THE UNIVERSITY *of* EDINBURGH

Joint action without and beyond planning

Olle Blomberg

Submitted for the degree of PhD by research

The University of Edinburgh

2012

I have read and understood the University of Edinburgh guidelines on plagiarism. This thesis was composed by me and is entirely my own work except where I indicate otherwise by use of quotes and references. No part of it has been submitted for any other degree or professional qualification.

Olle Blomberg

Acknowledgements

First of all, I want to thank my supervisors Till Vierkant, Jane Suilin Lavelle and Andy Clark. Till took over as primary supervisor at a very late stage, in August 2012 after Matt Nudds left the department, and provided very useful critical feedback at the mock viva that he organised in October (as did Suilin and Matthew Chrisman who were also present). Suilin has been my secondary supervisor since August 2011, but this official rank does not reflect her actual contribution. I have benefited greatly from her critical readings of my work and her constructive feedback since she came onboard. Andy Clark started out as my primary supervisor, but when the focus of my PhD changed after about one and a half years, he became my tertiary supervisor. Andy has been very supportive throughout my studies and gave me constructive feedback on material I wrote on Dynamical Systems Theory and bodily coordination that eventually became parts of chapter 5.

I also want to thank my two previous supervisors Matt Nudds and Natalie Gold, who each was my primary supervisor for about a year (unfortunately both left Edinburgh's philosophy department during my studies to take up positions elsewhere). Matt provided invaluable help with thinking through and fixing the overall structure of the thesis before he left, and he also helped me structure my thoughts on common knowledge in chapter 2 (plus saved me from making some fundamental mistakes in chapter 3). Before she left, Natalie provided valuable feedback on drafts of the paper that eventually became chapter 4 and guided my early explorations of the topic of common knowledge.

Others who have provided feedback (on writings or talks I have given) that in one way or another has influenced the content or shape of the thesis include (in no particular order): Orestis Palermos, Jonas Christensen, Eric Kerr, Andy McKinlay, Chiara Gambi, Anika Fiebich, Edward Baggs, Julian Kiverstein, Lucy Tomlinson, Michael Bratman, Liz Irvine, Ashley Taylor, Axel Seemann, Natalie Sebanz, Günther Knoblich, Georg Theiner, Auke Pols, and Deborah Tollefsen. Steve

Butterfill deserves to be mentioned and thanked separately. It was after a brief conversation with him at the first Mind Network meeting in Oxford in March 2010 that I finally decided to make joint action the focus of my PhD (I had then just written a short paper about young children and joint action). Steve later urged me to submit the paper that later became chapter 4 to a special issue of *Review of Philosophy and Psychology* that he edited together with Natalie Sebanz. He also invited me to a stimulating informal meeting about joint action with him and Axel Seemann in London, June 2011. Such encouragement and support have meant a lot.

I also want to deeply thank Richard Harper at Microsoft Research (MSR), Cambridge, as well as MSR itself. MSR awarded me a European PhD scholarship (2008-2011) based on a research proposal entitled "Embodiment, Cognition and Interaction Design" that I submitted together with Andy Clark. Despite a change of topic and direction, Richard, MSR and Andy have continued to support me. Richard also kindly took me on as a (generously paid) research intern at MSR's Cambridge lab from January to April 2012, where he pushed me to write up a first preliminary draft of the thesis. Richard's support is especially remarkable since he fundamentally disagrees with my approach to philosophy. Without the financial support and commitment from Richard and MSR, this thesis would simply not have been written. I hope Richard finds the thesis to at least be a source of stimulating disagreements.

Finally, I want to thank to my friends in Edinburgh and my fellow postgraduate philosophy students. I especially want to thank Mog Stapleton for her counsel in times of crisis. For games of squash, thanks to Shane Ryan, Dave Ward, Tom Ferrier, Cameron Boult and Lee Whittington. Special thanks to Dave for some truly epic ones.

Abstract

Leading philosophical accounts of joint activity, such as Michael Bratman's account of 'shared intentional activity', take joint activity to be the outcome of two or more agents having a 'shared intention', where this is a certain pattern of mutually known prior intentions (plans) that are directed toward a common goal. With Bratman's account as a foil, I address two lacunas that are relatively unexplored in the philosophical literature. The first lacuna concerns how to make sense of the apparently joint cooperative activities of agents that lack the capacities for planning and "mindreading" that one must have in order to be a party to a shared intention (consider, for example, the social play of young children or the cooperative hunting of non-human primates or social carnivores). The second lacuna concerns how participants (including adult human agents) are able to coordinate their actions "online"—that is, during action execution as a joint activity unfolds—without recourse to plans that specify in advance what they should do (consider the coordination involved when two friends meet and do a "high five"). Chapters 2 and 3 focus on the first lacuna, while chapters 4 and 5 focus on the second.

