilbertian picture of joint action, I am also putting forward a normative view. Unfortunately, space does not allow for a detailed consideration of all the arguments for and against this position.

§6 – Generalization of Problems in Joint Action Analyses

At the beginning of the first chapter I outlined the central problem that joint action theories attempt to solve, namely, what separates fortuitously correlated individual actions from genuine joint actions. We may now consider whether these three accounts that have been outlined satisfactorily answer this question. The aim of this section is to state the problems these accounts face in a general manner, so that we may consider whether the ideas concerning extended cognition may help to solve these issues.

As has been pointed out by Searle's critics, if Searle's account fails to adequately answer this question, it is on two related grounds, which we may bring together from the various objections under the labels the 'sharedness' problem and the 'coordination' problem. First, we have the 'sharedness' problem. Searle explicitly commits himself, as a result of his adherence to internalism, to the idea that an individual may be completely mistaken that the 'we' in her we-intention actually refers to a 'we' (Searle, 1990: 408). He suggests that rather than claim that this is not a case of a genuine collective intention, we should see this as a we-intention that is simply mistaken. This contradicts the intuitive idea that an individual cannot have a genuine we-intention unless there are actually other individuals who share this intention. It seems that on the basis of what it means to share, we are dealing with a phenomenon that is inherently relational. If that is the case, the existence of other agents is not only fortuitous, it is a necessary condition.

Second, we have the 'coordination' problem. Even if we accept this non-standard usage of the verb 'to share' and its unattractive consequences, it is not clear that Searle's account captures the cooperation or coordination necessary for joint action. Because each token of the we-intention exists in the minds of the participants, there is no sense of unified agency or coordination of intentions, and therefore it appears as if the resulting coordination of the actions of the participants is a matter of cognitive luck. They just happen to have a we-intention with the same content. Searle does address this issue, but in an unsatisfactory way. He states that collective intention presupposes, "a Background sense of the other as [a] candidate for cooperative agency; that is, it presupposes a sense of others as more than mere conscious agents, indeed as actual or potential members of a cooperative activity" (1990: 414). Under Searle's theory of the Background, Background capacities are not representational, but rather a set of preintentional

resources that enable intentional states to function. They are therefore biological phenomena to be spelled out in neurophysiological terms and cannot be accounted for in intentional terms. This only serves to mystify coordination, and therefore also joint action – if we take coordination to be an aspect of joint action - by taking it out of the explanatory realm of philosophy and the social sciences and placing it within the purview of the natural sciences.

Because Bratman attempts to understand joint intention in terms of a well-defined type of interdependence of individual intentions rather than positing an atomistic form of collective intention, he is subject to neither the 'sharedness' problem nor the 'coordination' problem. Both of these features are built into his account in his focus on the mutual responsiveness and support of the participants. At first glance then it appears that Bratman gives us a promising account of joint action. On further inspection however, it becomes unclear whether Bratman has provided an account that explains collective action or assumes it. I label this problem the 'circularity' problem. As Stoutland argues, Bratman's introduction of the technical notion of 'intend that' simply posits that intentions may have common contents. Stoutland claims that as this is a significant change to our normal concept of intention, especially with regard to the self-reflexivity condition[18], and takes for granted exactly the issue that is at question, this move should not be accepted. If, however, we grant Bratman's introduction of this notion and the claim that intentions may be understood in propositional terms instead of action-referential terms, it opens his account up to the charge of circularity leveled by Schmid (2009). The appeal to the collective action in the content of the intention assumes the existence of some collective which is conceptually prior to the intentions of the individual participants because it is necessary for those intentions to have content.

The second issue Bratman's account faces is a scope problem. By focusing on cases of small-scale, highly cooperative activities between participants with a high-level of cognitive functioning in situations where the participants are in a position to mutually influence their intentions and related attitudes, Bratman rules out many cases of joint action as defined in §1.3.

[18] The self-reflexivity condition is that intentions, unlike other intentional attitudes, necessarily refer to the one who has the intention. In traditional action theory it may be traced back to Baier (1970), and is frequently discussed Velleman (1989). For an analysis of this condition in the here-discussed theories see Searle (1990: 411-413) and Bratman (1992: 335-336). For discussion of the problems this leads to see Baier (1997: 24-25) and Stoutland (1997: 55-56).

As Tollefsen (2005) and Pacherie and Dokic (2006) have pointed out, participants with lower-levels of cognitive functioning, such as children, are capable of performing actions that involve a plurality of participants with a common goal, and as Baier (1997) has shown, many cases of joint action, in the sense defined in §1.3, involve participants who are not in a position to mutually and cooperatively mesh their subplans, for example, in the case of two people quarrelling. Bratman's account then fails to satisfactorily answer the question put forward in §1.3 because it takes the concept of sharedness that it needs to explain for granted, and fails to apply to all cases of joint action as here defined.

Gilbert's account shares the problem of circularity with Bratman's account, albeit in a significantly different way. In our consideration of what it means for a group to jointly commit to act as a body, we may want to conclude that this is itself a joint action, as forming an agreement appears to be an instance of collective intentional behavior. It is my suspicion that this complaint arises from the relative obscurity of Gilbert's account of expression of willingness and acting together as a body. In chapter 7, I argue that recent empirical research consistent with the philosophical framework of extended cognition provides a method for interpreting these two metaphors, which both clarifies them and resolves their apparent circularity.

Consideration of these objections appears to leave us with four desiderata of a joint action account. This list is not meant to be exhaustive, but rather a rough outline of the general problems facing current accounts. Any theory of joint action must account for the following four features:

1. Sharedness: Explain the way in which joint intention is shared among the participants
2. Coordination: Explain the way in which joint intention coordinates the behavior of the participants and allows them to plan their action in a mutually responsive manner.
3. Circularity: Do not presuppose any concept of collectivity that it seeks to explain
4. Scope: Apply to all action which involves two or more participants who have a common goal.

These preliminary conditions outline the general structure and basic theoretical position of a joint action theory. They serve to clarify the aims and convictions underlying the discussion I will put forward in the final section in relation to extended cognition.

If we return to the table outlined in §1.4, we will see that between summative accounts, as in Bratman and Searle, we have examined the basic possibility of accounting for joint action on the basis of combining individual subjects with either I-mode or We-mode intentions, or interrelated individuals as subjects with I-mode intentions in propositional terms and, if we accept the criticisms of Meijers, Stoutland, Baier and Schmid, this seems to show that these accounts are not satisfactory. This is not to say that no such account is logically possible. The objections brought forward here refer specifically to the proposed theories and another theory combining the same elements may be able to meet the conditions outlined so far. However, I take the difficulties faced by these accounts to constitute sufficient evidence that we should look to the remaining options in the table to develop an account of joint action and intention. The remaining possibility which has here been introduced is the relational, irreducible account of Margaret Gilbert. I have outlined some of the concerns with her account, and I will return to these concerns, and how I think the theoretical framework of the extended mind may satisfy them, in chapter 7.

§7 – Conclusion

In this chapter, I considered four types of joint action accounts: summative, individualist, reductionist, and collectivist. Summative accounts radically reduce joint intentional states to the intentional states of individuals in the I-mode. Summative accounts are rejected outright on the basis of thought experiments from Searle and Gilbert. Searle also serves as the example of an individualist account in this chapter, as he holds that all intentionality must take place within an individual brain. His account is, however, not reductionist because he understands we-intentions as irreducible to individual intentions and posits that we-intentions are an innate biological capacity. Bratman provides a reductionist account, in that all we-intentions are constructed from intimate interrelation of individual intentions and common knowledge, so that there are no intentions of the form ‘We intend’. Gilbert presents a collectivist account because she posits an irreducible plural subject that may bear intentional states of the ‘we’ form. I also survey the problems raised in opposition to these approaches in the literature and group the objections under four headings: sharedness, coordination, circularity and scope. I argue that the sharedness and

coordination objections apply to Searle's theory, that both Bratman's and Gilbert's theory are subject to the circularity objections and that in addition Bratman's theory faces a scope objection.

Part 2: Extended Cognition

Chapter 3: First-wave Extended Cognition[19]

The existing accounts of joint action surveyed in the last chapter each contain underlying commitments to various positions in the philosophy of mind. As we saw, Searle explicitly holds that the content of mental states must be internal, while Bratman and Gilbert reject this claim. On the other hand, Bratman holds that mental states must take the I-form, while Searle and Gilbert allows for mental states which take the we-form. One way to clarify these accounts and adjudicate between them is to focus more closely on the philosophy of mind upon which they rely. For example, one proposed motivation for the individualism in joint action accounts is the adherence to Cartesian assumptions in the theory of mind. Baier, speaking on just this point, decries the "Cartesian brainwash" infecting the collective intentionality literature and asks what it will take for us to accept the first person plural as the starting point for an analysis of joint action (1997: 18). Schmid continues this line of thought by tracing his distinction between formal and subjective individualism to specific Cartesian theses (2009: 34-35). He traces the formal individualism (here termed reductionism) of Bratman to Descartes' claim that he wishes to contemplate his mind in isolation from society, which leads Descartes to claim that intentionality can only be in the I-form. Subjective individualism (here termed individualism), which informs Searle's theory, on the other hand, derives from Descartes portrayal of the mind as the solitary place of representations. From this Descartes concludes that the mind is strictly independent of anything external to the individual mind, a commitment which Searle defends.

In this chapter, I consider an alternative conception of mind and cognition that has arisen in cognitive science and challenges these Cartesian assumptions. This new perspective does contradict many of the conditions of the Searlean and Bratmanian theories of mind which structure their accounts of joint action, but instead of following this line of thought and directly challenging existing theories of joint action, I will here introduce this new approach. In the following part, I make the conditional argument that if we are to accept the basic commitments of this conception of the mind and if we desire to use it as a basis for a theory of joint action, that joint action account must be irreducible and relational. But, before we turn to those claims, we

[19] I will use the term cognition to refer to certain processes that we take to be cognitive in our everyday pre-theoretical understanding and avoid the term mind, except where it is explicitly used by the theorist's under consideration. The issue of mind and its relation to certain cognitive processes is too large to be dealt with here.

must first consider what commitments this new conception of the mind involves. I begin by introducing the way in which this new strain of theorizing about the mind arose in cognitive science and the relating the main works in which this line of argument entered the philosophical debate. I then discuss the development of 'active externalism' by Clark and Chalmers (1998) and the subsequent modifications and clarifications made by Clark (2001, 2008). Specifically, I focus on the notions of coupling and parity.

§1 – Situated Cognition: a short history

What is now known as the Hypothesis of Extended Cognition (HEC) developed out of a movement in cognitive science comprised of several loosely-allied theses commonly referred to as 'situated cognition'. As a general characterization, the situated movement holds that thinking and learning are profoundly tied to both physical and social contexts and as such, suggests re-thinking the relationship between knowing and doing. Cognition is characterized as a relation involving an agent in a situation, rather than as an activity in an individual's mind. This relation takes place between an agent, her action, and the context and culture in which she performs the action.

As these introductory statements show, this approach strongly contrasts with Cartesian assumptions about the mind. This is not an accident. Situated cognition developed into a coherent movement in response to dominant theories of mind that informed cognitive psychology and directed it towards a concept of cognition as a process which takes place in isolation from the body and world. The philosophical background of these positions in cognitive psychology may be traced to theories of mind with strongly Cartesian elements.[20] The most prominent example of this is the "methodological solipsism" endorsed by Fodor (1980) and Searle (1983) which is reflected in the latter's theory of joint action. The idea is that the successful description of our mental states and cognitive processes must be achieved internally without reference to the agent's body or the surrounding world. Cognition is therefore a type of computation involving the

[20] While Cartesian dualism is usually rejected (mental substance is seen variously as either reducible to, identical to, realized by, determined by, or supervenient on physical phenomena), the mind remains isolated in several important ways discussed below.

processing of representations that is "wedged between perception (on the input side) and action (on the output side)" (Wilson & Clark 2009: 56). Hurly refers to this as the "sandwich model of cognition" (2001: 3-4). This computation is a process of symbol manipulation that relies on the syntactic (rule-based) and semantic (meaning) properties of the symbols. This theory is more prominently known as the computational theory of mind and is often understood using direct comparison of the mind to a computer, in which the mind is seen as the software run by the hardware of the brain. Importantly, because this understanding of the mind refers only to the internal manipulation of symbolic inputs and outputs, cognition is characterized as an *isolated* process which is separated from the world. Modern proponents of this view hold a more sophisticated position that moves away from the computer analogy, but maintains that all cognitive processes and mental phenomena are located in and explained by the neuronal processes of the brain. Following Varela et al (1991), we may label this view 'cognitivism'.[21]

Philosophical doubts about this position began with the work of Putnam (1975) and Burge (1979). They proposed arguments against the ability of isolated models of the mind to satisfactorily account for mental content. The most famous argument involves the "Twin Earth" thought experiment (Putnam 1975). Putnam imagines a planet that is identical to earth in every way, except that on Twin Earth instead of water with its chemical structure of H_2O, they have a liquid which is superficially exactly the same as water (has the same color, feel and so on) but has the chemical structure XYZ. To complicate matters, the inhabitants of twin earth also speak English and also call this liquid 'water'. Putnam's thought experiment takes place in the past, at a time when neither the inhabitants of earth nor those of Twin Earth have the theoretical knowledge necessary to differentiate between H_20 water and XYZ water. From this we may conclude that the people of Earth and Twin Earth have identical experiences with water.

Putnam then poses the question: does a resident of Earth mean the same thing when she says "water" as a twin-resident of Twin Earth means when she says "water"? As stipulated in the conditions of the thought experiment, the brains of both the resident and twin-resident are identical, yet it appears that when the resident says the word "water" it refers to H_2O and when the twin-resident says the word "water" it refers to XYZ. This suggests that the contents, or

[21] Other names for this position include: individualism (Wilson & Clark 2009; Menary 2010b), and brainbound (Clark 2008). For a summary of these positions see Rowlands (2010: 2-3).

intrinsic properties, of an individual's brain do not suffice to determine the references of the words that they use. Further, if we accept that the meaning of a word determines its reference, then the meaning of the word is not determined by the intrinsic properties of an individual's brain. This leads Putnam to conclude that "'meanings' just ain't in the head" (1975: 227).

The idea that intentional states require broad content because they depend on external context served as the first step in the challenge to the "cognitivist" position outlined above. The original challenge has developed in many directions, perhaps the most radical of which are the various strains of situated cognition with their claim that all knowledge is intimately related with action, and is structured, scaffolded and profoundly embedded in physical and socio-cultural contexts.

§2 – Varieties of Situated Cognition

'Situated cognition' is an umbrella term for a variety of distinct, yet interrelated, theses about the nature of cognition and theoretical frameworks for studying it. Situated views hold that mental processes are some combination of (1) embodied, (2) embedded, (3) enacted, and (4) extended, which leads to another term that perhaps better reflects the pluralist nature of this domain: 4E cognition[22]. The *embodied* thesis holds that cognition is significantly influenced by the type of body an organism has because cognitive processes are partly constituted by non-neural bodily structures. The *embedded* thesis holds that mental processes develop in an intimate interrelation with a specific environment and without the proper cognitive 'scaffolding' the functionality of these processes decreases. These two theses are often held together and do not strictly contradict the computationalist theory of mind, because while acknowledging the importance of the body and environment, they allow that actual cognition still takes place in the brain. The *enacted* thesis complicates these arguments by challenging the strict separation of action and cognition. It claims that mental processes are not just neural processes but also involve things that the organism *does*. Enactivists argue that mental processes are in part constituted by the actions of an organism in the world. The *extended* thesis holds that mental processes are not exclusively located in the brain, but extend out beyond the boundary of the organism. To complicate matters

[22] See Gallagher (2008) and Rowlands (2009b: 3).

further, the paradigm of *distributed cognition* shares similar theoretical commitments in that it claims that cognitive processes are best understood as distributed across individuals, tools and the environment, but, as we will see, challenges the account of cognitive agency given by some extended cognition theorists.

The relationship between these positions is too varied and complex to be dealt with here, as is their specific relationship to Cartesian theories of mind.[23] What we may draw from this brief discussion is that proponents of situated cognition variously hold that the neuronal processes of the brain are coupled with some combination of (i) body, (ii) natural environment, (iii) cognitive tools/artefacts, or (iv) social community/environment. As a result of our interest in the extension of the mind into the social community or environment, I focus on the extended thesis in the following. By taking cognition outside the structures and actions of a single organism, it allows for the possibility of a collective cognitive system that may be the bearer of cognitive states. It is this aspect of the situated cognition literature that I develop here.

§3 – Active Externalism

The classical source of the extended cognition debate in philosophy is a paper by Clark and Chalmers entitled "The Extended Mind" published in 1998, and this position is further developed by Clark (2004, 2008). The motivating question of the original paper is: where does the mind stop and the rest of the world begin? In answer to this question Clark and Chalmers (hereinafter C&C) put forward a view they call *active externalism*. This is the view that components of an individual's environment, to which the individual is connected in the appropriate way, may be just as much a part of the cognitive process as parts of the human brain. C&C argue that not only is the external environment causally active in cognitive processes but also, in the right circumstances, a *constitutive* part of a cognitive process. The central idea, in the original formulation of the HEC, is that when individuals use artefacts such as smartphones, computers or pencil and paper, the interaction between the individual and these artefacts may result in a *coupled system* that functions as a cognitive system. In other words, artefacts, to which an

[23] For an extended discussion see Rowlands (2010 Ch. 1, 2, 3).

organism in connected, are not just causally linked to cognitive processes, this system of connections *is* the cognitive process.

We may here distinguish between this thesis, associated with C&C and called variously active-, how-, or vehicle externalism, and content or passive externalism associated with Putnam (1975) and Burge (1979) described in §3.1. [24] As we saw, content externalism holds that the content of a mental state may be determined in part by environmental or causal factors. Active externalism can be seen as a radicalization of the Putnam and Burge view in two ways. First, not only the content of mental states may be constituted by external features, but also the processes of cognition themselves may be in part external. Second, in Putnam and Burge, the external components are passive, historical features, while in C&C they are present, active features. In the twin earth example, even though the content of the belief differs, the process remains the same and it remains inside the head. C&C argue that the driving force of the cognitive process need not be restricted to the inner biological realm. According to them, both cognitive contents and cognitive operations can be constituted and supported by a combination of inner biological states/processes and external non-biological states/processes.

§4 – Causal Coupling

Active externalism focuses on the efficacious features of the environment in the here and now. Essential to understanding how these active features combine to form a cognitive system is the notion of causal coupling. C&C introduce this concept in the following way, "the human organism is linked with an external entity in a two-way interaction, creating a *coupled system* that can be seen as a cognitive system in its own right" (1998: 8, italics in original). C&C then go on to specify three conditions which must be present in a coupled system between an organism and an external entity. First, all the components in the system must play an active causal role. Second, they must jointly govern behaviour in the same sort of way that cognition usually does. Third, if we remove the external component of this system, the competence of the entire system must drop, just as it would if we were to remove part of the brain. When these conditions are fulfilled,

[24] For taxonomy of the types of externalism and an extended discussion of the various objections and responses to each view see Hurley (2006).

this process counts as a coupled process, and the external factors are "just as causally relevant as typical internal features of the brain" (C&C 1998: 9).

It is important to note that these conditions imply that if we were to keep the internal structures of the brain the same and alter the environment, in a kind of extended twin earth thought experiment, both the competence and the behaviour of the organism could change radically. Critics may respond that a change in competence and behaviour is not enough to infer constitution, but C&C envision a relationship between mind and environment with a mutually constraining causal influence that develops over a period of time and transforms the nature of each component. The external and internal processes develop with and adapt to one another to create a system that jointly governs future behaviour. The most explicit description of this continuing influence between mind and world is given in Clark (2008: 24) where he states,

> Continuous reciprocal causation (CRC) occurs when some system S is both continuously affecting and simultaneously being affected by, activity in some other system O. Internally, we may well confront such causal complexity in the brain since many neural areas are linked by both feedback and feedforward pathways (e.g., Van Essen and Gallant 1994). On a larger canvas, we often find processes of CRC that criss-cross brain, body and local environment. Think of a dancer, whose bodily orientation is continuously affecting and being affected by her neural states, and whose movements are also influencing those of her partner, to whom she is continuously responding!

With this example Clark attempts to show that although we could break this system down into each of its component parts and analyze them in terms of internal and external, this task is complicated by the fact that the parts are continually influencing and being influenced by one another and unnecessary because analyzing this situation as a single system allows us to explain more thoroughly how each of these components coordinates to produce certain behaviours. According to Clark, if we can show that the brain, body and environment are reciprocally combined in a coherent arrangement in this way, then they form a coupled system.

§5 – The Parity Principle Introduced

Coupled systems themselves may be either cognitive or non-cognitive. In order to determine when a coupled system is cognitive C&C propose the parity principle. This principle has been the topic of much debate, both from critics of the HEC in general and from supporters of the HEC who argue that this formulation and use of the parity principle leads to misunderstandings of the

HEC or that some other condition is more appropriate.[25] We will return to this debate in the next chapter when we discuss second-wave arguments for extended cognition. For now, I will introduce the principle and describe its role in C&C's theory. C&C define the parity principle thusly:

> If, as we confront some task, a part of the world functions as a process which *were it done in the head*, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process (1998, 8, italics in original).

The appeal to the parity principle is meant to get us to disregard the apparent importance of the brain and focus purely on the functional role played by the various parts of a cognitive system. In other words, the parity principle claims that location is not the determining factor in the judgement of whether or not some process is cognitive and it does so on coarse-grained functionalist grounds. It states that if some process plays the right role and is integrated in the right way with other processes that we uncontroversially consider cognitive, then that process itself is cognitive, regardless of where it is performed. To do otherwise would be unwarranted bio-chauvinism.

§6 – Otto and his Notebook

The final step in C&C's argument is to extend this line of argument to mental states, such as belief, and therefore to a concept of the mind. Specifically, in their 1998 paper, C&C claim that "*beliefs* can be constituted partly by features of the environment, when those features play the right sort of role in driving cognitive processes. If so, the mind extends into the world" (12, italics in original). C&C argue for this point primarily on the basis of a thought experiment. This thought experiment has two characters, Inga and Otto. Both hear about a particular exhibit at the Museum of Modern Art in New York. When Inga hears about the exhibit she goes into her biological memory, recalls that the Museum is on 53rd St and directs her route accordingly. Otto, on the other hand, suffers from a mild form of Alzheimer's and records important names, dates, and addresses in his notebook. He takes the notebook with him wherever he goes and uses it frequently. When he hears about the exhibit, he turns to the relevant page in his notebook, sees

[25] See Adams and Aiziwa (2001, 2007, 2010a for the former and Wilson (2004) and Gallagher & Crisafi (2009), Sutton (2010), Menary (2010a) for the latter.

that the Museum is on 53rd St, and again, sets out accordingly. The question then is: what is the difference between these two situations? C&C answer that, in terms of causal-explanatory relevance, there is no difference, "what makes some information count as a belief is the role it plays, and there is no reason why the relevant role can be played only from inside the body" (1998, 14). In other words, the physical implementation of the causal role is irrelevant to its functional description. We can explain Inga's behaviour by appealing to her desire to see the exhibit and her belief that the Museum is on 53rd St., and we can explain Otto's behaviour by appealing to the exact same desire and belief, except here the belief supervenes on factors outside the human brain.

§7 – Coupling Conditions for Tools

In order to preclude concerns that every time I use my computer or smartphone it becomes part of my mind, C&C introduce a set of criteria to be met by non-biological artefacts for inclusion into coupled cognitive system. In a later reformulation which maintains the substantive content of the original, Clark outlines and clarifies these criteria thusly (Clark 2008: 79):

1. Availability and frequency: The resource should be reliably available and typically invoked (Otto always carries his notebook and won't answer that he doesn't know until after he has consulted it).
2. Automatic Endorsement: Any information received must be more or less automatically endorsed. It should not usually be subject to critical scrutiny (unlike the opinions of other people, for example). It should be deemed about as trustworthy as something retrieved from biological memory.
3. Accessibility: Information in the resource should be easily accessible as and when it is needed.
4. Previous Endorsement: Information in the resource is has been previously endorsed and this is the reason why it is in the resource.[26]

[26] This condition is sometimes left out of Clark's formulation (see for example Clark 2004).

C&C argue that these criteria produce an intuitively satisfactory set of judgements on putative tool-based cognitive extensions. They have been labelled the 'glue and trust' conditions because they emphasize that the external equipment must be habitually, fluently and transparently used, must be available as and when it is needed, and trusted as one would trust information recalled from biological memory.

§8 – Clarifications: Extended Cognition, Functionalism and the Brain

The thesis of extended cognition is thus primarily a claim about the nature and location of our mental states and processes that follows from a commitment to a traditional functionalist understanding of mental phenomena as multiply realizable. As a theory of mind, functionalism individuates mental states based on the role that they play within a system and not the internal constitution of the states themselves. Cognitive processing, under this view, is best described, not in terms of specific cellular activities or chemical properties, but in terms of its abstract structure or role. The functional role is determined by the causal relations to sensory stimuli, other mental states and ultimately behaviour. So, in simplified terms, pain may be a mental state that is caused by bodily damage, causes other mental states such as anxiety, and causes behaviours such as screaming or groaning. Under this theory, all beings with internal states that satisfy these conditions are capable of bearing this mental state. Cognitive states can therefore be realized in physically different systems using a variety of mechanisms (Wheeler 2010: 249). This is the claim of multiple realizability and is what is at work in the parity principle and the Otto thought experiment, where the information in the notebook functions equivalently to the information in biological memory in the belief-forming process.[27] It may be characterized as parity-based extended functionalism.

The radical aspect of the extended cognition claim lies in the assertion that not only is the content of a mental state individuated by external factors, but also sometimes certain external factors are

[27] It is also the point on which many critics have attacked the HEC. This discussion, which has generated a large literature which is too far afield to be dealt with here, but it depends on the "grain" of the functional similarity, that is, how similar external processes need to be to internal process for external and internal processes to play the same role in supporting behaviour. Rupert (2004) and Adams and Aiziwa (2009) argue for fine-grained or more specific similarity, while Clark (2008) and Wheeler (2010) argue for coarse-grained or abstract similarity.

required for the realization of the mental state itself. Its relation to cognitivist or internalist (or "sandwich") theories of mind is therefore clear; it is a direct rejection.[28] Instead, extended systems theorists "confront an image of the local mechanisms of human cognition quite literally bleeding out into body and world" (Clark 2008: 70). By articulating a detailed conception of the mind as comprised of mental states and cognitive functions that supervene on organized systems of processes and contents that criss-cross the boundary between brain, body and world, Clark provides a strong alternative to the internalism of 'cognitivist' conceptions of the mind.

There are two specific features of this account that are particularly relevant for this project as they will come up in the discussion of agency in relation to the HEC. First, C&C's account builds from a distinction between epistemic and pragmatic actions made by Kirsh & Maglio (1994). Pragmatic action is taken so as to alter the world in order to achieve a particular goal. Epistemic action in contrast is meant to alter the world so that one's cognitive, or problem-solving, load is eased, but may not involve any actual physical advance toward a clearly defined goal. Instead, these actions are designed to reveal information and make mental computation faster. The example that C&C employ deals with expert Tetris players checking the correct fit of the piece on the screen instead of in their heads. The important point of this example is that in evaluating the correct fit on the screen, players often move the piece further away from its destination, thus seemingly thwarting the explicit pragmatic goal. Nonetheless, the epistemic action supports the pragmatic action by more efficiently revealing the required information. Second, Clark states that "Possessing a contentful mental state is most plausibly a property of a whole active system" (2008: 76). Because the cognitive vehicles that are capable of bearing cognitive content can be distributed across brain, body and environment, the cognitive states they support are a systemic-level phenomena. In other words, the supervenience base of a cognitive state includes all the active parts of the system and therefore the cognitive state is realized by the interaction between all of the active parts of the system. It is therefore most correctly viewed as a property of the entire system.

[28] This rejection is somewhat softened by the partial nature of the HEC claim. It is not that all mental states are necessarily constituted by some part of the environment; it asserts only that some of them are. This leads Rowlands to claim that "Thus, contrary to popular belief, EMT [Extended Mind Thesis] is compatible with the possibility of a brain in a vat. It is just that, if EMT is true, the mental life exhibited by the brain would be somewhat truncated" (Rowlands 2009: 56).

The attribution of cognitive states to the entire system leads to a final point about the view of neural processing and the brain at work in this theory, which may be represented by the following five claims (Clark 2008: 136-137)[29]:

1. Neural activities are not blessed with a special property that makes them alone suitable for cognition and intelligence. The important characteristic they do show is functionality in supporting intelligent behaviour, which in turn relies on causal flow and does not differentiate between internal or external.
2. There is no single part of the brain or privileged inner component that is responsible for the "real thinking". Instead, mind and reason are "the emergent products of a well-functioning swirl of (mostly) self-organizing complexity".
3. The brain is "cognitively impartial". It does not differentiate between potential vehicles of cognition based on method or location, but rather on time and efficacy.
4. The structures that drive and shape our cognitive processes are actively created by "cycles of self-stimulating activity".
5. Control of the flow of this activity is fragmented and distributed, allowing different inner resources to call upon various external resources without this process reaching the level of conscious deliberation, or any intervention by an inner executive.

Clark's general point is that the brain is the essential core element of human cognitive activity, but it does not care about the nature or location of the processing and storage resources that it recruits to accomplish a cognitive task and it does not have a single 'center' of control. The brain itself is a distributed cognitive system that can be seen to support intelligent behaviour only if it is studied as a system. In relation to external components, the brain has an essential role because it 'recruits' the various components into a softly-assembled extended device. Soft- assembled systems are provisionally assembled units whose composition is temporary and highly-flexible depending on the tools and structures available in the current environment.[30] However, in the recruitment itself, the brain does not privilege internal components, and further, once this system

[29] Aspects of these claims are at work in much of Clark's discussion, and are components of much of the literature he cites, but they are explicitly laid out here together. For further discussion, see his "Principle of Ecological Assembly" (Clark 2008: 13) and his "Hypothesis of Cognitive Impartiality" (Clark 2008: 121).
[30] Thelan and Smith (1994).

is in place the flow of information required for the completion of a cognitive task occurs across the entire system.

§9 – First-wave Extended Cognitive Agency

As is clear from the Otto thought experiment, C&C treat extension of mind primarily as a capacity for enhancing the performance of individual minds and as a result focuses on the impact of tools on our thinking. This emphasis also presents itself when C&C consider the possibility of socially extended cognition. C&C recognize that the parity principle allows that other people could act as stand-ins for some internal cognitive processes and consider several examples. The first concerns a businessman and his secretary, in which the secretary could be an extension of her boss's memory. The second concerns a regular customer and a waiter at the customer's favourite restaurant, where the waiter could be an extension of the customer's beliefs about his favourite meals, or perhaps even his desire for good meals.

What is striking about the cases C&C suggest is that they conform so closely to the agent-tool case. Essentially, they treat the social extension of cognition as just another tool to enhance the capacities of an individual mind. This has the benefit of mitigating the intuitional dissonance concerning the locus of subjecthood and the ascription of responsibility. It is easier to accept extended cognition in cases where one part of the system is clearly the subject, and can therefore be assigned responsibility when something goes wrong.

Clark's ability to maintain an individualist conception of subjecthood results from his contention that while cognition may not be organism-bound, it is still organism-centered. He names this "Hypothesis of Organism-Centered Cognition", and expresses its content thusly:

> Human cognitive processing (sometimes) literally extends into the environment surrounding the organism. But the organism (and within the organism, the brain/CNS) remains the core and currently the most active element. Cognition is organism centered even when it is not organism bound (2008: 139).

The HOC therefore claims that even if cognition involves assembled hybrid systems, the core of cognitive agency remains with the neural networks of the brain. This move has two steps. First, he suggests that in order to individuate cognitive agents, we should not search for the location of the cognitive mechanisms, but rather for "a reliable, easily identifiable physical nexus of

perception and action, apparently driven by a persisting and modestly integrated body of goals and knowledge" (Clark 2008: 118). This process of identification of the cognitive agent precedes the question of the location and type of underlying mechanisms required for a particular cognitive task. Second, as we saw in §3.8, Clark makes the assumption that the brain is the essential core element of "all episodes of individual human cognitive activity" (Clark 2008: 118).

The extension of cognition results from the fact that the brain does not care where a particular cognitive process takes place. The brain maintains its privileged status by functioning as the recruiter for the various elements of an extended cognitive system. Clark states, "We may ask just how, and according to what principles, the various elements…came to combine into a specific soft-assembled information processing device. In this process of soft-assembly, the brain surely plays a very special role" (2008: 122). Once this system is in place it is responsible, as a whole, for cognitive processing, which boils down to the flow and transformation of information and serves as the "machinery of ongoing thought and reason" (2008: 122). By emphasizing this function of the brain, Clark navigates between the embedded mind hypothesis, which would give special place to the causal role of the environment, but deny that external factors *constitute* mental processes and the concept of the homunculus brain on the other hand. The brain controls the recruitment, but once the device is coupled together, all cognitive properties must be attributed to the device as a whole.

Clark sums up his account of organism-centered cognition as follows:

> …in rejecting the vision of human cognitive processing as *organism bound*, we should not feel forced to deny that it is (in most, perhaps all, real-world cases) *organism centered*. It is indeed primarily (though not solely) the biological organism that, courtesy especially of its potent neural apparatus, spins and maintains (or more minimally selects and exploits) the webs of additional structure that then form parts of the machinery that accomplishes its own cognizing (Clark 2008: 122)

This separation of the recruitment process from the cognitive process, or the process of assembly from the product of assembly, is meant to resolve the tension between viewing cognitive states as emergent properties of a system that has no central control, and where no individual part of the system is more important for the creation of this state from the rest, with the foundational role Clark attributes to the neural processes of the brain as the 'locus of cognition'.

§10 – Conclusion

In this section, I introduced first-wave extended cognition through its central concepts – functionalism, coupling and parity – and located within the development of situated cognition in cognitive science. I also emphasized the role that the brain plays in Clark's theory and how that influences his account of cognitive agency by dividing the *recruitment* of a cognitive system from the *functioning* of that system. Finally, I showed how this distinction allows Clark to claim that the individual remains the center of cognitive agency, and mentioned some effects of this commitment on Clark's view of socially extended cognition.

Chapter 4 - Second-wave Extended Cognition

In this chapter, I show how second-wave extended cognition developed out of Clark's account of cognition, introduce its four main concepts and consider how this development affects an account of cognitive agency.

§1 – Clark's two principles

Second-wave extended cognition begins with Clark's suggestion that there are two strands of extended cognition argumentation. The first depends on functional isomorphism and the parity principle. This is what we have looked at so far. The second places the emphasis on the *complementary* nature of internal and external resources. Despite the attention the parity principle has drawn in the critical literature, this second method may be seen as primary even in Clark's formulation. Clark states (1998: 99):

> Given this second line of argument (the one stressing complementarity), it is best to see functional isomorphism as at most part of a sufficient condition for cognitive extension…The more interesting and plausible argument, I feel, is the one which describes the seepage of mind into the world by stressing that "the brain's brief is to provide complementary facilities that will support the repeated exploitation of operations upon the world to provide computational processes (such as powerful pattern completion) that the world, even as, manipulated by us, does not usually afford.

Thus, the parity principle forms a sufficient condition for cognitive extension but not a necessary one and is better seen as "an informal test" for cognitive extension (Clark 2010, Sutton 2010). What is more interesting, according to Clark, is the way in which external elements play a role that is different from, but complementary to, inner elements. Clark elaborates on this complementary relation between distinct methods with respect to computation and storage by stating, "The brain need not waste its time *replicating* such capacities. Rather, it must learn to interface with the external media in ways that maximally exploit their particular virtues (Clark 1997: 220). The second principle shifts the debate from arguing for coarse-grained functional parity, which has proven divisive and seems to lead to a stalemate, to arguing for mutually beneficial distribution of cognitive tasks across complementary resources, which has resulted in a focus on the level of integration of these resources. In some recent work, Clark (2010) has reiterated and further emphasized the role of integration, proposing that it is the degree of integration in a unified system and complementarity between distinct resources that determines

whether a system is one in which the cognitive properties supervene on more than the neurological components of the system.

Second wave theorists do not necessarily reject the parity principle, for it is one way of characterizing the hybrid nature of cognition, but they do contend that there is a tension between a parity approach and their own. They argue that it is at best incomplete or misleading and at worst false because it disregards the importance of the differences between internal and external resources. They instead focus on Clark's second suggestion (Sutton 2010: 198-200; Menary 2006a: 333; Menary 2010a: 235). This has resulted in the development of four key concepts: complementarity, integration, manipulation and transformation.[31]

§2 – The Four Basic Principles of Second-wave Extended Cognition

i. Complementarity

Sutton defines the complementarity principle as follows (2010: 194):

> In extended cognitive systems, external states and processes need not mimic or replicate the formats, dynamics, or functions of inner states and processes. Rather, different components of the overall (enduring or temporary) systems can play quite different roles and have different properties while coupling in collective and complementary contributions to flexible thinking and acting.

Cognitive systems are here, in line with Clark's analysis, conceived of as shifting arrangements of various components which together perform certain cognitive tasks. Sutton argues that this type of complementarity must avoid arguing that the internal/external difference in components must fix the properties of certain components. In other words, we should not exclude the possibility that an external resource may be, for example, as dynamic and fluid as an internal process, because it is an external process. The properties of individual components are important in understanding how the cognitive whole functions, but must be discovered on their own and differences in these properties do not preclude their ability to form a functional whole. Another point which Sutton emphasizes is that the integration of the various parts into a cognitive whole

[31] These four concepts are developed primarily by Sutton (2010), Sutton et al (2010), Rowlands (2010), Menary (2010ab).

often alters or transforms the inner parts of the system. This leads us into the second principle, which is developed most thoroughly by Richard Menary.

ii. Integration

Menary supports Sutton's complementarity principle (Menary 2010b: 571) but places a twist on the line of argumentation by conceiving of extended cognition as based on *cognitive integration*, in which cognitive processes are seen as coordinated processes. The central principle is stated as follows:

> Cognition is the coordination of bodily processes of the organism with salient features of the environment, often created or maintained by the organism. A coordinated process allows the organism to perform cognitive tasks that it otherwise would be unable to; or allows it to perform tasks in a way that is distinctively different and is an improvement upon the way that the organism performs those tasks via neural processes alone (2010b: 563).

Menary argues that by focusing on integration we can avoid many of the criticisms of extended mind, as they apply mainly to the demarcation of functional equivalence required by the parity principle and incorporate much of the embodied and embedded research into an extended cognition account (Menary 2010: 229-231).

Menary also introduces three complementary ways that we may understand integration (2010: 233). First, we may understand it as 'biocausal coordination'. This method, which derives from dynamical systems theory, focuses on the reciprocal coupling between systems that are part of a larger system. There must be a symmetrical relation, which means that the two systems are mutually constraining of each other's behaviour, and they must have causal influence over one another for as long as they are coupled. This way of understanding complementarity is foundational for the joint action account I propose in chapter 7. The second method is 'embodied engagement', which focuses on the way body schemas, or unconscious sensorimotor programs for action, connect the organism and the environment. The third method is the 'manipulation thesis', which leads to the next subsection.

iii. Manipulation

Following Rowlands, Menary holds that the manipulation thesis is the claim that "cognitive processes are not located exclusively in the skin of cognising organisms because such processes are, in part, made up of physical or bodily *manipulation* of structures in the environments of such organisms" (Rowlands 1999: 23, quoted in Menary 2010a: 228, emphasis in original). Menary elaborates the way in which humans change their environment by directly restructuring their local environment with their bodies or creating tools or artefacts. This aspect of second-wave extended cognition shares many theoretical commitments with the embodied and embedded cognition thesis, but maintains that due to the intimate causal nature of the interaction we must consider the results of manipulation as, in part, constituting a cognitive system.

Menary identifies four types of manipulations of external cognitive vehicles: biological coupling, epistemic actions, self-correcting actions, cognitive practices. Biological coupling is related to the enactivist conception of cognition and involves concepts such as extended phenotypes and sensory motor contingencies. Exploring this in depth would take us too far afield; however, the next three types have a direct impact on the project of this thesis. We have seen epistemic action in the work of C&C. Menary relates the Tetris example of Kirsch and Maglio (1994), which he postulates as a paradigmatic example of epistemic action, which, as we saw, is the use of the environment as its own representation in order to more efficiently complete a cognitive task. He then introduces the concept 'self-correcting action'. These actions are similar to the epistemic actions, because they also direct and structure practical action when completing tasks, but they warrant a separate term because they do not invoke an explicit physical manipulation of the environment. Instead, they involve the use of linguistic structures as well as props and gestures in influencing cognition and future action.