In chapter 2, I focus on why participants must have mutual or common knowledge of each other's intentions and beliefs in order to have a shared intention: Why must these attitudes be "out in the open"? I argue that, if participants lack the concept of belief, then one of the two main motivations for the common knowledge requirement—to filter out certain cases that intuitively aren't cases of genuine joint activity—actually dissipates. Furthermore, a kind of "openness" that only requires of participants that they have the concept of goal but not that of belief can satisfy the other main motivation, to make sense of the idea that joint activities are non-accidentally coordinated. In chapter 3, I offer an account of a kind of joint activity in which agents such as young children and some non-human primates could participate, given what we know about their socio-cognitive capacities.

In chapter 4, I argue that 'shared intention'-accounts are unable to say much about spontaneous or skilful joint action because of the following widely accepted constraint on what one can intend: while an agent might intend—in the sense of

commit to a plan—*that* "we" do something together, an agent cannot intend *to* perform "our" joint action. I reject this constraint and argue that some joint actions (such as a joint manoeuvre performed by two figure skaters) are joint in virtue of each participant having what I call 'socially extended intention-in-action' that overlap. In chapter 5, I review empirical work on subpersonal enabling mechanisms for the coordination of joint action. The review provides clues to what it is that enables participants to successfully coordinate their actions in the absence of plan-like intentions or beyond what such intentions specify.

While what I address are *lacunas* rather than *problems*, an upshot of this thesis is that leading philosophical accounts of joint activity may have less explanatory scope than one might otherwise be led to believe. The accounts of joint activity and joint action that are presented in this thesis are arguably applicable to many of the joint activities and joint actions of adult human beings. The account also helps us avoid the false dichotomy between a very robust form of joint activity and a mere concatenation of purely individualistic actions—a dichotomy that accounts such as Bratman's arguably invite us to adopt.

TABLE OF CONTENTS

1 Introduction

We do countless things together: we shake hands, we walk, talk and play together, we eat and drink together, we lift heavy objects together, as well as collectively engage in more complex activities; we write academic papers, perform ballets and concerts, plan and run workshops and conferences, build airplanes, conduct scientific experiments, play multi-player online games and participate in religious ceremonies. The lists could, of course, go on and on and on. Such activities appear to be more than merely aggregates of individual actions. We may both have gone to see the same screening of the new Woody Allen film even if we didn't go to see the new Woody Allen film *together*. Even if it is true that "you and I lifted the heavy sofa", the conjuncts "you lifted the heavy sofa" and "I lifted the heavy sofa" are both misleading since each of us only made a partial contribution to the lifting of the sofa. This is not merely semantic nitpicking. Whether some collection of individual actions makes up a concerted joint action may make a big difference to how the actions are experienced in everyday life. Accidentally going to the cinema next to someone is typically much less enjoyable than going to cinema together with him or her. Similarly, if you are the butt of many jokes told by your colleagues one week, it makes a big emotional difference whether or not this is merely an unlucky co-occurrence or a concerted effort by your colleagues to bully you (Roth, 2011).

Philosophers have attempted to bridge this gap between parallel and joint activity by suggesting various ways in which the actions of individuals are "glued" together into a unified activity. By and large, philosophers have focused on the idea that a joint action is the outcome of a "shared intention" (or a "joint" or "collective intention") that unifies and coordinates the joint action. Philosophers are not the only academics that have taken an interest in joint action though—or to use another term, "cooperative action". Other creatures than adult human beings appear to do things together too. Consider, for example, a group of lionesses encircling and attacking their prey, or the social pretend play of two toddlers. Behavioural ecologists and ethologists have thus studied the cooperative abilities and behaviours of social carnivores such as hyenas and lions (Drea & Carter, 2009; Stander, 1992); marine

biologists have studied the cooperative hunting of social cetaceans such as killer whales (Nøttestad, Fernö, & Axelsen, 2002); primatologists and comparative and developmental psychologists have studied the cooperative abilities and behaviours of non-human primates (e.g. Boesch & Boesch, 1989; Chalmeau & Gallo, 1995; Melis, Hare, & Tomasello, 2006; Petit, Desportes, & Thierry, 1992) and young children (C Brownell, 2011; Carpenter, 2009; Hamann, Warneken, & Tomasello, 2011; Warneken, Chen, & Tomasello, 2006). Concepts such as 'joint action' and 'cooperative action' are rarely explicated in detail by these empirical researchers, but they are not referring to mere aggregates of individual actions. They take cooperation to involve several agents acting in pursuit of a "common goal", a "joint goal", or perhaps, a goal that is "shared".