Here Menary relies on the work of Lev Vygotsky, particularly Vygotsky's diachronic account of the role of language in infant development (Vygotsky 1978). According to Vygotsky's theory, higher cognitive capacities, such as reasoning, are initially developed as social phenomena in what he terms the 'intermental plane'. The idea here is that a child is guided in her development of problem-solving techniques by both her own egocentric speech about the activity and the speech of her parents. In this sense the development of the child's problem-solving ability is

intermental. Over time, the success of the child in performing certain activities reduces the need for the parental support of action and the egocentric speech of the child is internalized. Therefore, the *intra*mental flavour of cognition arises only after the development of *inter*mental capacities. Menary argues that the speech, whether inter- or intramental is not merely an epiphenomenon of the activity, but provides a cognitive structure for the activity itself by playing an important role in the organization of the sequence of supporting actions, the search for solutions and the recognition of failures. Menary claims that "self-corrective speech, whether private or public, is used to structure, direct and correct actions that lead to the completion of cognitive tasks" (Menary 2010b: 570).

The last quote leads us to the final aspect of Menary's account of normative manipulations, namely, the concept of a 'cognitive practice', which is simply the way representational systems are manipulated according to particular *cognitive norms* in order to complete *cognitive tasks* (Menary 2010: 238). Examples of cognitive tasks are problem solving, planning, and making inferences and a norm is cognitive, rather than moral or social, when it is directly tied to completing one of these activities. Menary identifies four types of norms which help to guide cognitive practices. Purposive norms direct the activity towards an end or goal. Corrective norms relate how representations may be used to correct a particular activity in pursuit of an end. Manipulative norms regulate the manipulation of inscriptions of a representational system and interpretive norms concern the interpretation of the inscriptions of a representational system (Menary 2010a: 239).

iv. Transformation

In his exposition of the developmental foundation of cognition, Menary also presents the concept of *cognitive transformation*. This concept concerns the way in which manipulative norms are acquired against the background of the social context of the individual. Menary investigates how "the normative and social structure of the environment, mediated by learning and training histories, has a direct transformatory effect on the body" (Menary 2010b: 572). Through an analysis of body schemas, particularly with respect to the way in which these schemas structure

our interactions with the environment and are developed through a process of learning, practice and habituation, Menary concludes that our bodies have been transformed to be capable of creating external representational structures, such as writing. One example that Menary discusses at length is the ontogenetic development of mathematical abilities, such as the internalization of a public numeral system, which results in the manipulation of new operations and the creation of new representational formats. According to Menary, this process of development is the transformation of our cognitive capacities. He states, "The deeply transformative power of our learning histories in the cognitive niche is one that reformats the representational capacities of the brain in terms of public symbol systems" (Menary 2010b; 576).

§3 – Agency in Second-wave Extended Cognition

The central tenets of second-wave extended cognition do not in themselves commit to any particular account of cognitive agency. Despite emphasizing the relational and social nature of both the development and practice of cognition, they are compatible with holistic but still individualistic accounts of agency and action. There is, however, some intuition that the organism-centered account of Clark is misleading. Menary appears to question the first-wave account of cognitive agency by including the following statement in a discussion of the problems with first-wave cognition:

> Extended-mind-style arguments based on the parity principle have encouraged critics to think in terms of an internal cognitive system that is extended outward into the world. Hence, on one interpretation, it implicitly endorses a picture of a discrete cognitive agent some of whose processes get extended out into the world (Menary 2010: 234).

In this particular section however, he does not further elaborate on this claim. In an earlier work Menary argues, "We are not just coupling artefacts to pre-existing cognitive agents; the organism becomes a cognitive agent by being coupled to the external environment" (2006a: 342). This suggests that while Menary accepts the basic idea of an individual being the center of cognition, he holds a stronger thesis about the necessity of extension for the existence of a cognitive system. Cognitive agency is therefore still organism-centered, but the organism itself is not a pre-existing or predetermined cognitive agent. Nonetheless, this account of cognitive agency does not appear to provide any advantage to the theory of cognitive agency put forward by Clark for an analysis of joint action because it maintains the central individualist commitment.

Other second wave theorists appear to prefer alternative accounts of cognitive agency. In a footnote (Sutton 2010: 215-216 fn1), Sutton states, "My take on EM, based on the 'complementary' rather than the 'parity' of inner and outer resources, brings it closer to the related theory of 'distributed cognition'". Distributed cognition and its benefits and drawbacks for an account of joint action will be discussed in the next chapter. I will attempt to show how this suggested relation between the cognitive agency postulated by distributed cognition and second-wave extended cognition appears to challenge the adherence of second-wave extended cognition to Clark's account of cognitive agency.

Another instance of a second-wave theorists questioning the HOC may be found in the following quote from Rowlands in his discussion of the ownership of a cognitive process: "For our purposes, the notion of a subject can, I think, be understood quite broadly. For example, I do not wish to rule out the possibility that the subject in question might be a group rather than an individual" (Rowlands 2010: 135). This opens the door for the postulation of group cognitive processes and states we will see in §5.3.

Cognitive agency in second-wave extended cognition appears to be undetermined. The focus on the social and intermental aspects of complementarity and integration in part undermines the HOC proposed by Clark, but does not necessarily reject the claim that cognition is still somehow centered on the organism, if understood holistically, as is demonstrated by Menary. Nonetheless, an alternative view of cognitive agency, arising from the connection of second-wave extended cognition and distributed cognition, raises the possibility of a collective cognitive system that is not focused on a particular organism but rather views both the assembled cognitive system and the process of assembly is itself distributed. Second-wave extended cognition appears to be compatible with both of these accounts of cognitive agency.

§4 – Conclusion

In this chapter, I have shown how: (1) the development of second-wave extended cognition changes the points of emphasis of the discussion to complementarity and integration, rather than functional isomorphism, (2) offers its own account of cognitive tasks practices and norms, (3) intensifies the role of intersubjective processes in the creation of cognitive systems, particularly in Menary's use of Vygotsky, and (4), may be used to challenge the account of cognitive agency associated with first-wave extended cognition.

Chapter 5 – Distributed and Group Cognition

Taking up Sutton's suggestion, in this chapter I introduce the theoretical commitments of distributed cognition and consider their relationship to second-wave extended cognition. Distributed cognition is often lumped together with versions of extended cognition, both by proponents and critics of the extended cognition view.[32] They share several theoretical presuppositions, such as the general tendency to test the boundaries of cognition and give cognitive status to external scaffolding, tools and cognitive artefacts, and the emphasis on the historical and cultural dimensions of cognition. I attempt to show here how the incorporation of tools and artefacts into cognitive systems may be understood in conjunction with the concepts of complementarity and integration, and how the focus of cognition over time relates to the concept of transformation.

§1 – Distributed Cognition

Although the beginnings of distributed cognition may be traced to earlier works, perhaps the most important example of the distributed cognition framework comes from the cognitive anthropologist Edwin Hutchins in his 1995 book *Cognition in the Wild*, which is an ethnographic study of "pilotage" or navigation near land. The term "cognition in the wild" refers to the grounding of this work in a particular methodological approach. Hutchins attempts to remove the study of cognition from a controlled laboratory setting to the actual processes of cognition used in everyday environments and settings. Hutchins refers to this approach as *cognitive ethnography* (Hutchins 1995: 371).

Hutchins relates his extended intuitions and perhaps more radical ambitions in terms of cognitive agency in his statement of the purpose of his book:

> This book is an attempt to put cognition back into the social and cultural world. In doing this I hope to show that human cognition is not just influenced by culture and society, but that it is in a very fundamental sense a

[32] See especially, Adams & Aizawa (2007); Menary (2010a); Sutton (2006).

> cultural and social process. To do this I will move the boundaries of the cognitive unit of analysis out beyond the skin of the individual person (Hutchins 1995: xiv).

This quote suggests the importance of determining the proper *cognitive unit of analysis* for distributed cognition accounts. It is in this area that distributed cognition most emphatically accepts the basic tenets of extended cognition (active externalism, rejection of specific location of cognition), and at the same time, rejects the individualist framework of Clark's HOC. This motivation has subsequently been worked out by Hutchins and Jim Hollan into a principle of distributed cognition analyses. Hutchins relates this principle as follows, "The central claim of the distributed cognition framework is that the proper unit of analysis for cognition should not be set a priori, but should be responsive to the nature of the phenomena under study" (Hutchins 2011: 426). What determines the cognitive unit that is to be studied is relative to the explanatory needs of a particular project. At times this may mean that the correct boundary is drawn at the skin of an organism, and at other times even this may be too wide, for example when we are attempting to isolate processes within the brain, or subsections of the brain. At other times, the boundary of the organism is too restrictive and we may need to see an extended cognitive system as the proper unit of analysis. While this statement is a full endorsement of the admonition of first-wave extended cognition not to arbitrarily assume that cognition is limited to the individual brain, it departs from the HOC of Clark. According to this principle it is not that we start from an individual's brain and then move outwards on the basis of recruitment and soft-assembly to the formation of a cognitive system. Instead the starting point is to seek "a system that can dynamically configure itself to bring subsystems into coordination to accomplish various functions" (Hollan et al. 2000: 175).

In *Cognition in the Wild*, which is a study of navigation on a US naval ship, the bridge of the ship is taken to be the proper unit of cognitive analysis. Hutchins maintains that the task of pinpointing the ship's location involves a great deal of coordinated behaviour and problem solving, tasks which are distributed among individuals in the navigational team, parts of the ship and specific navigational instruments. He points out that even though this task requires multiple, and quite distinct, representational subsystems, no single subsystem is solely responsible for the representation of the ship's location. Hutchins claims that because no one human could accomplish all the things that must be done in order to complete the cognitive task (repeatedly determining the ship's location as it nears a port), we must view the individual humans as

components of a larger cognitive system. In addition to the collective cognitive processes of the individuals involved, Hutchins holds that a variety of mechanisms, including instruments and other artefacts, contribute to the overall cognitive system. Therefore, the representation of the ship's position is best seen as realized by the relevant individuals, navigational instruments and their coordinated actions (Hutchins 1995: Ch. 3 and 4).

This theoretical framework has been subsequently developed and applied to various cases of "cognition in the wild", such as jury decision making, problem solving and so on, by Hollan, Hutchins and Kirsch. They differentiate between three main types of cognitive distribution which highlight several of the connections between distributed cognition and second-wave extended cognition: socially distributed cognition, coordination of internal and external structures, and cultural cognition (Hollan et al. 2000: 176-180; Hutchins 2001: 2068; Hollan & Hutchins 2010: 241-244).

The central thesis of socially distributed cognition is that social organization is a cognitive architecture. Social organization regulates the directions and fluidity of information flows in collections of people according to certain patterns. It therefore plays an important role in the transmission and transformation of information, shaping the cognitive processes of the group in question and providing the structure according to which the group is able to implement the cognitive processes required to complete a cognitive task.

The coordination of internal and external structures proposed by Hutchins and Hollan emphasizes the role of causal coupling between individual neuronal processes, external artefacts and symbol systems for the emergence of cognition both developmentally and in specific cognitive processes. This reinforces the view that tools, artefacts and instruments are not merely stimuli but are integrated elements in a hybrid cognitive system. Distributed cognition here understands the coordination of internal and external structures along the lines of complementarity and integration. Tools and artefacts contribute to the completion of cognitive tasks because they are able to perform cognitive functions in ways that are significantly different from, but essentially complementary to, those of the neuronal processes of the individuals. Hutchins points out that "it is essential to distinguish the cognitive properties required to manipulate the artefact from the [cognitive property] that is achieved via the manipulation of the artefact" (2001: 2070).

The third type of distributed cognition concerns the close relationship between culture and cognition. In other words, it concerns cognition that is distributed over time (Hutchins 1995: 354; Hollan et al. 200: 178; Hollan & Hutchins 2010: 241). Hollan et al. describe this intimate dual connection as follows:

> This means, on the one hand, that culture emerges out of the activity of human agents in their historical contexts, as mental, material and social structures interact, and on the other hand, that culture in the form of a history of material artefacts and social practices, shapes cognitive processes, particularly cognitive processes that are distributed over agents, artefacts, and environments (2000: 178)

According to this view, culture is seen as a cognitive process and cognition as a cultural process. Culture may be understood as cognitive distribution over time because it acts as a receptacle for accumulated knowledge and resource for learning, problem-solving and reasoning. It therefore structures cognition by allowing for the implementation of cognitive practices that have been previously developed.

More recently, Hutchins has developed the concept of *cultural practice* and elaborated on the role of this concept in structuring cognition. This concept shares many theoretical similarities to the concept of transformation in second-wave cognitive extension, such as the importance of bodily knowledge for cognitive processes, the cultural transmission of this bodily knowledge, and the transformation of cognition that arises from these factors. Hutchins holds that cultural practices have two functions in the organization of human interactions with the world. They achieve this organizational success "first by furnishing the world with the cultural artefacts that comprise most of the structure with which we interact. Second, cultural practices orchestrate our interactions with the natural phenomena and cultural artefacts that produce cognitive outcomes" (Hutchins 2008: 2018). External representation systems, such as numbers and writing, are products of these cultural practices. Further, they are not static mental representations of knowledge, but rather embodied skills with intimate effects on our sensorimotor activities (Hutchins 2008: 2012). Hutchins discusses the example of seeing constellations. Stars are perceived without respect to a cultural background, but the determinations of constellations "exist only by virtue of someone enacting it via cultural practice that allocates visual attention in a particular way" (Hutchins 2011b: 441). We see material reality according to patterns and turn them into representations based on these patterns by enacting their meaning through cultural

practices. Hutchins argues that in understanding how our cultural practices inform our lower-level sensorimotor processes we may also develop an account of how higher-level cognitive processes, such as reasoning, emerge from the lower-level sensorimotor ones. He states, "Cultural practices orchestrate the coordination of low-level perceptual and motor processes with cultural materials to produce particular higher-level cognitive processes" (Hutchins 2011a: 434).

§2 – Agency in Distributed Cognition

As part of his analysis of cultural practices and larger distributed systems, Hutchins warns that "there is a danger of attributing to the individual cognitive properties that belong to the larger distributed system" (Hutchins 2008: 2011). This concern derives from distributed cognition's unique approach to the study of cognition. From the start, distributed cognition rejects the goal of extending the individual mind and with it certain aspects of the privilege that neuronal processes enjoy in Clark's account. Instead, distributed cognition, from the beginning, takes a wider point of view and sees the entire system, including its cultural and historical dimensions, as the starting point for a study of cognition.

Hutchins has also developed an account of agency in distributed cognition specifically in response to Clark's organism-centered account.[33] He argues that Clark's view is too individualistic and too centered on the brain. As we saw in §4.3, Clark justifies his organism-centered view by dividing the assembly and recruitment process, which takes place in the brain, from the resulting cognitive system, which is extended. Hutchins rejects this distinction by claiming that "the assembly process itself is extended and orchestrated by the cultural practices that constitute the cognitive niche" (Hutchins 2011b: 442). The assembly and recruitment processes themselves are here seen as distributed in the cultural landscape.

Hutchins' argument against Clark relies heavily on the role of culture in cognition. He claims that Clark fails to realize that the cultural world is dynamic, for example in the way that it includes the

[33] Hutchins account is not the only attempt to develop a view of cognitive agency for distributed systems. Giere (2006) develops a deflationary account that limits the agency to the individual human components of a distributed cognitive system. However, as he adheres to Clark's account of HOC (§3.9), his account offers little beyond what has already been said about Clark for the purposes of this thesis.

actions of other people, and therefore quite unlike Otto's notebook. He proposes that, "a straightforward way to deal with [cognitive assembly] is to abandon the assumption that the biological brain is the essential element. Doing so, of course, requires that one look elsewhere for the apparently impartial forces that assemble cognitive systems" (2011a: 439). Hutchins argues that we may look to the dynamic character of culture to discover how the assembly and recruitment process of cognitive systems to include other possible sources in addition to the brain and body. Hutchins proposes a competing account of the assembly of cognition in his *hypothesis of enculturated cognition*: "The ecological assemblies of human cognition make pervasive use of cultural products. They are always initially, and often subsequently, assembled on the spot in ongoing cultural practices" (Hutchins 2011b: 445). This view emphasizes the dynamic nature of culture in the orchestration of the assembly of extended cognitive systems. Hutchins concern is therefore that an individualist account of cognitive agency threatens to isolate the activity of the brain from the dynamics of cultural practices, which he takes to unnecessarily privilege the organism.

By rejecting the organism-centered view and arguing that the cognitive unit of analysis must be chosen based on the task that is performed, Hutchins' account of distributed cognitive agency raises the possibility of groups of people being proper cognitive agents. As a motivation for his treatment of the bridge of a ship as a single cognitive system, Hutchins states:

> In terms of the energy budget of a human group and the efficiency with which a group exploits its physical environment, social organizational factors often produce group properties that differ considerably from the properties of individuals. Clearly, the same sorts of phenomena occur in the cognitive domain. Depending on their organization, groups must have cognitive properties that are not predictable from a knowledge of the properties of the individuals in the group (1995: xiii).

This sort of distributed agency implies the existence of collective cognition because collectives may have cognitive properties that supervene on, and are not simply reducible to, the actions and properties of the members of the collective. The collective may bear cognitive states over and above its constituent individuals. We see here an argumentative strategy similar to the one used by Gilbert against summative accounts of collective attitudes which we will see again in a more formal instantiation in the discussion of group cognition. The collective properties (attitudes/processes) are not simply the aggregation of the individual properties because the collective properties may radically differ from the individual properties. This is one aspect of

Hutchins navigational study.[34] The task of locating the ship's position involves a significant amount of coordinated behaviour and problem solving, which is distributed among various individuals, parts of the ship and navigational tools. Many of these subsystems are themselves representational, for example the particular individuals, and yet, no single subsystem is responsible for the representation of the ship's location. According to Hutchins, the final representation, which guides the further action of the ship's crew, belongs to the whole and is best explained as being realized by, or supervening on, the relevant individuals, artefacts and tools and their coordinated behaviour. Hutchins attributes mental states to certain collectives on the ground, so that such a collective functions as a highly integrated single unit that manipulates representational media with the aim of producing intelligent results. He maintains that groups may hold such mental states, or be the subject of such mental processes, as remembering (1995: 196), perceiving (1995: 182, 194), having skill or expertise (1995: Ch. 4 and 5), holding hypotheses and being biased in the evaluation of these hypotheses (1995: 239-61).

In this analysis Hutchins introduces a specific standard for judging whether the system is the proper unit of cognitive analysis, namely, does the cognitive architecture of the system allow for the propagation of representational states across various media and does it bring these representations into coordination in order to facilitate intelligent goal-directed behaviour (Hutchins 1995: 117). This standard significantly diverges from the argumentative strategies of collective intentionality theorists, because it introduces specific empirical requirements for collectives that go beyond the ascription of intentionality to the collective. I will review the problems with using this approach to determining the appropriate unit of cognition in relation to collective intentions and joint action in §7.7.

§3 – Group Cognition

The concept of cognitive agency in distributed cognition has been recently developed further based on a series of empirical studies. Theiner et al. (2010) present an overview of much of this work with a prescient philosophical discussion. Their central thesis is that "groups of people can

[34] For another oft quoted use of this theoretical framework see Tribble's study of the cognition required for the performance of plays in Elizabethan England (Tribble 2005).

manifest cognitive capacities that go beyond the simple aggregation of the cognitive capacities of their individual members’" (Theiner et al. 2010: 378). Not only do they claim that we may ascribe cognitive states and processes to groups[35], but they also claim that in many cases it is necessary to do so because recent research shows that cognitive capacities commonly ascribed to individuals are also present at the level of groups.

In arguing for this position they employ the following strategy. First, Theiner et al. claim that studying group cognition in this way is a special case of using the theoretical framework of extended cognition. They argue that we may expand the claim that systems that extend beyond the biological boundaries of the organism are capable of cognition. Once the boundaries of cognitive systems have been broken down, they claim that the next step should be to turn our focus away from the demarcation problem, that is, determining the “mark of the cognitive”, and instead focus on particular processes, which are “cognitive” according to our pre-theoretical understanding of the distinctive roles these processes play in producing intelligent behaviour. They propose processes such as memory, learning, and problem solving. They then develop a series of principles for identifying cognitive processes and emergent capacities in groups, and argue that empirical research in distributed problem solving shows that these capacities can be instantiated in groups. They spell this strategy out as follows (Theiner et al. 2010: 379):

> Some mental properties seem clearly projectible to groups – groups solve problems that individuals cannot, for example – but others, like consciousness, seem equally unprojectible. Because of the heterogeneity among different cognitive and mental predicates, we believe that abstract arguments about group minds or extended minds should be replaced by specific discussions tied to particular properties: group memory, group problem solving etc….We should rather be asking whether specific cognitive models that work at the level of individuals also work at the level of groups.

The central concept in their account is emergence. At the most basic level, emergence requires non-aggregativity, that is, the properties found at the higher level of organization are not simply the aggregate of the individual parts. As we have seen, this is the same concept employed by Gilbert and Hutchins. However, following Wimsatt (1986), they offer a more technical definition, which they express thusly (2010: 382):

> A property P of a complex system S is aggregative if and only if (i) P(S) is invariant under the inter-substitution of parts of S, or any other parts taken from a relevantly similar domain; (ii) P(S) remains qualitatively similar (differing only in value) under the addition or subtraction of parts; (iii) P(S) is invariant

[35] For this section I will use the term group in the colloquial sense in order to maintain continuity with Theiner et al.

> under the decomposition and re-aggregation of parts; and (iv) there are no cooperative or inhibitory interactions among parts.

Aggregativity is then the mark by which we determine whether a property is emergent. For Wimsatt, emergence is not all or nothing; there are degrees of emergence based on how many of the four conditions a property meets. They label properties that fail all four conditions "minimally decomposable" (2010: 382) and develop several requirements for group cognition based on this concept. First, group cognition is an emergent phenomena and is not the result of mere aggregativity. Second, group cognition involves the differentiation of roles of the members and an organizational structure. Third, the emergence of group level cognitive processes signals an important explanatory kind when analyzing the behaviour of the system.

Theiner et al. also introduce two methodological principles which they propose as a guide to identifying group cognition. The first is a twist on the parity principle, which they designate the "social parity principle" (2010: 384). It takes the following form:

> Social Parity Principle: If, in confronting some task, a group collectively functions in a process which, were it done in the head, would be accepted as a cognitive process, then that group *is* performing that cognitive process

They quickly point out that this principle is meant to serve as a baseline default metaphysical position and not a demarcation of cognitive or non-cognitive processes. For that, Theiner et al. present the second principle, namely that we have evidence for group cognition when the capacities for problem solving that can be ascribed to the group are distinct from those possessed by any of the individual members.

In support of this claim, they analyze four types of empirical research for instances of group-level cognitive processes: stigmergic path formation, the collective coloring problem, division of labour in groups and transactive memory systems. I will now consider three of these examples, leaving out the issue of collective coloring.

i. Stigmergic Path Formation

This case study concerns the ability of groups to create path systems that are mutually advantageous to the members of the group without the members themselves being aware of the collective effect. Early trail blazers though a jungle, for example, use their machetes to build paths. These paths are then used and modified by later trekkers, who are guided by a compromise between reaching their destinations and taking the paths where travel is the easiest. This phenomenon has been studied in the "Active Walker" computational model. In this situation, there is a classic hard information processing problem, namely, the creation of a set of paths that connect a set of points using the minimal amount of path length. The pertinent result of the computational model is that good approximate solutions may be found under certain parameters without any cognizant human planner or participant being aware of the problem the group is solving.

TGA argue that this is an example of emergent group level problem solving because the problem of the group diverges from the problem of the individuals, and therefore, the capacities ascribed to the group must differ from those applied to the individuals. Theiner et al. state that "the path systems are solutions to a problem for the group, not the individuals…The group as a whole can be aptly construed [as] solving a problem (Theiner et al. 2010: 385-6) Additionally, were this problem solved by an individual, we would count it as a cognitive process. This process is a paradigmatic example of correlated behaviour that leads to a specific outcome but which nonetheless falls outside the issue of joint action because the behaviour is not characterizable in intentional terms. I will return to the relationship between arguments based on these types of unintentional consequences and joint action in §7.7.

ii. Division of Labour in Groups

In an attempt to study the mechanism through which the members of a group coordinate their activities in order to reach an explicit group outcome, Roberts and Goldstone (2009) developed a game called "Group Binary Search". In this game, one hundred and six participants were divided into eighteen internet-connected groups. Their task was to determine a random number chosen by a computer between 51 and 100. Each participant entered a number between 0 and 50. The

computer summed all the numbers of the group members, compared this sum to the number it had chosen, and then gave feedback to the group, telling them whether their number was too high or too low, and in one variation of the experiment, by how much. The guess and feedback process was repeated until the group guessed the correct number. Each group played this game multiple times.

The relevant result is that the group's performance improved over time. With each iteration of the game, the group needed fewer rounds to guess the correct number. One of the reasons for this improvement is that the group members became strategically differentiated. Certain participants adopted 'conservative' roles in which they did not react to the feedback of the computer, while others adopted 'reactionary' roles and did respond to the computer feedback. The increase in predictability over the course of the experiment allowed the group to more efficiently solve the problem.

The point that Theiner et al. make is that groups may self-organize in order to solve a problem that involves coordination by differentiating roles. This group self-organization separates the individual problem-solving capacity from the groups'. The efficiency that the group develops is not the aggregate of the people's efficiency, and allows for the invocation of the social parity principle as it undertakes a cognitive process that would be considered cognitive if it had taken place in an individual.

iii. Transactive Memory Systems

The notion of a 'Transactive Memory System' (TMS) was introduced by Wegner (1986, 1995) and his colleagues (Wegner et al. 1985) in order to study the functional organization of memory in couples, families, and small workgroups. This notion relies heavily on the analogy between individual mental processes and group processes. Wegner states,

> The study of transactive memory is concerned with the prediction of group (and individual) behaviour through an understanding of the manner in which groups process and structure information…transactive memory draws deeply on the analogy between the mental operations of the individual and the processes of the group…A transactive memory system is a set of individual memory systems in combination with the communication that takes place between individuals (Wegner 1986: 185-206)

The experiments in this study attempt to understand the patterns of transactions that occur when people remember together as a group and the conditions under which these transactions lead to collaborative inhibition or facilitation.

The results show that groups containing members that know each other well, have had a long time to adapt to each other, and trust one another, seem to be able to form transactive memory systems that exceed the capabilities of the individuals considered separately. In fact, Theiner et al. claim that by dropping the individualist constraint, the TMS perspective reveals the general requirements for any memory system, namely, that (a) information is stored from experience, (b) there is subsequent access to this information, and (c) the accessed information is used to deal with related situations more effectively. Theiner et al. claim that "These functional requirements do not require that memories be the exclusive property of individuals, and in fact there are documented cases of collective memories satisfying these requirements better than individual memories" (2010: 389).

Theiner et al. then integrate these case studies into a general argument, which proceeds as follows:

> 1. The 'ability to cognize' is best thought of as the collection of a number of fairly well-understood capacities such as memory, attention, learning and problem-solving, which in the individual case are considered 'cognitive' processes.
>
> 2. There are cases in which groups collectively perform a function that involves the implementation of processes that, were they done in the head, would be considered cognitive processes.
>
> 3. Group level capacities for problem solving (or task performance) are distinct from those possessed by any individual members of the group
>
> 4. These processes and capacities also satisfy the following requirements: (a) they are emergent, in the sense of non-aggregativity, (b) they involve the differentiation of cognitive roles, (c) they introduce a level of cognitive organization that is an important explanatory kind with respect to the behaviour of the system.

Conclusion: groups can constitute cognitive systems in their own right.

§4 – Conclusion

In this chapter, I presented an alternative to Clark's account of organism-centered cognitive agency based on distributed cognition. I introduced the central tenets of distributed cognition as they are found in Hutchins and his hypothesis of enculturated cognition, which emphasizes the role of culture in distributing cognition over groups of people and over time. I then considered the philosophical argument for group level cognitive processes derived from recent empirical work in cognitive science that employs the distributed cognition framework. I pointed out that this work extends the argument of extended cognition to include coupled cognitive systems that involve individuals as their proper parts on the basis of the methodological principle that we should not *a priori* determine the bounds of cognitive systems, but rather consider which processes systems perform and then consider whether these processes conform to our pre-theoretical understanding of cognitive processes.

Part 3: Extended Cognition in Joint Action

Chapter 6 – First-wave Extended Cognition in Joint Action Theory

We have seen that the motivation for many arguments for the HEC is shared by the philosophers involved in the joint action debate; namely, they want to delimit the use of mentalistic discourse for the explanation and prediction of action. In the Otto experiment, Clark is concerned with the explanation for why Otto is at 53rd street at the Modern Art Museum, rather than 55th street, and turns to his mental states for an explanation, just as an action theorist would. Theiner is concerned with understanding whether the cognitive processing, planning and problem-solving that goes into solving collective tasks is best described as belonging to the entire group, a concern reflected in Gilbert's questions about the role of a plural subject in coordinating action. Moreover, the general proposal of the HEC is that cognitive systems are continually changing, context-, situation- and task- dependent collections that perform some cognitive task. Presumably, collective intentional behaviour, that is, the coordinated and intentional bringing about of some effect in the world, involves the completion of types of cognitive tasks, such as problem-solving, by collections of people that are equally variable and situation-dependent. The preceding part laid out several strains of the general idea of the HEC (the HEC of Clark & Chalmers, second-wave EC, distributed cognition and group cognition), each of which may have some application to joint action. In this part, I draw out some of the consequences of these theories for one another. This chapter deals with the HEC of C&C, which is the strain of the HEC with the most literature in relation to joint action. I will also consider the secondary literature on this issue, particularly the work of Deborah Tollefsen and Holger Lyre. The next chapter deals with joint action in relation to second-wave extended, distributed and group cognition.

§1 – The Importance of Coupling

The unifying component in all iterations of extended cognition is the concept of causal coupling. Let us return to some of Clark's statements on coupling to clarify what role this concept has. In a paper with Wilson (Wilson & Clark 2009: 68), we find the following claim:

> One key failure of the arguments supporting the extended-mind story, they [Adams & Aiziwa] suggest (this volume), is the failure of those arguments to distinguish mere causal influence from constitution. Now merely coupling a resource to an agent does not, of course, make it part of the agent. But this does not show the nature and degree of intercomponential coupling to be irrelevant to the question of constitution. What makes my hippocampus part of my cognitive system, it seems fair to say, has a great deal to do with how it is

> informationally integrated with the rest of my cognitive system. We can imagine a case in which, despite being firmly located in my head, there is zero integration and hence the onboard hippocampus fails to form part of my active cognitive system. Contrariwise, we can imagine a hippocampus in a distant vat whose activity is so well integrated as to unproblematically count as part of my cognitive apparatus (see, e.g., Dennett, 1978 - a classic treatment titled "Where Am I?"). Coupling, we conclude, does not in and of itself render a tool or resource part of the agent's cognitive apparatus. But the right kind of coupling (one resulting in deep functional integration) is a major part of what determines the scope and bounds of an agent's cognitive apparatus.

What is at issue here is the coupling-constitution fallacy, a debate which lies outside the scope of this thesis. However, we may take away several points from this quote. The first is the comparison between the brain and an extended system. Wilson and Clark point out that we may take the interactions of different subsections of the brain as mere causal influence or construe them as constituting a single system, and when we make this distinction, it is not the location of the particular subsystem of the brain that matters, but rather the degree of informational integration.

The central point is that coupled systems, in the context of extended cognition, are highly complex, non-linear interactions between neurological and external resources. The complexity of these interactions precludes the possibility of analyzing these systems based on a separation of the internal and external parts, although analysis of the individual parts in isolation remains important. Further, the processes these systems perform adhere to our understanding of 'cognitive processes'. This point is also present in Menary (2007), where he argues that continuous reciprocal causation and dynamic feedback loops between neurological and external resources are necessary for extended cognitive systems because they underlie the integrated nature of these hybrid systems. This underlying reliance on coupling is also present in the discussion of distributed cognition and group cognition, as under this account the relevant aspect remains the deep functional integration of the system. Group cognition does, however, change the focus on the importance of coupling *between people* rather than between neurological processes and external tools.

The last sentence of the quote emphasizes that coupling alone is not enough. What we need is the "right kind of coupling". The concept of coupling comes from dynamical systems theory, which is an established framework used in the physical sciences for modeling phenomena that change over time. Under the broadest definition, any system that changes over time is a dynamic system. This definition encompasses almost all physical systems. Where the technical apparatus of

dynamic systems theory exhibits the most explanatory power, is in systems which involve a certain level of complexity in their changes over time. An example of a coupled system in this context is two wall-mounted pendulums placed in close proximity on a single wall. The two pendulums tend to become synchronized over time because the vibrations running along the wall synchronize their swing times. This process is highly conducive to a dynamic systems analysis in which the motion of each pendulum is modelled by a set of non-linear differential equations. The pendulums are coupled because the motion equation for each pendulum includes a term that represents the influence of the other pendulum's current state. Many physical systems evolve according to this type of continual interaction where the variables of the mathematical description of one system act as the control parameters of the other system. As a result, the evolution of the system is best analyzed as a single larger system.

This background underscores the difference between coupled systems and *cognitive* coupled systems. The application of dynamic systems theory to cognitive systems is often a matter of the use of the conceptual scheme rather than a strict mathematical analysis involving mutual control parameters (Rupert 2009: 132-134). This is a divisive issue[36] in cognitive science and for simplicity we may follow Clark in positing a type of "dynamic computationalism" (Clark 2008: 27-29), which preserves much of the traditional language of representation and computation. The upshot is that what determines whether two distinct subsystems may be considered as a single larger system is the extent to which they mutually constrain one another over time through their constant causal interaction. Clark's conditions for cognitive tools are an attempt to provide a set of requirements for the level of interaction necessary to be considered a single system.

Nonetheless, the issue of coupling between people remains open, not only with respect to the issue of cognitive coupling, but also with respect to the nature of the difference in the interactions of various potential cognitive extensions. The cognitive coupling of individuals will likely require a different set of conditions as the issue of mutually constraining interaction is fundamentally different in cases involving two individuals than in cases involving bodily movements or external

[36] Divisive because it appears that the mechanistic nature of these explanations limits the usefulness of this framework for the representational and computational aspects of cognition. Indeed, many early dynamicists denied that cognition involves representation and computation (Thelan and Smith 1994; Kelso 1995; van Gelder 1995). This position is rejected by many theorists who attempt to incorporate dynamic aspects into a more traditional connectionist account (Clark 1997 ch. 6, 2001 ch.7, 2008 ch.1, esp. §1.8 and §1.9).

tools. In the rest of this chapter I explore the relationship between the search for the right kind of coupling between individuals and an analysis of joint action, against the background of first-wave extended cognition.

§2 – Collective Intentionality and Extended Cognition

The ways in which the theory of the mind present in extended cognition affects collective intentionality, and more specifically joint action, is not entirely clear, except in relation to Searle's account. The focus on the dynamical interactions between neural structures, body and world as pervasive and intimate, but more importantly for us, also action-orienting and behaviour-guiding, results in a conception of mind that includes external features actively participating in an organism's mental activity and cognitive processes. It is therefore in direct opposition to internalist conceptions of mind. The contrast of this active externalism to Searle's adherence to strict internalism precludes any possibility of his account benefitting from accepting the HEC[37], and as our concern here is elaborating the possibility of a benefit to joint action theories, Searle's account is left as a foil for the appropriateness of extended cognition for externalist accounts of joint action.

The possibility implied by distributed and group cognition of coupling between individuals is present in C&C in the 1998 paper (17-18) where they suggest that cognition could be socially extended to include systems with multiple agents or intimate relations between mental states of individuals. In the section, "Beyond the Outer Limits", C&C pose the question, "What about socially extended cognition? Could my mental states be partly constituted by the states of other thinkers? We see no reason why not, in principle" (17). This statement has inspired two possible interpretations in the domain of joint action: individualist and collectivist. On the collectivist view, taking minds outside of the head supports the possibility of the formation of collective cognitive systems with collective mental states, which is in line with distributed and group cognition. This removes some of the major reasons for the adherence to certain forms of

[37] As I noted, some extended mind theorists do not hold that accepting HEC necessarily involves rejecting brains-in-vats (Rowlands 2009: 56) and therefore does not completely reject the possibility of an internalist account such as Searle's. Nonetheless, it certainly does not support it.

individualism in the collective intentionality literature. One may argue that locating the mind outside of individuals and thereby creating the possibility of legitimate collectives with their own mental states and processes removes many of the motivations for these views. An analysis of joint action may then follow from the formation of a collective subject.

On the other hand, the extension of mind may be interpreted as strictly the extension of the mental states of an individual mind, as we have seen in Clark's account of cognitive agency. According to this view, every case of an extended mind requires a single 'locus of cognition', which is identified with the brain, ruling out the possibility of a collective agent. However, single mental states, such as a particular intention, may be externally individuated[38], and therefore linked to another extended system forming a type of intimate interrelation. If this is the case, the HEC may provide support for individualist accounts of collective intentionality in that it contributes a plausible story for how the intentions of individuals can be appropriately combined to produce a collective outcome.

In each of these two applications there are two issues which need to be considered, and which must not be conflated. First, an application of extended cognition to the literature on joint action must itself serve as an adequate account of joint action. And second, we may judge the interpretation based on its coherence in relation to extended cognition. That is, does the interpretation under question accurately portray the underlying account of extended cognition it requires?

§3 – Collectivist Interpretation

The collectivist view is put forward by Tollefsen (2006). She begins her account by demarcating collective systems from solipsistic systems (141). Solipsistic systems are composed of an agent and an artefact, such as Otto and his notebook, whereas collective systems are composed of coupled agents. She then argues for the plausibility of collective systems on the basis of a thought experiment that parallels the Otto-Inga example from C&C. Olaf and Inga are a married couple

[38] See Putnam (1975) and Burge (1979) for a discussion of social externalism.

that have been together for more than 30 years and the spend most of their time together. Olaf is rather forgetful and has difficulty remembering names, dates appointments and so on. Inga, on the other hand, has a sharp mind and reliable memory. As a result, Olaf relies on Inga to provide him with the missing information he needs to function on a daily basis.

This experiment is meant to show that collective systems may be functionally equivalent to solipsistic systems, because here Inga plays the same role for Olaf as the notebook plays for Otto. Like Otto's notebook, Inga fulfills the conditions required for a causally coupled system. She plays an active causal role in jointly determining Otto's behaviour and if we were to remove her from the system, Otto's competence and performance would suffer. Further, Inga fulfills the conditions[39] placed on non-biological artefacts:

1. Availability and frequency: Inga is readily available to Olaf, and Olaf typically consults her on a variety of details.
2. Endorsement: The information Inga provides is more or less automatically endorsed. In fact, Olaf uses Inga to check the reliability of his biological memory.
3. Accessibility: The information provided by Inga is easier to access than the information in Otto's notebook. Olaf merely has to ask a question, whereas Otto has to locate the information by searching through the pages of his notebook.

Tollefsen also emphasizes that the C&C argument relies heavily on the concept of functional equivalence. Otto and his notebook form a coupled system because his use of the notebook is functionally equivalent to his use of short-term memory. She notes that this separates the issue of the role that his notebook plays from how it carries out this function. It is only in how the notebook plays this role that the difference between his biological memory and notebook is evident, but according to C&C it is also irrelevant. This line of argument applies equally well to the extension from solipsistic systems to collective systems. Inga plays the same role for Olaf that Otto's notebook plays for Otto; the difference arises in how they perform this function, which is, as C&C claim, not pertinent.

[39] These conditions are adapted from Clark's (2004) work and differ from their expression in Clark (2008).