What these empirical researchers are referring to arguably cannot be what most philosophical accounts of joint action are accounts of. Up until very recently, philosophers have exclusively aimed to elucidate the mundane joint activities of adult human beings, and have worked with examples such as two people painting a house together, going for a walk together or making a hollandaise sauce together. According to most influential accounts, for two agents to genuinely do these things together, they must have quite sophisticated capacities for planning, mindreading and communication (e.g. Alonso, 2009; Bratman, 1992, 1993; Gilbert, 2009; Pettit & Schweikard, 2006).[1] Such accounts thus make participation in genuine joint activity too conceptually and cognitively demanding for agents such as social carnivores, non-human primates or very young children.[2] For example, according to Michael Bratman's account of what he calls "shared intentional activity"—an account that I will use as a foil throughout the thesis—for us to be walking together, our walking must be caused by a mutually known pattern of mental states that includes the intention of each of "that we go for a walk together" and the intention of each that

[1] Notable exceptions include the accounts of Searle (1990), Tollefsen (2005), Tollefsen and Dale (2012), Pacherie and Dokic (2006), Pacherie (2007; 2011), Butterfill (2012), and Gold and Sugden (2007a).

[2] But see Carpenter (2009) for a defense of the idea that even 1-year-old infants meet the demands of Bratman's account of shared intentional activity.

we do this by way of the other's intention "that we go for a walk together" (I present this account in detail in section 1.2.1). For an agent to be a participant in such a shared intentional activity, she must thus have the concept of "our activity", be able to have intentions about the intentions of others, and be able to have mutual or common knowledge with other agents. There thus appears to be a lacuna in the philosophy of joint action.[3] This lacuna may lead to a myopic and misleading picture of joint action and cooperation, according to which an apparently joint activity either really is the genuine thing—a case of shared intentional activity or "true collaboration" (Tomasello & Hamann, 2012)—or merely a collection of purely individualistic but possibly coordinated activities that are performed in parallel. One aim of this thesis is to help philosophers and cognitive scientists avoid this false dualism.

The second aim of this thesis is to address another lacuna in the philosophical literature, namely how participants together manage to coordinate their actions while executing a prior shared intention. Leading philosophical accounts of joint action have tended to explicate the "jointness" of joint activity in terms of a prior shared intention—a shared plan—that distributes roles and tasks between the participants. But they have nothing to say about the intentional structures and subpersonal mechanisms that might enable the online coordination involved in carrying out such a shared plan. (By coordination that is "online", I mean coordination during action execution, which is performed as the joint activity unfolds. By contrast, "offline" coordination is coordination that is specified or carried out in advance of action execution, for example by recourse to plan-intentions that specify in advance what each participant should do.) Nor do these accounts have much to say about the possibility of unplanned but spontaneously coordinated executions of joint actions, such as that of two people on a street suddenly rushing to catch and steady a third pedestrian who has just tripped and fallen.

[3] See Tollefsen (2005), Butterfill (2012) and Pacherie (2011) for earlier attempts to fill the lacuna. Both Tollefsen and Butterfill intend their accounts to be applicable to the joint activities of young children in particular. I touch on aspects of Butterfill's (2012a) account in chapter 3, and I discuss and elaborate Tollefsen's (2005) account in chapter 4.

In the next section, I consider what the rationale is for taking shared intention to be the linchpin of genuine joint action. I then go on, in section 1.2 to discuss two accounts of what shared intention might be: Bratman's account of shared intentional activity and Christopher Kutz's account of "joint action as such". Despite the different starting points of these accounts, Bratman and Kutz agree about what the sociopsychological characteristics of a core set of joint activities are, namely the set of small-scale egalitarian joint activities such as that of two people painting a house or cooking dinner together. In light of the cognitive and conceptual demands that these sociopsychological characteristics impose on participants, I then give a more detailed exposition of the two lacunas that I will be concerned with in section 1.3. Finally, in section 1.4, I provide a brief roadmap of the thesis.