This still leaves us with the question of whether they actually do perform the same function. C&C propose that the functional equivalence is to be determined by the coarse-grained causal dynamics of the case, where causal dynamics is understood in folk psychological terms from our perspective as interpreters of behaviour. The relevant question for C&C is: will Otto's behaviour be guided by the information in his notebook, as it would be if the information came from his biological memory? In Tollefsen, the question transforms into: will Olaf's behaviour be guided by Inga's information, as it would be either from a notebook or his own memory? As is suggested by the fact that Inga fulfills the same criteria, the answer appears to be yes. Olaf will do and say the same things, and form the same mental states if the information comes from Inga, his notebook, or his biological memory. Although this idea entails a coarse-grained sense of causal dynamics and therefore threatens to embroil us in the debate about the parity principle, it fits nicely with our understanding of cases of individual non-extended behaviour. The folk psychological notions that we use to predict and understand behaviour are successful despite the fact that we know very little about the fine-grained causal mechanisms involved in the brain modules that are responsible for this behaviour.

Tollefsen notes that, somewhat unlike the Otto-notebook system, the Inga-Otto system could be understood simply as the transference of information between two different systems. To allay this concern, she argues that the retrieval and reconstruction of information may be jointly accomplished. Perhaps, while attempting to reminisce about an old acquaintance, Olaf remembers the first name and Inga the last name, or Olaf remembers where they met and Inga remembers why, and together they put together the memory. Through a process of joint deliberation and discussion Olaf and Inga can reconstruct information that plays an active role in their collective cognitive processes and in determining their behaviour. This is particularly relevant for the issue at stake in this thesis, in that certain types of cognition are here construed as processes that involve a plurality of agents.

Tollefsen presents one method of conceptualizing collective systems of individuals that are capable of bearing propositional attitudes based on the idea of causal coupling. In this picture, many cognitive processes, including those involving propositional attitudes such as belief, are carried out by an integrated system of neural, bodily, and environmental factors, where these environmental factors may also include other agents. In relation to a theory of joint action based

on joint intention, Tollefsen's account faces several problems. This is not a reason to reject her account, as, in that paper, she is not concerned with providing such an account. It is, however, important for the purposes of this thesis to point out its inappropriateness for an account of joint action. Specifically, I argue that the conditions for inclusion into a collective cognitive system as instantiated by Olaf and Inga fail to satisfy two of the conditions required for a complete theory of joint action as outlined in §2.8. In §7.8, I also relate some concerns with the relation between this account of collective cognitive systems and the notion of group cognition put forward by Theiner et al. (2010).

The problem with understanding Tollefsen's concept of collective cognitive systems in relation to a theory of joint action is essentially a scope problem, as the coupling conditions outlined are too restrictive. Responding to the fear that artefacts are too easily 'decoupled' from an organism to be included in our concept of the mind, C&C create their agent-tool conditions with the express purpose of emphasizing the extent to which the instrument is available and the frequency with which it is used, which results in the focus on trust, reliability and accessibility over an extended time. These characteristics do not represent the typical relationships between people who partake in joint actions. Cases such as Olaf and Inga are rather rare. It is not standard that another person is regularly present for the majority of our daily activities, and is constantly being asked the type of questions that lead to the joint formation of a belief, or the joint reconstruction of a memory. For example, we may take Clark's example of the dancers from §3.4. By all accounts of joint action, they have a joint intention to perform the dance. However, they may only interact during this performance and the practices that are required to perfect it. Dancer 1 is not readily accessible and frequently accessed as a source of information, nor is the information she provides immediately endorsed, except for the time period in which they are performing the dance. They then fail all of the conditions for being part of a coupled system outside of a very specific and limited time period. For this time, on the other hand, they appear to be intimately physically and cognitively coupled.

Participation in a collective system in Tollefsen's sense is therefore not a necessary condition for the formation of a joint intention, nor for the potential group cognitive processes required to perform such an action, because there are many cases of collections that have a joint intention or group processes that do not fulfill these requirements for a coupled system. Tollefsen's account

therefore fails to satisfy the scope condition (§2.8). This result is not surprising for two reasons. First, the conditions here presented for the creation of an agent-agent system are, in the pertinent respects, the same as those for an agent-tool system, which as we saw are boiled down to "a high degree of trust, reliability, and accessibility" *over an extended time*. The intuition informing this move is that Inga is an imperfect, although perhaps in some ways superior, type of computing instrument, which is not the case in the standard cases of joint action. Second, Tollefsen is not attempting to provide a theory of joint action at all, but merely to show that the HEC leads to the logical possibility of collective cognitive systems where the mental states/processes of individuals who are integrated in the proper way may serve as active components of the mental states/processes of other individuals, and the collective as a whole may be attributed mental states. Regardless of the appropriateness of her account for joint action, the logical possibility of a collective cognitive system is an important result for what follows.

§4 – Individualist Interpretation

As we will see, there is a tension between this collectivist interpretation and Clark's account of organism-centered cognitive agency because the collectivist interpretation does not consider collective cognitive systems as the extension of the cognition of a single organism. If we are going to apply the HEC as imagined by Clark to the issue of joint action, we must then develop an individualistic account that is consistent with Clark's human-centered theory of cognitive agency. Such an account has been proposed by Holger Lyre.[40] He argues that the conditions that Tollefsen applies are wholly unsuited for analyzing social groups that perform joint actions, because very rarely do joint actions result from collections of people who would meet these requirements. Therefore, in order to analyze joint action, we need to import a separate set of conditions for creating coupled systems between multiple agents, who themselves may be seen as cognitive systems in their own right.

Lyre recognizes that Clark's account of cognitive agency is thoroughly individualist and therefore that the nature of the HEC concerns the cognitive enhancement of individuals. On his

[40] Nothing published. These views are attributed to him on the basis of a presentation given at the University of Vienna on 18.4.2013.

view, each of the possible extensions of cognitive processes requires its own separate conditions for coupling; that is, the neural structure of the brain may be linked to some combination of the body, natural environment, cognitive tools, the social environment or other people, but each of these extensions requires its own set of coupling conditions. Lyre states (2013: 3):

> in order to avoid rampant extension, it is important to specify restricting and limiting coupling criteria that must be met by external components of a cognitive systems…such criteria will roughly comprise accessibility, robustness, and reliability of the external components (and they must be spelled out in detail for all levels – a task still open for proponents of EC [Extended Cognition]).

This argument holds strong intuitive weight. The way the neural processing is connected to my physical gestures is surely significantly different to the way it is coupled to a notebook. This claim is also supported by theorists within the extended cognition literature (Sutton 2006, 2010; Barnier et al. 2008; Wilson and Clark 2009). Sutton, for one, suggests that the introduction of specific criteria for various potential cognitive extensions not only takes the explanatory weight off the parity principle, but also gives extended cognition a wider explanatory scope. He claims that (2010: 198):

> the existence of a number of distinct dimensions on which particular cases [of potential cognitive extensions] can differ is the sign of a promising, multidimensional space for doing EM-inspired cognitive science, by developing taxonomies or typologies of external resources in use, or of coupled systems (2010: 198).

Lyre's proposal is then an attempt to develop one type of cognitive coupled system with this more general and inclusive framework.

The conditions of C&C would then only apply to cognitive tools, which is the way these conditions are originally envisioned in their paper. Lyre proposes that shared intentionality constitutes one of the most decisive mechanisms of coupling for extended social cognition (2013: 8, fn.1). Based on the idea that the processes the take place in social interaction are driven and partially constituted by environmental scaffolding, Lyre argues that the vehicles of cognitive extension may be found in the social environment, which he takes to include linguistic practice, gesture, touch, facial expressions and other social signals and cues, such as joint attention, and behaviour reading. He then suggests that Bratman provides the requisite analysis of shared intentionality to provide the link between social and extended cognition. His aim is to, "propose a new understanding of the mechanisms of extended social cognition – with Bratman's well-known conditions as coupling conditions" (2013: 8, fn. 1).

Lyre's proposal that we take Bratman's conditions for joint intention as the appropriate coupling conditions for extended social cognition is a more radical instrumentalization of the other participants in a joint action than is found originally in Bratman. According to Lyre, the mental states of the other participants are merely instruments for the enhancement of my own cognitive processes. They allow me to possess intentions of the form "I intend that we *J*", by explaining how the content of my intention may include another person's cognitive states. This approach necessitates a reversion into the C&C conception of social cognition as the extension of an individual's cognitive processes with the help of a partner. The first issue with this proposal is that treating the other participants as tools to enhance individual function seems to misrepresent what is going on a in a joint action. There is an important sense in which I cannot do something with my tools *together*. Inga might perform the same function for Olaf that the notebook performs for Otto, but in the Inga-Olaf case there is a sense of sharedness in each participant arising from the necessary coordination of actions, beliefs, goals and intentions, that simply cannot exist between Otto and his notebook. It rings false to characterize Inga as merely a cognitive enhancer for Otto and this is what Lyre's account proposes.

This concern does not yet however amount to an argument. I argue that this proposal fails for two reasons. The first is related to the concern above and is discussed here; the second is discussed in the next section which deals with the relationship of both Tollefsen' and Lyre's accounts to dynamic systems theory. The central problem with Lyre's approach is that, as a theory of joint action, Lyre's proposal fails to rectify a central problem with Bratman's theory dealt with in Chapter 2. If we conceive of the social extension as Lyre suggests we are left with the problem of circularity. Imagining each participant as coupled with separate mental states of another participant does not account for the collectivity in the content of the mental state; it merely enhances the ability of the subplans to mesh. Being able to incorporate another person's individual mental states does not provide an expression of the collectivity involved in an intention of the form 'I intend that *we* J' because it does not account for the 'we' that must be in place for there to be intentions of this form. It still seems to be the case that we need an intention of the form 'We intend to J' before we may form an 'I intend that we J' intention, and couching the 'I intend' in extended cognition language does not clear up this conceptual difficulty.

§5 – Extended Cognition in the Individualist and Collectivist Interpretations

In this section, I consider the two interpretations of coupled systems of individuals against the background of dynamic systems theory. I argue that if we recognize the importance of total state explanations inherent in dynamic systems theory, we must adopt a version of the collectivist interpretation, but not necessarily the interpretation proposed by Tollefsen.

The main argument for this position is as follows:

> (1) A basic principle of coupled systems is that the total state is the emergent property of the dynamic interactions of the system taken collectively
>
> (2) If we accept that joint actions may be performed by collections of individuals who constitute a coupled system, we either need to attribute the relevant states to the system as a whole, or introduce a distinction, such as Clark's in the case of neural processes, that somehow gives priority to one component or process.
>
> (3) There is no reason to introduce such a distinction in the case of genuine small-group joint action.
>
> Therefore,
> (4) If we are going to use the concept of coupled systems to understand joint action, we need to adopt the collectivist interpretation and attribute the relevant states and processes to the entire system.

Proponents of the individualist interpretation may argue that (2) is false on the grounds that the coupled system is constituted by an individual and the mental state(s) of another participating individual. I take this to be implausible because a joint action would then involve a large number of separate coupled systems. The neural processes of each participant would have to form a distinct coupled system with each of the relevant neural processes of each of the other participants resulting in a complex of interrelated coupled systems that nonetheless do not themselves constitute a single system. There may of course be subsystems within the larger system, which also may be important for a thorough understanding of a joint action. However, if we consider the individuals to be coupled the total state supervenes on the larger single system.

With regards to (3), the distinction between the recruitment of the system and the system as a whole, as is suggested by Clark for individual cognitive systems, will not work in the case of joint action. The recruitment of participants (components) of the system is not necessarily organized by any particular participant, as may be the case with the brain in an individual cognitive system. There may be such cases, but they are not paradigmatic instances of joint action and therefore should not be considered fundamental to the initiation of a joint action. Further, the simultaneous and mutual influence of the participants on one another limits the value of treating each participant as a prime-cognizer that extends her cognition by treating the mental states of the other participants as tools to enhance her own cognitive capacities.

The collectivist application proposed in (4) supports the claim that if the behaviour in question is group behaviour displaying the appropriate level of organization and intelligence, then the cognition required for the performance of this behaviour, should equally be attributed to the system as a whole, not to the interrelated properties of separate individual parts. Therefore, an application of this thesis to cases of joint action supports the ascription of cognition to the collective.

This claim seems to support Tollefsen's account of collective cognitive systems. However, her direct application of the coupling conditions for cognitive tools to cases involving multiple people causes problems for the range of cognition that her account allows collectives to have, and the kinds of collective that may have them. This issue will arise again in the next section in relation to joint action, but it is also a problem for group cognitive properties in general. In the discussion of group cognition in §5.3, we saw that a range of groups may undertake certain cognitive processes such as memory, attention and problem-solving. Many of the collectives considered do not meet the coupling conditions adopted by Tollefsen, because the members of these collectives are only together for the duration of the experiment, and are therefore significantly different from cases such as Olaf and Inga (with the exception of certain couples in the Transactive Memory System experiments). It would be premature to rule out many cases of potential group cognitive processes simply because they fail to meet conditions designed for cognitive tools. Tollefsen's use of the coupling conditions from Clark are therefore not only

inappropriate for a theory of joint action, they are inappropriate for group or collective cognition in general.

§6 – Conclusion

In summary, although both the existing attempts to apply the HEC to an analysis of joint action face important problems both as theories of joint action and interpretations of extended cognition, they reveal several valuable lessons for further attempts. First, Tollefsen's account presents an example of the logical possibility of collective cognitive systems and shows that a suitable analysis of the coupling conditions of agents requires a separate set of conditions from those that C&C use for non-biological tools. Lyre's analysis strengthens the second claim by generalizing the need for distinct sets of coupling conditions to all putative extensions, whether bodily, environmental, tool-based or social, and shows that an attempt to provide these conditions according to Clark's account of cognitive agency treats joint action as a matter of instrumentalization of the other participants.

Ultimately, the problems for both accounts lie in their adherence to different aspects of Clark's account. Tollefsen straightforwardly adopts Clark's coupling conditions, which limits the range of her account of collective cognitive systems, both for joint action and many group level cognitive processes in general. She is, however, correct in her rejection of Clark's hypothesis of organism-centered cognition, which brings her account into line with the concepts of dynamical systems theory. Lyre has the opposite problem. He recognizes that the coupling conditions Clark proposes are only relevant for cognitive tools, and therefore rejects them outright, but adheres to Clark's account of cognitive agency, which, in the case of coupled individuals who perform joint actions, leads to an implausibly complicated account of joint action and a conflict with the idea of total state explanation.

Chapter 7 – Plural Subjects as Collective Cognitive Systems?

In this chapter, I focus the analysis of the relationship be the HEC and joint action by exploring one specific area where the framework of HEC, specifically the use of the concept of coupled systems, may be beneficial to a particular joint action theory, namely the plural subject theory of Margaret Gilbert. This approach leaves many potential aspects of second-wave, distributed and group cognition out, but represents the most profitable usage of the idea of collective cognitive systems for joint action. I argue that we should discard the project of supplying necessary and sufficient conditions taken from the collective intentionality literature for all cases of group cognitive processes on the grounds that conditions meant to specify cases of joint action are not directly applicable to the empirical research into the nature of cognition, because they employ different standards for the ascription of intentional states. However, I also explore the relationship between joint commitment and cognitive coupling and argue that a coherent view of joint action as a dynamic process involving various levels of coupling is consistent with, and provides support for, Margaret Gilbert's theory of plural subjects. To do this I introduce the idea of low-level coupling in several domains, which is consistent with, if not based on, Menary's idea of dynamic biocausal integration. Finally, I argue that based on the nature of joint actions there is significant overlap between plural subjects and collective cognitive systems.

§1 – Coupling Conditions for Group Cognition

Despite the problem that first-wave extended cognition faces when applied to cases of multiple agents, the general extended cognition framework in its distributed/group cognition form may hold more promise. Further, as this discussion also touches upon the issue of coupling between individuals, many of the results of our previous discussion about conditions of coupling remain relevant. In a 2010 paper, Wilson discusses such a framework in the context of collective vision. He argues that an account of coupling between individuals should satisfy three points. First, two (or more) elements are *coupled* just in case they exchange information by means of reliable, two-way causal connections between them. Individuals who are collectively coupled are therefore interdependent in their cognitive and behavioural activities. Second, two (or more) coupled elements form an *integrated system* in situations in which they operate as a single causal whole

within the causal nexus – with causes affecting the resultant system as a whole, and the activities of that system as a whole producing certain effects. Third, an integratively coupled system shows *functional gain* just when it either enhances the existing functions of its coupled elements, or manifests novel functions relative to those possessed by any of its elements. Wilson emphasizes that the two aspects of functional gain are important for an account of collective cognition, because they imply that the cognitive interdependence between people has both individual-level as well as group-level effects.

This account of the conditions for coupling between individuals shifts the focus to the second-wave cognition concepts of integration and complementarity. In Menary's definition of cognitive integration, he considers three levels of potential integration. One is the dynamic integration of mutually constraining systems over time. This idea will be central to the account developed her. The task for us is to outline the way in which collections of individuals come to demonstrate these features, and the extent to which these two features must be present in order for there to be a genuine collective cognitive system. In short, is this all there is to the "right kind of coupling" for individuals, and if so, how does it come about? And also, does this type of coupling lead to genuine collective cognition in all cases of joint action?

§2 – Normative Commitments as 'Glue and Trust'

Wilson's discussion provides us with the following three standards as the starting point for considering another set of coupling conditions between individuals: information exchange, integration (in terms of causal relations), and functional gain. These serve as a solid basis for analyzing coupling between individuals, however, they remain only a starting point, as they fail to provide any insight into the conditions under which such coupling may come about and be maintained.

What is missing from Wilson's standards is an account of the 'glue and trust' that bonds the individuals of the extended cognitive system together, which allows them to demonstrate properties such as informational integration and functional gain. As we have seen, the conditions

of C&C for the cognitive integration of tools are too restrictive, but their focus on accessibility, availability and reliability are surely correct, if the extended system is to perform according to Wilson's standards. If individuals are to be coupled together, they must display coordination of their various cognitive processes, have frequent access to each other's cognitive states and the trust that the cognitive states of the others are not deceptively intended. I propose that between individuals, these conditions are frequently met using normative commitments, and therefore, instead of looking to Bratman, as Lyre does, we should look to Gilbert to find the potential specific coupling conditions for individuals, because for individuals coupling is not purely a question of frequent access, previous endorsement and so on, but essentially one of trust and reciprocation. It may be the case that for us to have these responses to other people we need something more fundamental, something with a normative component, namely, mutually binding commitments, or in other words, Gilbert's concept of joint commitment. Under this approach, normative commitments are understood as practical methods for the mutual constraint of the behaviour of the participants in the joint action over time.

The conditions of C&C are too restrictive because they purport to be in continuously in effect, in a way that precludes the possibility of people conforming to them except in very rare cases. Simply put, only extremely interdependent couples spend that much time together. This works in the case of tools because there is a specific 'prime-cognizer', an individual cognitive agent, that can carry the tool around, and who, in a sense, *drives the cognition*. One reason why these conditions do not work for collections of individuals is that there is no one particular individual driving the cognition. The bias comes through in the examples we have seen in Clark, and is briefly commented upon in the discussion of Tollefsen. However, in the case of joint action, there is another force which *drives the cognition*, namely, the action itself. After individuals jointly commit to performing some action, they are committed to the shared cognitive processes that result in the completion of that action.

There are many ways to conceive of such processes. One such way is to see these processes in terms of group cognition, as they may include the group processes of problem-solving, memory, attention and so on. The conditions of availability, frequency and reliability are satisfied *during the course of the action*. Actions unfold over time and collections that perform joint actions also unfold over time with a complex pattern of interaction, in which the cognitive states and sub-

actions of one individual serve as constraints on other individuals, informing their sub-actions and sub-states. This highlights the appropriateness of using concepts taken from dynamic systems theory in the explanation of these actions. Further, it is understood among the participants in a joint action that making these states available to the other participants is required for the completion of the action. Individuals working together on a task possess different kinds of knowledge and perform various sub-tasks that are required by the action. As a result, they engage in interactions that allow them to pool their various resources and share information. In addition, individuals in a cognitive system have overlapping and shared access to knowledge that enables them to be aware of what the others are doing. This enables the coordination of expectations to emerge from verbal and non-verbal communication, habitual roles and shared symbolic representation. The joint action serves as the explanation of the sub-actions; it also serves as the explanation of any collective cognition that is required for the performance of the action. If we conceive of the scope of the requirements of glue and trust in relation to specific tasks, they become satisfiable for collective cognitive systems, but in order to do so, these collections must jointly commit to some goal and be entwined in a network of obligations, entitlements, rights and commitments to ensure that the requisite integration and interrelation is maintained over the course of the action. As we will see however, the issue of coupling between individuals contains more than coupling through joint commitment.