Note that I sometimes talk of a "joint *action*" of several agents, sometimes of their "joint *activity*". Similarly, I sometimes talk of an agent's contribution as that agent's "action", sometimes as her "activity". These terms do not pick out a hard and fast distinction, but I will typically use the term 'action' when I am concerned with a relatively simple (individual or joint) intentional performance, and the term 'activity' when I am concerned with a relatively complex and extended (individual or joint) intentional performance. For example, a particular turn in a dance is a joint action while the whole dance is a joint activity.

1.1 From joint action to shared intention

One way of making the move from joint action to shared intention starts from the observation that all the bodily movements that are part of a joint action are bodily movements of individuals, so what glues those movements together into a joint action must be some "internal component" (Alonso, 2009, pp. 444–445). Along similar lines, one may note that the observable behaviours of agents who perform actions in parallel and of agents who are engaged in joint activity may be indistinguishable to an outside observer (Gräfenhain, Behne, Carpenter, & Tomasello, 2009, p. 1430; Searle, 1990, pp. 402–403). Perhaps we happen to walk

alongside each other to the cinema, end up next to each other in the ticket cue to the box office, and we then accidentally end up sitting beside each other in the cinema (although by that point, each of us would probably suspect that the other was a stalker!). It is conceivable that at least some joint activities could be behaviourally indistinguishable from their merely-acting-in-parallel counterpart activities. For example, Searle stipulates that a case where people in a park run for shelter in the rain could be behaviourally identical with a case where people in the same park together perform an outdoor ballet (Searle, 1990, pp. 402–403). Again, this suggests that there must be an internal component that makes the difference between a mere 'multi-agent activity' and a joint activity. (I will use the term 'multi-agent activity' to refer the actions of two or more individuals considered as an aggregate, which may not be unified by anything besides their spatiotemporal proximity.) Alonso claims that this internal component is a shared intention: "the key property of joint action lies in […] the participants' having a 'collective' or 'shared' intention to so act." (pp. 444–445) Searle draws the same conclusion.[4]

The move from joint action to shared intention is often made by way of an analogy with individual action. One role that intentions play in individual agency is to imbue actions with intentionality, that is, to make behaviour purposive or goal-directed. It may be thought that just as an individual action is intentional under a description (Anscombe, 1969), a joint action is intentional under a description of the action as joint (Pacherie 2011; Tuomela 2007). Since it is often assumed that an action is

[4] Alonso and Searle have very different views on what a collective or shared intention is though. Note that the important difference-making internal component need not be an intention. It may also be something that occurs upstream of intention-formation. For example, Gold and Sugden (2007a) argue that the intentions of the participants in a joint action may be ordinary intentions that merely specify the agent's own actions. What distinguishes joint action from a collection of individual actions on their view is the kind of *reasoning process* that leads up to the formation of these intentions. According to Gold and Sugden's team-reasoning account, joint actions are the outcome of intentions that have been formed in response to the question "What should *we* do?", rather than the question "What should *I* do?" (Bacharach, 2006; Gold & Sugden, 2007a, 2007b; Pacherie, 2011). For each of two agents to frame their strategic situation in terms of the former question is for each of them to conceive of themselves as a member of a group of which the other agent is also a member.

intentional under a particular description in virtue of the intention that cause the action, it is natural to think that a joint action is intentional under a particular description of the action as joint in virtue of a shared intention that cause the joint action (Pacherie, 2011, p. 175). The activity of "you and I going to the cinema together" may be jointly intentional under that description, but the same activity may not be jointly intentional under the description "you and I leaving home together so that Alf and Bo can prepare a surprise party in your honour." (Kutz, 2000, p. 21) The former, the thought goes, is jointly intentional in virtue of the fact that we have a shared intention to go to the cinema together.

It is fair to say that the starting point for the proponents of all accounts of joint action discussed in contemporary analytic philosophy is some causal theory of action (but see Seemann, 2009a for a rare exception). Both in the individual and the joint case, it is assumed that an intention fails to explain an action if it does not cause the action in the appropriate way (Davidson, 1963). Little has been written about what the appropriate causal relation between shared intention and joint action is supposed to be though. Bratman says that the causal connection from the shared intention to the joint action "will go by way of mutual responsiveness of each to each in a way that tracks the joint action." (1992, pp. 338–339, 2009a, p. 159, also 2009b, p. 53) Without this appropriate causal connection, the shared intention doesn't explain the joint action according to Bratman. In this thesis, I assume that some version of the causal theory of action is correct.