§3 – Levels of Coupling

There are several levels at which coupling between individuals may occur. The hypothesis presented above suggests that high-level commitment to a shared goal is the proper mechanism for coupling between individuals in many cases, but there are several other possibilities, such as lower-level cognitive processes like attention and posture, that must be considered. The role of normative commitment in coupling must therefore be clarified.

Tollefsen et al. (2013) introduce the concept of an ‘alignment system’ based on linguistic and psychological research in conversation and interpersonal interaction. An alignment system is a loosely connected set of cognitive processes that facilitate social interactions. This system may

include both low-level and high-level cognitive coupling. In the psychological literature, alignment refers to the dynamic 'matching' of the behaviour or cognitive states of two or more people over time, for example, in their gestures, gaze, word choice or posture. Tollefsen et al. distinguish between genuine alignment which involves continual mutual adaption, and mimicry, which is the simple matching of behaviour in a single instance and is often asymmetrical. The importance of the concept of an alignment system is that it involves the coordination of behaviour over time and is achieved through mutual responsiveness. As a result, it is a genuine instance of coupling, which mimicry is not. Nonetheless, mimicry is the bedrock upon which alignment systems may be built. Tollefsen et al. argue that (2013: 51),

> Although we conceive of alignment as distinct from mimicry, our capacity to form coupled systems likely relies on the basic ability to mimic the behaviour of others. Low-level mimicry and basic priming mechanisms that may generate mimicry probably help to start and sustain mutual adaptiveness (Tollefsen and Dale 2012). What we wish to highlight is that the integration of these low-level processes, contextual variables, high-level cognitive plans and so on, sustains a robust pattern of interaction between human beings when they interact.

Tollefsen et al. report empirical research that suggests that during natural conversation many sub-personal alignment processes are at work and that the degree of alignment influences the success of many interpersonal processes, such as learning, information exchange and communication. Further, the research suggests a connection between behavioural alignment (posture, eye movements) and linguistic alignment (syntactic and semantic). Tollefsen et al. conclude that "Behavioural alignment seems to give rise to alignment in conversation, which, in turn, gives rise to a mutual understanding and deeper understanding of one another, which amounts to an alignment of overall interactive comprehension" (2013: 52). According to this view, interaction between collections of individuals over time is greatly facilitated by low-level cognitive processes and higher-level cognitive plans.

Alignment processes also contribute to the successful completion of collaborative tasks. Tollefsen et al. point to research that suggests that aligning with others in synchronized movements such as dancing and marching improves both the perceptual and motor ability during cooperative tasks and enhances the general rapport and pro-social behaviour of the participants. This also holds for cases in which the behaviour is *complementary* rather than matching. In a study by Richardson et al. (2007), as reported by Tollefsen et al. (2013: 53), participants were asked to move objects from one part of the laboratory to another. They observed that when the objects reached a particular size, participants spontaneously organized themselves into a

complementary perceptuomotor unit and moved the object together. Upon further analysis, Richardson et al. discovered that the point at which the participants formed a collective to pick something up was very similar to the individual case of a person deciding whether to use one hand or two. In the mathematical analysis they put forward, the dynamics of the joint action case match the dynamics of the individual case in surprising detail. As this bears a direct relation to Gilbert's theory of acting as a single body, we will return to this issue at greater length in section §7.5.

Coupling based on lower-level processes occurs unintentionally. The participants in these studies are often not aware that they are bodily or linguistically entwined. This raises the question of the relationship between the lower-level coupling explicated in this research and the higher-level coupling suggested by the concept of joint commitment. It may be argued that larger social behaviours may be brought about purely by sub-personal processes, making recourse to higher-order cognitive processes such as joint intention or commitment explanatorily superfluous, or at least secondary. This contrasts not only with Gilbert's analysis, but all accounts of joint action, as they all rely on higher-order cognitive states that are shared in some way. We will discuss the potential difficulties of applying this research to philosophical accounts of joint action in §7.6.

In an earlier paper, Tollefsen and Dale (2012) present the general thesis that alignment systems underwrite all joint action, but that this does not preclude the importance of high-level coupling. They explore four possible areas of the interaction between these various levels of coupling: mutual amplification, dimensional compensation, misalignment-needs-commitment and illusion of we-will.

Mutual amplification suggests that in experiments that induce alignment at a lower-level, we should also see increased we-intentional coupling (coupling of a higher-order) including joint intentions, feelings of solidarity and mutual expectations. In addition, when commitment is initiated in prior planning for a joint task, mutual amplification predicts an increase in lower-level alignment when compared to cases where no such commitment is introduced. The concept of dimensional compensation refers to the way in which commitment may arise from certain aspects of lower-level alignment. Tollefsen and Dale suggest the possibility that when lower-level alignment is restricted by experimental conditions, collectives may turn more directly to explicit

commitment to coordinate and guide their actions. This leads to the idea of 'misalignment-needs-commitment', which states that commitment may be necessary to identify and correct the ways in which individuals become misaligned. Finally, Tollefsen and Dale suggest that joint actions may be subject to something they, following Wegner (2003), call the 'illusion of conscious will'. The hold that in certain situations, "low-level alignment systems, anchored to surface synchronies in contingencies, could lead to illusory cognitive contracts of the kind described in philosophical theories" (2012: 403). They recognize the ambivalence in their final two suggestions and conclude that "the theory predicts two quite different aspects of higher-level deep commitments. In the former case [alignment-needs-commitments], they [deep commitments] are necessary components of performing complex non-aligned patterns of behaviour or joint activities; in the illusion of the we-will, they are epiphenomena, misattributed to original intentions" (2012: 403). They do not resolve this apparent tension and as I result I will return to this issue in the next sections.

§4 – Coupling and Joint Action

Tollefsen and Dale (2012) also consider the relation between this view of coupling and an account of joint action. They suggest that incorporating the empirical research that points to coupling as a mechanism of cohesion between individuals serves several purposes for the literature on joint action, which I combine into two related groups. First, joint action theories informed by this research are better equipped to deal with how people come together to act jointly, that is, how joint actions are implemented, executed and initiated, and second, the empirical research may help to adjudicate between the various theories of joint action (2012: 389-391).

Tollefsen and Dale attempt to demonstrate the first role they see by using their concept of alignment systems to supplement Searle's account of joint action (2012: 397-401). They argue that lower-level cognitive processing fills out Searle's concept of the Background. According to Searle, the Background is a set of non-intentional capacities that is presupposed for seeing the other as a candidate for cooperative agency. Tollefsen and Dale argue that, "One way to conceive

of this set of capacities, however, is to understand them as structures or features of an alignment system" (2012: 398). Their central point is that alignment systems provide the necessary structure for the formation and continuation of we-intentions. In relation to Searle's example of the ballet troupe (§2.1), they see the relationship thusly (2012: 398),

> The ballet troupe's higher-order we-intentions will inform their lower level processes and explains how their perceptual and motor systems can function together to achieve their goal. Similarly, the presence of an alignment system explains how we-intentions can be formed on the fly, so to speak, without prior planning or agreements.

Further, the lower-level alignment explains the way in which we-intentions in individual minds lead to unified agency. The coordination of minds and bodies over time is facilitated by the alignment system, and therefore does not need to be present in the we-intentions themselves, which, according to Tollefsen and Dale, allows Searle's account to avoid the criticisms raised by Meijers (2003) and Schmid (2009) discussed in §2.7.

The choice of Searle's account as the joint action theory to which they apply their concept of alignment systems seems ill-advised. As we have seen, Searle's two conditions for collective intentions stipulate that all the relevant cognitive states and processes are had by individuals, and that they exist completely within individual minds. The concept of an alignment system, when construed as a potential candidate for group cognition, as it is in the 2013 paper, violates both of these conditions in the extreme. First, in order to couple in the way Tollefsen and Dale describe, individuals have to stand in particular relations to one another. In fact, much of the work done by the concept of an alignment system involves describing these interrelations in detail and according to Tollefsen and Dale, this interrelation is structurally necessary for the formation of a we-intention. In Searle's theory, however, the necessity of any external relation is categorically ruled out. Second, Searle denies that any group or collective entity could possibly have cognitive states. Tollefsen et al., on the other hand, argue that "an alignment system can provide the sort of integration necessary for cognitive systems" (2013: 61). The concept of an alignment system therefore seems to be inconsistent with Searle's two fundamental requirements for statements about collective intentionality.

Nonetheless, in their discussion, Tollefsen and Dale present a novel concept of joint action, which may be consistent with another joint action theory. Rather than focus on joint action as a product that is static in time, they conceive of it as a dynamic, self-organizing process that

unfolds over time. This brings their conception into line with Hollan et al.'s admonition that when choosing the proper unit of cognitive analysis we must find, "a system that can dynamically configure itself to bring subsystems into coordination to accomplish various functions" (2000: 175, see §5.1 for discussion). It also places the emphasis on the way in which joint actions are initiated and sustained. To explain this, they elaborate on the types of lower-level cognitive processes that lead to coupling between individuals and their relation to higher-level cognitive states. They have developed a consequential area of overlap between group cognition and collective intentionality, and suggested several potent areas where the use of empirical research on the coupling of lower-level cognitive processes could be highly valuable for theories of collective intentionality. In the next section, I argue that their framework is consistent with, and solves both of the problems we outlined for, Margaret Gilbert's theory of plural subjects.

§5 – Plural Subjects and Alignment Systems

Tollefsen and Dale argue that while deep commitments and joint intentions have an important role in joint action, they are not sufficient for joint agency to be successful because they must be supplemented by alignment processes at a lower level. In addition to not being a sufficient condition for joint action, Tollefsen and Dale do not fully commit to positing joint intention as a necessary condition. Additionally, they do not differentiate clearly between the role of 'deep commitments' and 'we-intentions'. They do, however, claim that "we-intentions of some form seem to be the mark of joint *agency*" (2012: 400, emphasis in original). These pronouncements leave much to be explained concerning the role of intention and commitment in joint action.

Much of this may be clarified by subsuming the role of lower-level coupling processes and alignment systems under Gilbert's account instead of Searle's. The role of gaze, posture and other such processes fills out Gilbert's claim that joint commitment involves an 'expression of readiness' by providing detailed mechanisms that introduce a collection of people to each other as possible subjects of such an expression and expands on the ways such an expression may take place. Gilbert recognizes that an expression of readiness "may take various forms, …[which] correspond to those contexts in which people come to be doing something together", may be

manifested in action, rather than verbally explicit agreements, and "may emerge gradually, over time" (2006: 139-140). Gilbert's account of common knowledge includes the proviso that it "is not to say that each party to a joint commitment must have been directly *aware* of each party's expression of readiness" (2006: 139). One example she discusses concerns a spontaneous group forming to help a crash victim, in which it is clear that an expression of readiness may emerge from a more basic level of shared behaviour, but it is not a joint action until it is present (2006: 139-140).

These features of the 'relevant expressive behaviour' may be explained by the various lower-level processes that Tollefsen and Dale describe. For example, Tollefsen and Dale consider studies which show that many of the processes that support conversation are sub-personal alignment processes, such as bodily posture (Shockley et al. 2003), verbal cues (Schockley et al. 2007), and various other levels of linguistic organization, from diction to sentence structure (Tollefsen et al. 2013: 51). They also discuss the role of priming in alignment, which predicts that the cognitive accessibility of many behaviours, such as a chosen sentence structure, is induced by hearing another person use it, and thereby increasing the probability of producing a similar behaviour oneself. Many of these processes are at work in the information exchange that leads up to an expression of readiness, and therefore may be a relevant explanation for how an expression of readiness comes about, the forms it may take, and the way in which it is received. The central point here is that under this view, because lower-level alignment processes play a significant role in creating and explaining an expression of readiness, and an expression of readiness is necessary for a joint commitment, lower-level cognitive coupling may serve in many cases as the foundation for the existence of high-level cognitive commitments.

This leaves open the question of the *necessity* of low-level alignment processes in the case of explicit agreement. Gilbert holds that an expression of readiness may be either implicit or explicit, and while it seems that in the implicit case low-level coupling will play a significant role, in the case of explicit agreement based on an exchange of emails, for example, we may not require these processes. It appears that low-level coupling and alignment systems are not necessary conditions for all expressions of readiness because we may have an expression of readiness that does not include lower-level coupling (email, phone conversation). Nonetheless, they may still be necessary conditions for a non-verbal, non-explicit expression, such as in the

case of the bystanders rushing to aid a crash victim, because these types of agreements seem to need to take place in person and involve a gradual non-explicit exchange of information.

In any case, low-level coupling is surely not a sufficient condition for joint action. There are many cases of lower-level coupling that do not meet the requirements for joint action as here defined, some of which are discussed by Tollefsen and Dale. They discuss, for example, cases such as 'joint following', in which subtle directional cues cause two or more people to wander in a direction that neither intended, simply by "following" each other. This case involves many surface-level coupling processes and leads to a collective outcome, but it is not a case of joint action. We, therefore, need something more than a concept of coupling derived from empirical research to explain genuine joint action, because the story of alignment processes leading to an alignment system does not adequately differentiate between collective intentional behaviour and simply aggregative collective outcomes. In order to make this differentiation, we need to maintain the appeal to a sufficient condition for a joint action. Here we may simply appeal to the account given by Gilbert. As we have seen, a joint action involves joint commitment to intend to X as a body, and therefore in our language, higher-level coupling. Under this view, surface level-coupling may begin before there is a joint action; it may 'jump start' a spontaneous joint action, but there is no joint action until there is, in addition to dynamic alignment coupling, higher-level coupling based on joint commitment. Low-level coupling may be a necessary condition for implicit (as opposed to explicit) expressions of readiness, which in turn is sufficient for a joint commitment, but it is joint commitment that remains necessary and sufficient for a joint action.

One important aspect of this view is that it emphasizes the dynamic nature of joint actions, which in turn focuses the analysis on the fact that they must be started, sometimes spontaneously, and that they must be sustained, and that these two processes are not necessarily the same. The inability of joint action theories to differentiate these two aspects and account for spontaneous joint action has been criticized from within the joint action literature by Baier (1997: 42-43). It also figures largely in the criticisms of Gilbert that her account is circular because it requires collective intentionality for the creation of a joint commitment. This circularity may be partially explained away by appealing to alignment processes and alignment systems. We have seen that by providing elaboration of how joint commitments may come about through visual, linguistic or bodily coupling, and clarifying how higher-level commitments and lower-level alignment interact

to initiate and sustain joint action, alignment processes specify mechanisms through which plural subjects come about by means of an expression of readiness, which involve more basic forms of collective phenomena. This appeal to lower-level, non-intentional phenomena renders the circularity non-vicious, although, it must be noted, only for cases of implicit expressions of readiness. Explicit expressions of readiness involving direct conversation still seem to involve some collective intentionality.

Further, dynamic explanations of self-organizing behaviour demystify the idea of acting together as a body, which helps clarify how joint actions are maintained. The key concept here, which links the analogy of an individual acting through his body and a plural subject acting as a body, is the idea of 'coordinative structures'. In the case of a single individual, 'coordinative structure' refers to the ability to restrict a large number of degrees of freedom, in for example muscle groups, which allows them to self-organize into cohesive functional units without direct central control. The idea is that the degrees of freedom are reduced through mutual constraint among body parts so that specific, task-oriented muscle patterns may be differentiated, for example, between shooting a basketball and hitting a golf ball. In a recent flurry of literature (Richardson et al. 2007; Shockley et al. 2009; Ramezoni et al. 2011, 2012), this concept has been applied to the completion of joint tasks. The general claim is that joint tasks and the social interactions of individuals participating in the completion of these tasks involve the introduction of gradually developing mutual constraints across two or more people's bodily and cognitive states as a kind of 'joint coordinative structure'. This body of literature suggests that interpersonal coordination is governed by the same, or at least highly similar, general laws and principles that govern the coordination of an individual's actor's movements. However, several of these studies deserve closer attention and I would like to highlight three of them here.

First, Richardson et al. (2007) studied both the intentional and unintentional interpersonal coordination of two people sitting side-by-side in rocking chairs. They performed two experiments. In experiment 1, the subjects were told to intentionally coordinate their rocking, while in experiment 2, they were told to rock at a pace they felt comfortable. In both experiments, visual information was controlled. The subjects were either told to directly attend to what the other participants was doing (focal condition) or to look straight ahead (peripheral condition). They found that in both cases the movements were constrained by the self-organizing dynamics

of a coupled oscillator system and that "interpersonal rhythmic coordination is constrained by the same dynamical entrainment processes as intrapersonal rhythmic coordination" (2007: 869). By studying rocking chairs, which involve incidental rhythmic movement, they are able to argue that rhythmic synchrony may be automatic. On the other hand, they also discuss literature (Schmidt, O'Brien, and Sysko 1999) in which more deliberate movements are involved and in these cases alignment only occurs when the social context demands it. This suggests that based on the movement we are dealing with alignment may either be automatically initiated or initiated by a higher-level phenomenon. They conclude that "stable coordination has less to do with the physical qualities of the coupling medium [visual, auditory...] and more to do with whether the coupling provides the appropriate information about the movements in question" (2007: 884). Interestingly, in experiment 1, the coordination observed was the same for the peripheral and focal conditions, while in experiment 2, the level of coordination dropped in the peripheral condition. This suggests that sharing a joint intention plays a role in lower-level coupling, by changing the participant's attention and information gathering ability.

Shockley et al. (2009) review the literature on interpersonal coupling and coordination and suggest that "interpersonal coordination is a case of an emergent coordinative structure", thus connecting individual bodily actions and the actions of a plural subject, acting as a single body. There is, however, a conflict between the role of representations in coordinating behaviour in plural subject theory and in dynamic systems analyses of interpersonal coordination, which we see when considering their discussion. They suggest two possible accounts of interpersonal coordination: shared representations and coordinative structures. The idea informing their account of shared representations is that behavioural coordination, from body posture to gaze direction, is tied to the coordination of the representations of the participants, which may serve to facilitate the understanding of the actions of others, predict future actions, and lead to an understanding of what joint actions are possible (Shockely et al. 2009: 313). This approach characterizes cognitive representations as the central motivation and mechanism of coordination, and is therefore consistent with plural subject theory. In contrast, understanding interpersonal coordination as a coordinative structure, that is as a self-organized, softly-assembled (temporary) set of individual components that behave as a single functional unit (Bernstein 1967), does not necessarily involve an appeal to the cognitive representation of one's own or another's actions. What viewing coordinative action in this way does show, however, is that "transitions between

stable coordination modes across individuals exhibited the same hallmark dynamics found in individual interlimb coordination" (Shockley et al. 2009: 313). They also speculate that conversational coordination reflects a functional reorganization of body and eye movements to support joints actions, which leads them to suggest that "to understand the organization of this joint-action system, the assumption that each participant can be studied in isolation must be relaxed" (2009: 313).

While the coordinative structure account does not itself accord any special role to shared representations, work by Garrod and Pickering (2004) attempts to show that both of these accounts are consistent with each other. They argue that alignment at each level serves to bring conceptual representations together, and they offer an account of how higher level representations become aligned through priming. This is, however, only a beginning, as there is much empirical work to be done to show how alignment at lower levels interacts with alignment at the level of conceptual representation. Nonetheless, the consensus among researchers in this area is that the ability to engage in and sustain reciprocal relations is regulated by both cognitive and perceptual motor processes, which work together to enable coordination during joint action (Sebanz et al. 2006).

Finally, Ramenzoni et al. (2011) explored the role of task demands on the degree and stability of interpersonal coordination. They found that joint coordinative structures were influenced by the nature of the task performed and the constraints it placed on joint and single performance. In this study, participants performed a joint task in which one person held a stick with a circle attached at the top, while the other held a pointer through the circle without touching the borders. They then changed the size of the circle to vary the task difficulty. What they found is that when the joint task required more precision (smaller targets), the coordination between the actors increases, which they took to provide further evidence in support of the hypothesis that interpersonal and intrapersonal coordination are governed by the same basic principles. They conclude that (2011: 456):

> If interpersonal perceptual-motor coordination embodies joint attention and the coordination of cognitive activity involved in joint tasks, then perceptual-motor processes may provide a foothold for understanding the fundamental constraints that also characterize cognitive coordination during joint action .

This suggests that there may be a fundamental connection between the coordination of motor processes as understood by the analysis of coordinative structures, the coupling of lower-level cognitive processes outlined by Tollefsen and Dale, and higher-level cognitive processes such as joint commitment.

Working out the exact relation of these three areas is an empirical matter, but they all are consistent with the conceptual framework Gilbert supplies with the ideas of expressing readiness, acting as a single body, and forming a plural subject. Further, the concept of joint coordinative structures provides empirical evidence that two or more people completing a joint task may literally act in ways similar to that of a single body.

§6 – Potential Conflict

Despite the apparent coherence of the Gilbert's conceptual account and the empirical literature, there is one major problem with the conceptual coherence of using subpersonal dynamic patterns to explain the existence and functioning of full blown propositional attitudes.

In their discussion of the 'illusion of the we-will' (2012: 402-403), Tollefsen and Dale suggest that by controlling the lower-level processes in a laboratory setting, researchers may be able to create commitments in individuals, which they falsely attribute to prior intentions to form the commitment. From this consideration, they conclude that high-level commitments "are epiphenomena, misattributed to original intentions" (2012: 403). This style of argument in which dynamic structures supersede higher-level phenomena is also present in Shockley et al.'s juxtaposition of shared representations and coordinative structures. Gilbert's theory accounts well for this concern. If the joint commitment is somehow induced by laboratory settings, it does not mean that it is irrelevant for the performance of some joint task by the plural subject thusly committed. Regardless of its origin, the joint commitment functions in the same way and if there is a commitment to a joint intention, that intention functions in the same way. However, because joint commitment *precedes* joint intention under Gilbert's theory, the fact that the intention to form the plural subject was mistaken does not affect the subsequent joint intention. If anything, it

raises an interesting problem concerning the role of individual intentions in the formation of a plural subject. By construing 'expression of readiness' in nonintentional terms, we seem to have opened the door to the idea that joint commitments may be formed without that being the intention of the individuals. Gilbert states that "personal decisions and intentions create commitments of a kind I shall call *personal commitments*", and that "a joint commitment is not a concatenation of personal commitments. Thus it is not formed by virtue of the formation of a personal commitment by each of the parties" (2009: 180). Instead, Gilbert relies heavily on the idea of openly expressing personal readiness jointly with the others. This appears to suggest that the expression of readiness must at some point be recognized, but it does not rule out the idea that it was induced by laboratory settings. We may therefore mistakenly attribute the origin of our expression of readiness to a personal intention, but that does not preclude our ability to openly express our readiness at some time, and for this expression to lead to a genuine joint commitment.

§7 – Plural Subjects and Group Cognition

Up until this point I have developed the following theses:

1. Collections of people can be coupled by low-level and high-level cognitive processes or states.
2. Low-level coupling is a necessary feature of an implicit 'expression of readiness', which is a necessary step in the creation of a joint commitment. However, because there may also be explicit expressions of readiness that do not involve the presence of all the participants, it appears that low-level alignment is not itself necessary.
3. Low-level coupling is insufficient for full-blown joint action (based on the joint following example).
4. Joint commitment therefore remains a necessary and sufficient condition for a joint action.
5. Low-level coupling and high-level coupling (normative commitment) interact, are mutually sustaining, and are both involved in the execution of a joint action.

6. Plural subjects are integrated systems, which are coupled along multiple levels of cognition and may themselves instantiate certain cognitive processes.

I have not, however, addressed the direct relationship between plural subjects and distributed cognition as developed by Hutchins, nor group cognition as it is developed by Theiner. We may still ask whether *all* plural subjects are also collective cognitive systems. It may be the case that low-level alignment and joint commitment to a goal, while sufficient to create a coupled system, are not sufficient for a *cognitive* system in the relevant sense in every case. It may be argued that there is no reason to think that all groups that perform joint actions must thereby engage in some type of group level cognitive process. A thorough defence of any position on the issue of determining what constitutes a cognitive system is outside the scope of this thesis. However, I would like to present some considerations on the nature of joint action and normative commitment, which are relevant to group cognition.

There is a disconnect between empirical research of distributed cognitive systems such as transactive memory systems, stigmergic path-formation and so on, and conceptual analyses of joint intention and belief, which results from the different aims of the projects which inform them. A consequence of this is that we should not expect complete overlap. It would be premature to suggest that all collective cognitive systems require normative commitment. Theiner et al.'s example of stigmergic path-formation is clearly not an example of intentional group behaviour, because none of the participants of the study are aware of the problem they are solving. In fact, it is on the basis of the unintended consequences of the individual's behaviour that Theiner et al. argue that we should view this as an example of group cognition.

The reason for the separation between the collectives considered in the distributed and group cognition literature and the collectives considered in the joint action literature develops from the different conditions they place on collective behaviour. The concern for joint action is whether the behaviour of the collective can be characterized in intentional terms. When this is the case, joint action theorists conclude that the mental life of the collective can be understood using concepts such as belief, desire and intention. On the other hand, for cognitive scientists working on distributed cognition, these mental predicates require certain features which they argue are paradigmatic of cognition, such as the propagation of information across various subsystems and

components. Only then, according to Hutchins for example, may the collective be said to be a distributed cognitive system.

This separation presents a barrier to reconciling these theories, but not an insurmountable one, and not one that precludes any useful crossover of specific ideas and theses. We may, for example, hold that highlighting the difference between non-intentional and intentional behaviour is also important for empirical research. In the Richardson (2007) study for example, the presence of intentional coupling changed the information flows and visual patterns which affected the degree of lower-level coupling. Also, restricting group cognition to intentional behaviour allows proponents of group cognition to avoid one prominent objection. In his discussion of empirical arguments for group minds, Rupert points out that in the stigmergic path formation case from Theiner et al., the group problem-solving ability is better construed as an accidental by-product of individual level-intelligence (Rupert 2012: 632). If, however, we construe group cognition as an intentional phenomenon this objection loses force because the cognition displayed is not accidental and therefore not simply a "by-product", although we do at the same time lose cases of stigmergy as examples of group cognition. Nonetheless, the fact remains that in researching the nature of cognition it would be misguided and premature to hold that normative commitment is a necessary condition for a collective cognitive system because it does not seem conceptually impossible that a group could display cognitive properties unintentionally.

It would also be rash to claim that all alignment systems, in Tollefsen and Dale's sense, are collective cognitive systems. Tollefsen et al. also discuss a case that we saw from Theiner, that of Transactive Memory Systems. They argue that alignment plays a significant role in transactive memory because "alignment at the syntactic, semantic, and perceptual level will, as it does in conversation, produce cues that support and sustain the interactions involved in the transactive memory system" (2013: 56). However, they also point out that not all alignment systems involve robust group cognition: "we would not contend that all cases of conversation or collaboration involve group cognitive systems. We would instead argue that some such activities may qualify by having important dynamic alignment properties" (2013: 57). Low-level cognitive alignment is therefore one way in which individuals may integrate to form a collective cognitive system, but it clearly is not independently sufficient for group cognition. My contention is that the concept of

an ‘alignment system’ is the most readily applicable concept in the empirical literature for an understanding of joint action.

What alignment does is provide a plausible story for how the integration necessary for a cognitive system is developed. I contend that in cases in which the alignment leads to joint commitment to intend as a body to perform some action, joint commitment also plays an important role in developing the requisite integration. If we return to the conditions of Wilson for cognitive coupling between individuals, we see that plural subjects satisfy them for the most part. In forming a plural subject, individuals commit to exchanging the information reliably and as needed, each individual action in service to the joint action effects all other participants in the plural action and the plural subject itself may be viewed as a causal whole. Furthermore, in cases of necessarily joint action, by joining a plural subject the participants produce an effect that none of them individually could do. While this does not require that every case of joint action demands the formation of a collective cognitive system, as we may still imagine many cases of minimal joint actions, such as spontaneously pushing a stalled car, in which it would be highly unintuitive to claim that we form a genuine collective cognitive system, it does suggest that many joint action types which require planning and strategy over time do involve group cognitive processes. The key point is the task differentiation we saw in Ramenzoni (2011). Group cognition is task-variable; that is, it requires a certain kind of task. Teams, or collections that, while falling short of robust institutional existence, repeatedly use complementary problem-solving strategies, develop cognitive props (plays, codes, signals, ‘inside’ language) and divide their cognitive labour, all of which is intentional behaviour involving joint commitment in order to efficiently come to solutions, are perfect targets for the explanatory framework of distributed and group cognition.

Viewing collectivities as full-blown cognitive systems according to the distributed cognition framework requires that the group develop sophisticated mechanisms and architectures that allow for the transfer of collectively constructed representations across various media in order to react to features of the environment in an intelligent way. Although some collectivities that undertake joint actions, such as the navigational crew of a large ship, may meet these requirements, most do not develop such mechanisms and architectures. Nonetheless, all joint actions do seem to involve some minimal level of collective cognition, dynamic coupling, shared representations and basic alignment patterns. This result suggests that further research in this area will require a taxonomy

of the various levels of coupling and collective cognition according to the complexity of the task that the collectivity undertakes and that, on a more positive note, researching plural subjects as (some type of) collective cognitive systems may be plausible and explanatorily fruitful.

Conclusion

I began this thesis by considering the prevalence of our ascriptions of mental properties to collections of people. Since that introduction I have outlined two strands of research which both argue that such ascriptions are justified, but for significantly divergent, if related, reasons. Collective intentionality theorists do so on the grounds that the behaviour of certain collectivities is best characterized as intentional, while group/distributed cognition theorists do so on the grounds that we may find complex computational frameworks in collectivities that allow for the production of intelligent behaviour by information flows through integrated computational systems. In this conclusion, I will attempt to relate the most important points of overlap that may allow for a productive exchange between these similarly interested bodies of literature.

Within these two approaches, there is a range of causal interaction between individuals that leads to some type of collective cognitive activity. Based on the discussion of lower-level and higher-level cognitive coupling presented here, we can then draw a distinction between cognitive processes, which are understood on the basis of our pre-theoretical understanding of the distinctive roles they play in producing intelligent behaviour, that emerge in systems of individuals, from *intentional* collective cognitive systems, in which the cognitive processes are best applied to the collective. According to empirical research in distributed and group cognition, these capacities such as memory, attention, learning and problem-solving can be instantiated by groups. In these explanations, individuals are coupled in the appropriate way and the cognitive process emerges. Group cognition in the wide sense may then be seen as a special case of the claim that cognitive activities span a web of brain, body, and social and technological scaffolding, which is the general claim of extended cognition. It may also be seen as in line with the concepts of cognitive integration and complementarity emphasized by more recent arguments for extended cognition. These types of cognitive processes potentially attributed to groups may not be enough for us to conclude that these collections of individuals bring about some outcome in the world, together. The type of group cognition required for supporting plural action is a further subset of this special case. It is important to separate these two types of group cognition based on the types of predicates they ascribe to the subject.

I have here attempted to apply Gilbert's conditions to the idea of coupling employed in the studies of other collective cognitive systems in distributed/group cognition and social psychological research based on dynamic systems theory in order to explain one way in which a collective can be properly integrated and capable of being attributed the relevant mental states, in a folk-psychological causal-explanatory sense, required for undertaking a joint action. This conclusion, however, does not satisfy the requirements of current theories of distributed cognition for full-blown distributed cognitive systems.

The introduction to distributed cognition in this thesis came by way of the Hypothesis of Extended Cognition, and most importantly, by the specific approach used by the HEC to understand all cognition. The reasoning underlying the HEC in general is that mental phenomena must be understood as a property of a system that develops over time. A single neuron or collection of neurons in a subsystem of the brain does not alone have mental properties. Only when these collections reach a certain level of integration does a mental state emerge. Within the brain a cognitive task is often addressed by a temporarily assembled coalition of distributed neural components, held together by a fundamental 'functional connectivity' (Clark 2008, 137). One central claim in extended cognition is that the language of systems, components, interfaces and bandwidth is necessary for understanding cognition, and that this language does not privilege the biological in any significant way. The only relevant factors are the functional roles played by the components and the 'bandwidth' over the interfaces (points of contact between components), or in other words, the level of integration and fluid causal exchange (Haugeland 1998). This points to the fact that the HEC is not so much about the extension of an individual mind, which is the impression that comes from C&C's focus on tool use, but rather about the dynamic interactions of systems and the patterns that emerge, which becomes clear when the focus is turned slightly to its related use in distributed cognition.

The extended cognition framework, with the language of dynamic systems and emergent non-reductionist properties, is particularly appropriate for the analysis of joint action when it is itself extended to cover group cognition. It provides a plausible story for the attribution of literal cognitive states and processes to groups of people, based on a compatible argumentative strategy centered on the idea of non-aggregativity. Brains are assumed to have cognitive properties that no individual neuron has, or is capable of having, and this consideration, in the group context,

suggests that collectives have the potential to display emergent cognitive properties that no individual member has, or might even be capable of having. The individual neurons, the body and cognitive artefacts create an integrated system in which cognitive states emerge, just as collections of individuals, who achieve the requisite level of integration are capable of possessing cognitive properties and bringing about joint outcomes together.

On the other hand, arguing for collective cognition on the basis of informational flows between subsystems in a larger cognitive architecture, which is the strategy of distributed cognition, precludes the possibility of ascribing such cognitive properties to all collectives that bring about some outcome intentionally. If the complexity of the task is too low, the cohesion of the group too temporary, and the communication strategies of the collective too simplistic, no collective cognitive processes may be required or present. This is where the argumentative strategies of the two approaches (distributed cognition and collective intentionality) diverge, and where we find the central problem for further work in this vein. Much distributed cognition literature seeks to empirically demonstrate the existence of mental states and cognitive processes on the basis that the group produces intelligent behaviour which is governed by an integrated computational system. The sophistication required of the collective to process information in the appropriate way demands an extensive cognitive architecture. Many collectives that perform joint actions lack this architecture. They may not therefore be said to have collective cognitive properties according to distributed cognition. Collective intentionality theorists, in contrast, argue that these collectivities may be ascribed intentional states because their behaviour is intentionally characterizable. This disconnect must be overcome in order to apply the findings of this empirical research to the philosophical problem of joint action.

While this disconnect has not been addressed here at length, we may nonetheless maintain that if we want to understand certain aspects of joint action, we may look to empirical studies in group cognition of collectives performing joint actions, as these studies provide important information about the practical requirements of collective intentional behaviour and the strategies employed by collectivities to meet these requirements. These studies suggest that when there is collective cognition, it is an emergent, dynamic activity that is not attributable to any single participant in the action, but instead to the shared (in a non-aggregative fashion) cognition of the collective, which in turn suggests that the collective as a whole is the proper unit of analysis.

Here we may turn to Gilbert's theory of plural subjects. It provides a conceptual framework which identifies 'landmarks' along the way, and informs the 'architecture' of the cognitive creations of the group, such as the joint intentions and beliefs, which in turn, inform the cognitive processes (memory, problem solving) that the group performs in the completion of the joint action. In order for these collectives to have the requisite togetherness to form an intentional collective cognitive system, they must explicitly form this under the conditions of joint commitment. This applies to the highly complex and integrated collectives studied in the distributed cognition literature. Commitment to acting as a body may then be seen as a commitment to achieving the type of organizational integration required for the attribution of folk-psychological mental states, but not necessarily for mental states involving the creation and propagation of collective representations in the distributed cognition sense. The explicit and intentional nature of this commitment separates a significant section of cases from those where causally coupled collections of people bring about some effect simply by chance. In turn, the notions of group belief and action in Gilbert's account receive the support of empirical research, especially with respect to her concepts of 'expression of readiness' and 'acting as a single body'. The collections of agents that perform joint actions have clear lines of interaction and processes of alignment that are mediated by normative commitments and form the foundation of coordinative structures that lead to group cognitive systems.

Conceiving of joint action and cognition in this way provides us with the conceptual tools to accept Baier's suggestion to take the first person plural seriously. Baier points out that "the threat of the general will lies in its assumption of a super-singularity, not in its non-suspicion of the first person plural" (1997: 19). She suggests that the entailments of the first person plural require either a collective subject, or a type of shared collective ownership of action for interrelated individuals, which does not require the complete repudiation of individual autonomy. When performing a joint action, in which individuals must stand in a certain relation to one another, we may interpret these individuals as sharing an intentional state, or having an intentional state collectively, and still interpret them as both the bearers of the intentional state required for their individual contribution to the joint action and the intentional agents responsible for putting themselves into that particular relational structure. As Baier states, "We do not need to analyse our 'we shall…'s either into a set of 'shall's of ordinary 'I's and 'you's, as Bratman attempts to

do, or into the ‘shall’ of some super individual, some ‘moi commun,’ as Rousseau does. We can simply ask why the first person singular should be deemed more fundamental than the first person plural” (Baier 1997: 41). Under accounts based on an extended or distributed view of cognition, neither the 'we shall’s' of our intentional state, nor of our cognitive processes need be analyzed as sets of ‘I’s and ‘you’s, nor do we need to postulate any entity in a strong ontological sense over and above the individuals.

I take Baier’s point to be a rejection of aggregativity. She stresses that group beliefs and intentions are not aggregates of individual beliefs and intentions, and group cognitive processes, such as problem-solving, are not the aggregate of individual cognitive processes. The central claim developed here is that the use of dynamic systems analyses in extended cognition, social psychology and cognitive science, and the pervasive use of concepts from dynamic systems theory that entails, especially the concepts of coupling, self-organization and soft-assembly, provides a strong empirical background for the conceptual rejection of aggregativity, making it the ideal empirical background to support Baier’s claim and Gilbert’s theory. Further, this approach highlights the connections between Hutchins’ and Hollan’s account of distributed cognition, Theiner et al.’s account of group cognition, and Tollefsen’s account of alignment systems. Taken together this empirical research may be used to develop more sophisticated philosophical explanations of joint action and provides reasons to prefer relational, non-reductionist philosophical accounts, such as Margaret Gilbert’s.

The first steps suggested here represent a limited beginning in the rethinking of the barriers between individual and collective brought about by the empirical results of the extended, distributed, group and situated cognition framework, results which may be brought to bear not only on issues of joint intention and action, but also social ontology and group agency.

Bibliography

Ackerman, M.S. & Halverson, C. (1998) "Considering an organization's memory", *Proceedings of Computer Supported Cooperative Work,* New York: Kluwer Academic Publishers, 39-48.

Adams, F. & Aizawa, K. (2001) "The Bounds of Cognition", *Philosophical Psychology*, 14, 43-64.

Adams, F. & Aiziwa, K. (2007) *Bounds of Cognition*, Oxford: Blackwell Press.

Adams, F. & Aiziwa, K. (2010) "Defending the Bounds of Cognition", *The Extended Mind*, ed. Menary, R. Cambridge, Mass: MIT Press. p. 67-80.

Anscombe, E. (2000) *Intention* (reprint), Cambridge, MA: Harvard University Press.

Baier, A. (1970) "Act and Intent", *Journal of Philosophy* 67: 648-658.

Baier, A. (1997) "Doing things with Others: The Mental Commons" in Alanen, L., S. Heinämaa, and T. Wallgren, (eds.), *Commonality and Particularity in Ethics,* New York: St. Martin's Press, Inc.

Bernstein, N.A. (1967) *Coordination and regulation of movements,* New York: Pergamon Press.

Blomberg, O. (2011) "Socially Extended Intentions-in-Action", *Review of Philosophy and Psychology*, Vol. 2, 335-353.

Bratman, M. (1992) "Shared Cooperative Activity", *The Philosophical Review* 101: 327–341.

Bratman, M. (1999) *Faces of Intention*, Cambridge: Cambridge University Press.

Bratman, M. (2009) "Modest sociality and the distinctiveness of intention", *Philosophical Studies* 144:149–165.

Burge, T. (1979) "Individualism and the mental", in French, Uehling, & Wettstein (eds.). *Midwest Studies in Philosophy,* Minneapolis: University of Minnesota Press, Vol. 4, 73-121.

Clark, A. (1997) *Being There*, Cambridge, Mass: MIT Press.

Clark, A. (1998) "Authors Response: Review Symposium on *Being There*", *Metascience*. 7: 95-103.

Clark, A. (2001) *Mindware*: *An Introduction to Cognitive Science*, Oxford: Oxford University Press.

Clark, A. (2004) “Momento’s revenge: the extended mind revisited”, reprinted in R. Menary (ed.) *The Extended Mind* Cambridge, Mass.: MIT Press.

Clark, A. (2008) *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, New York: Oxford University Press.

Clark, A. (2010) “Coupling, Constitution, and the Cognitive Kind: A Reply to Adams and Aiziwa”, in R. Menary (ed.) *The Extended Mind*, Cambridge, Mass.: MIT Press.

Clark, A. & Chalmers, D. (1998) “The extended mind”, *Analysis*, 58 (1), 7-19.

Davidson, D. (1980) *Essays on Actions and Events*, Oxford: Oxford University Press.

Flor, N.V. and Hutchins, E. (1992) “Analyzing Distributed Cognition in Software Teams: a Case Study of Collaborative Programming During Adaptive Software Maintenance”, in *Empirical Studies of Programmers: Fourth Workshop*, eds. J. Koenemann-Belliveau, T. Moher. and S. Robertson, Norwood, NJ: Ablex.

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

Gallagher, S. (2008) “Intersubjectivity in perception”, *Continental Philosophy Review*. 41: 163-178.

Gallagher, S. & Crisafi, A. (2009) “Mental Institutions”, *Topoi*, 28, 45-51.

Garbis, C. & Waern, Y. (*1999*) "Team coordination and communication in a rescue command staff: The role of public representations", *La Travail Humain*: 273-291

Garrod, S., & Pickering, M.J. (2004) “Why is conversation so easy?”, *Trends in Cognitive Sciences*, 8(1): 8-11.

Giere, R. (2006) “The Role of Agency in Distributed Cognitive Systems”, *Philosophy of Science* 73: 710-719.

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

Gilbert, M. (1990) “Walking Together: A Paradigmatic Social Phenomenon”, *Midwest Studies in Philosophy* 15:1–14.

Gilbert, M. (1996) *Living Together*, Maryland: Rowman & Littlefield.

Gilbert, M. (1997) “What Is It for Us to Intend?”, in *Contemporary Action Theory, vol. 2: The Philosophy and Logic of Social Action*, ed. G. Holmstrom-Hintikka and R. Tuomela, Dordrecht: Kluwer Academic Publishers, 65–85.

Gilbert, M. (2006) *A Theory of Political Obligation*, Oxford: Oxford University Press.

Gilbert, M. (2009) “Shared Intention and Personal Intention”, *Philosophical Studies* 144:167–187.

Goldstone, R.L. & Gureckis T.M. (2009) “Collective Behavoir”, *Topics in Cognitive Science* 1: 412-438.

Goldstone, R.L. & Roberts, M. E. (2006) “Self-organized Trail Systems in Groups of Humans”, *Complexity* 11: 43-50.

Hollan, J. & Hutchins, E. (2010) “Opportunities and Challenges for Augmented Environments: A distributed Cognition Perspective”, *User Friendly Environments: From Meeting Rooms to Digital Collaborative Spaces*, ed. Lahlou Saadi, 237-259.

Hollan, J., Hutchins, E., Kirsh, D. (2000) “Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research”, *Transactions of Computer-Human Interaction*. 7(2): 174-196.

Huebner, B. (2008) “Do You See What We See? An Investigation of an Argument against Collective Representations”, *Philosophical Psychology* 21: 91-112.

Huebner, B., Bruno, M., & Sarkissian, H. (2010) “What Does the Nation of China Think about Phenomenal States?”, *Review of Philosophy and Psychology* 1: 225-243.

Hurley, S. (2001) “Perception and Action: Alternate Views”, *Synthese*.129: 3-40.

Hurley, S. (2006) “The varieties of externalism”, in R.Menary (ed.) *The Extended Mind*. Cambridge, Mass.: MIT Press.

Hutchins, E. (1995) *Cognition in the Wild*, Cambridge, MA: MIT Press.

Hutchins, E. (2001) “Distributed Cognition”, *International Encyclopaedia of the Social & Behavioral Sciences*, eds. Smelser, N. & Baltes, P., Amsterdam: Elsevier, 2068-2072.

Hutchins, E. (2008) “The role of cultural practice in the emergence of modern human intelligence”, *Philosophical Transactions of the Royal Society B*, 363: 2011-2019.

Hutchins, E. (2011a) “Enaction, Imagination, and Insight”, *Enaction: Toward a new paradigm for cognitive science*, eds. Stewart, J., Gapenne, O., and Di Paolo, E., Cambridge, Mass: MIT Press, 425-450.

Hutchins, E. (2011b) “Enculturating the Supersized Mind”, *Philosophical Studies*, 152: 437-446.

Hutchins, E. & Klausen, T. (1996) “Distributed cognition in an airline cockpit”, in Engeström, Y., and Middleton, D., eds. *Cognition and communication at work,* New York: Cambridge University Press, 15-34.

Knobe, J. & Prinz, J. (2008) "Intuitions about Consciousness: Experimental Studies", *Phenomenology and the Cognitive Sciences* 7: 67-83.

List, C., & Pettit, P. (2011) *Group Agency: The Possibility, Design, and Status of Corporate Agents*, Oxford, UK: Oxford University Press.

Menary, R. (2006) "Attacking the Bounds of Cognition", *Philosophical Psychology*. 19(3): p. 329-344.

Menary, R. (2010a) "Cognitive Integration and the Extended Mind", *The Extended Mind*, ed. Menary, R. Cambridge, Mass: MIT Press, 227-243.

Menary, R. (2010b) "Dimensions of Mind", *Phenomenology and the Cognitive Sciences*. 9(4): 561-578.

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

Perry, M. (2003) "Distributed Cognition", in *HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science*, ed. Carroll, J.M., San Francisco: Morgan Kaufmann Publishers, 193-225.

Pettit, P. (1993) *The Common Mind: An Essay on Psychology, Society, and Politics*, Oxford: Oxford University Press.

Putnam, H. (1975) "The meaning of 'meaning'", in K. Gunderson (ed.) *Minnesota studies in the philosophy of science*. Minneapolis: University of Minnesota Press, Vol. 7, 131-193.

Quinton, A. (1975) "Social Objects", *Proceedings of the Aristotelian Society* 76:1-27.

Rakoczy, H. (2006). "Pretend play and the development of collective intentionality", *Cognitive Systems Research*, 7, 113-127.

Rakoczy, H. (2007) "Play, games, and the development of collective intentionality", in C. Kalish &M. Sabbagh (Eds.), *Conventionality in cognitive development: How children acquire representations in language, thought and action. New Directions in Child and Adolescent Development*. No. 115 (pp. 53-67), San Francisco: Jossey-Bass.

Ramenzoni, V.C., Davis, T.J., Riley, M.A., Shockley, K. & Baker, A.A. (2011) "Joint action in a cooperative precision task: nested processes of intrapersonal and interpersonal coordination", *Experimental Brain Research*, 211: 447-457.

Ramenzoni, V.C., Riley, M.A., Shockley, K., & Baker, A.A. (2012) "Interpersonal and intrapersonal coordinative modes for joint and single task performance", *Human Movement Science* 5: 1253-1267.

Richardson, D.C., Marsh, K.L, Isenhower, R.W., Goodman, J.R., & Schmidt, R. (2007) "Rocking Together: Dynamics of intentional and unintentional interpersonal coordination", *Human Movement Science*, 26(6): 867-891.

Roberts, M.E. & Goldstone, R.L. (2009) *Adaptive group coordination proceedings of the thirty-first annual conference of the cognitive science society*, Amsterdam: Cognitive Science Society, 2698-2704.

Rowlands, M. (2009) "Enactivism and the Extended Mind", *Topoi*, 28: 53-62.

Rowlands, M. (2010) *The New Science of the Mind*, Cambridge, Mass.: MIT Press.

Rupert, R. (2009) *Cognitive Systems and the Extended Mind*, Oxford: Oxford University Press.

Schmid, H. B. (2003) "Can Brains in Vats Think as a Team?", *Philosophical Explorations* 6: 3, 201-217.

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

Schmidt, R.C., O'Brien, B., & Sysko, R. (1999) "Self-organization in between-person cooperative tasks and possible applications for sport", *International Journal of Sport Psychology*, 30: 558-579.

Schumpeter, J.A. (1908) *Das Wesen und der Hauptinhalt der theoretischen Nationalökonomie*. Leipzig: Dunker und Humblot.

Schweikard, D.P. (2008) "Limiting Reductionism in the Theory of Collective Action" in *Concepts of Sharedness: Essays on Collective Intentionality*, ed. H.B. Schmid, K. Schulte-Ostermann, Nikos Psarros, Piscataway, N.J: Transaction Books, 89-119.

Schweikard, D. P. and Schmid, H. B., "Collective Intentionality", *The Stanford Encyclopaedia of Philosophy* (Summer 2013 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2013/entries/collective-intentionality/>.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006) "Joint action: bodies and minds moving together", *Trends in Cognitive Sciences*, 10: 70-76.

Searle, J. (1983) *Intentionality*, New York: Cambridge University Press.

Searle, J. (1990) "Collective Intentions and Actions", in *Intentions in Communication*, edited by P. Cohen, J. Morgan, and M. Pollack, Cambridge, Mass.: MIT Press, 401–415.

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

Searle, J. (1997) "Replies to Critics", *Philosophy and Phenomenological Research* 57: 449-451.

Searle, J. (1998) “Social Ontology and the Philosophy of Society”, *Analyse und Kritik* 20: 143-158.

Shockey, K., Baker, A. A., Richardson, M.J., & Fowler, C.A. (2007) “Articulatory constraints on interpersonal postural coordination”, *Journal of Experimental Psychology: Human Perception and Performance*, 33: 201-208.

Shockey, K., Richardson, D.C., & Dale, R. (2009) “Conversation and coordinative structures”, *Topics in Cognitive Science*, 1: 305-319.

Shockley, K., Santana, M.V., & Fowler, C.A. (2003) “Mutual interpersonal postural constraints are involved in cooperative conversation”, *Journal of Experimental Psychology: Human Perception and Performance*, 29: 326-332.

Stoutland, F. (1997) “Why Are Philosophers of Action so Anti-Social?”, in Alanen, Heinamaa, and Wallgreen eds. *Commonality and Particularity in Ethics*, New York, N.Y.: St. Martin’s Press.

Sutton, J. (2006) “Distributed Cognition: Domains and dimensions”, *Pragmatics & Cognition*. 14(2): 235-247.

Sutton, J. (2010) “Exograms and Interdisciplinarity: History, the Extended Mind, and the Civilizing Process”, in *The Extended Mind*, ed. Menary, R. Cambridge, Mass: MIT Press. p. 189-225.

Sutton, J., Harris, C.B., Keil, P.G., Barnier, A.J. (2010) “The psychology of memory, extended cognition, and socially distributed remembering”, *Phenomenology and the Cognitive Sciences*. 9: 521-560.

Sytsma, J. & Machery, E. (2009) “How to Study Folk Intuitions about Phenomenal Consciousness” *Philosophical Psychology* 22: 21-35.

Sytsma, J. & Machery, E. (2010) “Two Conceptions of Subjective Experience”, *Philosophical Studies* 151: 299-327.

Taylor, C. (1985) *Philosophical Papers Vol. 2*, Cambridge: Cambridge University Press.

Thelan, E. & Smith, L. (1994) *A Dynamic Approach to the Development of Cognition and Action*, Cambridge, Mass: MIT Press.

Theiner, G., Allen, A., & Goldstone, R. (2010) “Recognizing group cognition”, *Cognitive Systems Research*, 11(4): 378-395.

.

Tollefsen, D. (2004) “Collective Intentionality”, *Internet Encyclopaedia of Philosophy*, URL = < http://www.iep.utm.edu/coll-int/>.

Tollefsen, D. (2005) “Let’s pretend! Children and joint action”, *Philosophy of the Social Sciences*, 35(1), 75–97.

Tollefsen, D. (2006) “From extended mind to collective mind”, *Cognitive Systems Research* 7 (2-3), 140-150.

Tollefsen, D. & Dale, R. (2012) “Naturalizing joint action: A process-based approach”, *Philosophical Psychology*, 25(3): 385-407.

Tollefsen, D. P., Dale, R., & Paxton, A. (2013) “Alignment, Transactive Memory, and Collective Cognitive Systems”, *Review of Philosophy and Psychology*, 4: 49-64.

Tomasello, M. & Carpenter, M. (2007) “Shared intentionality”, *Developmental Science*, 10, 121-125.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005) “Understanding and sharing intentions: The origins of cultural cognition”, *Behavioral and Brain Sciences*, 28, 675-691.

Tomasello, M., & Rakoczy, H. (2003) “What makes human cognition unique? From individual to shared to collective intentionality”, *Mind and Language*, 18(2), 121-147.

Tribble, E. (2005) “Distributing cognition in the globe”, *Shakespeare Quarterly*, 56: 135-155.

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

Varela, F.J., Thompson, E. & Rosch, E. (1991) *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, Mass: MIT Press.

Velleman, J.D. (1989) *Practical Reflection*, Princeton: Princeton University Press.

Velleman, J.D. (1997) “How to Share and Intention”, *Philosophy and Phenomenological Research* 57: 29-51.

Weber, M. (1922/1978) *Economy and Society: An Outline of Interpretive Sociology*, Berkley, CA: University of California Press

Wegner, D.M. (1986) “Transactive memory: A contemporary analysis of the group mind”, in eds. Mullen, B. & Goethals, G.R., *Theories of group behaviour*, New York: Springer Verlag, 185-208.

Wegner, D.M. (1995) “A computer network model of human transactive memory”, *Social Cognition*, 13: 319-339.

Wegner, D.M. (2003) “The mind’s best trick: How we experience conscious will”, *Trends in Cognitive Sciences*, 7: 65-69.

Wegner, D.M., Giuliano, T., & Hertel, P. (1985) "Cognitive interdependence in close relationships", in ed. Iches, W.J., *Compatible and incompatible relationships*, New York: Springer Verlag, 253-276.

Wilson, R. A. (2004) *Boundaries of the Mind*, Cambridge: Cambridge University Press.

Wilson, R. A. (2010) "Extended Vision", in *Perception, Action and Consciousness*, eds. Gangopadhyay, Madary, M., Spicer, F., Oxford: Oxford University Press, 277-290.

Wilson, R.A. & Clark, A. (2009) "How to Situate Cognition: Letting Nature Take its Course". *The Cambridge Handbook of Situated Cognition I,* ed. Aydede M. & Robbins O. New York: Cambridge University Press, 55-77.

Wimsatt, W.C. (1986) "Forms of Aggregativity", in eds. Grene, M.A., Donagan, A., Perovich, A.N., Wedin, M.V., *Human nature and natural knowledge*, Dordrecht: Reidel, 259-291.

Appendix 1: Abstract

The proposed aim of this thesis is to analyze the potential connections between the hypothesis of extended cognition (HEC) and collective intentionality and to begin the development of an account of joint action building on the idea of a coupled system. The HEC is the philosophical interpretation of a new movement in cognitive science, with related strands alternatively named, "situated cognition", "4ea cognition" "distributed cognition" or "social cognition", which challenges the previously dominant models of mind by suggesting that the mind is not bounded to the properties of the skin or skull. The philosophical idea is that if cognitive processes are *constituted* by processes that involve features of the body and environment, then the supervenience base of mental states extends beyond the boundaries of the skull.

The view that the mind encompasses aspects of the social environment, including other minds, allows for both an individualistic and a collectivistic application to the literature on joint action and collective intentionality. On the collectivistic view, taking minds outside of heads supports the possibility of the formation of collective cognitive systems, which are, in a sense, group minds, with collective mental states. This removes some of the major reasons for the adherence to certain forms of individualism in the collective intentionality literature. Some collective intentionality theorists hold that joint intentions must be analyzed in terms of the intentions of individuals, finding the 'jointness' of the intention in either the content or the mode of the relevant propositional attitudes rather than the subject. One may argue that locating the mind outside of individuals removes many of the motivations for these views, by creating the possibility of legitimate collectives with their own mental states. An analysis of joint action may then follow from the formation of a collective subject.

Because both the HEC and collective intentionality are relatively new topics by philosophical standards, neither of these positions has been comprehensively formulated. The central question of this thesis is whether we can develop a coherent method of conceptualizing collective systems of individuals that are capable of bearing propositional attitudes based on the idea of causal coupling. The proposed account defends the collectivist application of the HEC to collective intentionality based on the analysis of joint action put forward by Margaret Gilbert. It is argued that Gilbert's theory is in a unique position to fill this roll because it involves the concept of a

plural subject, to which intentional attitudes can be ascribed. Gilbert's account of the formation of a plural subject reveals the correct set of conditions for the formation of a coupled agential system which guides its group behaviour intelligently.

In order to make this application plausible a concept of coupling will need to be systematically developed for this special case of socially distributed cognition. Much of the confusion surrounding this topic derives from the somewhat nebulous explanation of coupling that has been given in the literature. Under the standard account, coupling is equated with continuous reciprocal causation over time, but there are many cases of these types of couplings between individuals that are not instances of joint action. This issue is clarified in order to show how Gilbert's conditions can be added to the base account of causal coupling. It is argued that each set of extensions (body, environment, cognitive tools and other minds) requires its own specific set of extension conditions and that Gilbert's concept of joint commitment to acting as a body provides the extra condition which allows us to analyze intentional joint action using the language of coupled systems and the conceptual framework of the extended mind.

~

Ziel dieser Arbeit ist die mögliche Zussamenhänge zwischen ‚Erweiterte Kognition' und ‚Kollektive Intentionalität' zu analysieren. Erweiterte Kognition ist die philosophische Interpretation einer neuen Bewegung in der kognitiven Wissenschaft, welche die dominanten Modelle des Geistes in Frage stellt, in dem sie behauptet, dass der Geist nicht auf den menschlichen Körper begrenzt ist. Die zentrale Idee dahinter ist, wenn kognitive Prozesse von Prozessen, die Eigenschaften von der Umwelt beinhalten, konstituiert sind, dann inkludiert die Superveniencebasis von mentale Zustande weitere Elemente ausserhalb des Gehirns (und des Körpers).

Die These, dass der Geist Aspekte der sozialen Umwelt und vielleicht andere Geiste umgreift, offnet viele Möglichkeiten für eine Analyse gemeinsamer Handlungen. Kollektiv gesehen unterstutzen erweiterte kognitive Prozesse eine Analyse gemeinsamer Handlung, die kollektive, mentale Zustände benutzt, um zu erklären wie wir Dingen absichtlich und miteinander erreichen. Viele Theorien gemeinsamer Handlungen besagen, dass wir dieses Phänomen nur durch

individuelle Absichten erklären können. Wenn wir die Thesen der erweiterten Kognition als wahr annehmen, sind viele Motivationen für diese individualistischen Theorien weg. Anstelle von solche Theorien können wir dann eine Theorie bauen, die gemeinsame Handlungen durch genuine gemeinsame Absichten erklärt.

Weil beide Forschungbereiche (Erweiterte Kognition und Kollektive Intentionalität) relativ neu sind, wurden sie nicht umfassend formuliert und daher bleiben viele Fragen über den möglichen Gewinn dieser Thesen offen.In meiner Masterarbeit habe ich mich umfassend mit einer dieser offenen Fragen beschäftigt: können wir Ideen der erweiterten Kognition benutzen, um eine kohärente Methode zu entwickeln, die notwendige Rationalität einer gemeinsame Handlung durchzufuhren als eine vereinigten Gruppenprozess, statt koordinierte individuelle Prozesse, zu verstehen.

Appendix 2: Curriculum Vitae

Matthew Rachar

Email: matthewrachar@gmail.com
Phone: +43 680 110 7822

Education

2011 – Present	M.A., Philosophy University of Vienna
2004 – 2009	B.A., Philosophy University of British Columbia Vancouver, Canada

Scholarships and Awards

2008	CIS Academic All-Canadian Trek Excellence Scholarship for Continuing Students David McLean Men's Basketball Award
2007	CIS Academic All-Canadian David McLean Men's Basketball Award Jack D Edmondson Memorial Basketball Award
2006	CIS Academic All-Canadian David McLean Men's Basketball Award Bob Hindmarch Award
2005	CIS Academic All-Canadian Bob Hindmarch Award (Basketball)
2004	David McLean Men's Basketball Award

Employment

Editor/Proofreader	Dr. Herwig Grimm

	Messerli Forschungsinstitut September 2013 – Present
Basketball Player	BK Klosternueburg Dukes Phone: +43 664 886 242 44 www.bk-klosterneuburg.at September 2009 – Present
English Teacher	Niederösterreiche Landesschulrat Volkschule Gerasdorf (Elementary school) Hauptschule Gerasdorf (Middle School) October 2009 – June 2010

Languages

English (Native), German (Fluent: writing, reading, speaking)