1.2 Two accounts of joint action

In this section, I present Bratman's constructivist account of "shared intentional activity" and Kutz's minimalist account of "joint action as such". These accounts represent two candidate answers to the question of what a shared intention is.[5] One purpose of presenting these accounts is to introduce some philosophical puzzles that arise in the debate about what joint action is. By juxtaposing these two accounts, I

[5] Kutz doesn't gloss what he is doing as giving an account of what shared intention is, but his account can be construed in this way without misrepresenting it.

also want to make an important methodological point, namely that the label 'joint action' or 'joint activity' can be appropriately applied to many different phenomena, and it is far from obvious that an informative and interesting account of joint action *as such* is possible. The two accounts have different conceptions of what the phenomenon to be explained is, and Bratman and Kutz aim to meet slightly different constraints in developing their accounts. Despite this, the accounts share many assumptions regarding what is characteristic of small-scale egalitarian joint activities such as that of two people going for a walk together.

I have chosen to present Bratman's account because it outlines an intentional construction that in my view manages to capture many features that seem to be characteristic of many joint activities of everyday life. Furthermore, the construction is built out of ordinary personal plan-intentions and beliefs, components that are arguably already available within a satisfactory account of individual human agency. It is also unusually explicit, and thus provides a clear picture of what the conceptual and cognitive requirements are for participating in a shared intentional activity. In the rest of the thesis, I will primarily use Bratman's account as my point of reference. Another reason for picking Bratman's account in particular is that it has been influential not only in philosophy (Alonso, 2009; Pettit & Schweikard, 2006), but also in developmental psychology (Carpenter, 2009; Tomasello & Hamann, 2012)[6] and in artificial intelligence research on multi-agent systems (Georgeff, Pell, Pollack, Tambe, & Wooldridge, 1999; Wooldridge, 2000). Kutz's work has not had this interdisciplinary impact, but it is a distinctive account that can throw contrasting light on the Bratman's account.[7]

---

[6] The work of Margaret Gilbert and Raimo Tuomela has also influenced the design of experiments in developmental psychology (Gräfenhain, Behne, Carpenter, & Tomasello, 2009; Hamann, Warneken, Greenberg, & Tomasello, 2011; Hamann, Warneken, & Tomasello, 2011; Warneken, Chen, & Tomasello, 2006).

[7] Of course, other accounts could probably have been suitable in this introductory chapter too: For example, the accounts offered by Gilbert (1990, 2006, 2009), Raimo Tuomela and Kaarlo Miller (1988), Seamus Miller's (1995, 2001), John Searle (1990), Philip Cohen and Hector Levesque (1991), Abraham Roth (2003, 2004), Philip Pettit and David Schweikard (2006), or Natalie Gold and Robert Sugden (2007a; see also Pacherie, 2011).

1.2.1 Bratman's constructivist account of "shared intentional activity"

The target of Bratman's account of joint action is "cases of small scale shared intentional agency in the absence of asymmetric authority relations." (2009a, p. 150) This includes most types of everyday activities that one thinks of as being done together with someone else, such as you and I painting a house together, singing a duet, playing basketball, engaging in conversation, dancing tango, or going for a walk or traveling to New York together (Bratman, 1997, pp. 50–51, 1999, p. 130, 2009a, p. 150). Excluded are cases such as the moderated televised debate between the leaders of two political parties (where the participants are not merely acting as individual agents but as spokespersons for larger organisations), a strike initiated by a union, or the industrial manufacturing of an airplane (cases which both involve asymmetrical relations of authority and power). Bratman's account has been also been used by others to make sense of the joint activities young children, such as children's social pretend play, or two children building a block tower together (Carpenter, 2009; Tomasello, Carpenter, Call, Behne, & Moll, 2005; Tomasello & Hamann, 2012). However, Bratman's focus is on the shared intentional activities of "adult humans in a broadly modern world" (2009a, p. 153). His aim is to use the resources of his account of individual human agency (see Bratman, 1987, 2007), to account for the most "robust" cases of joint activity that these resources allows him to capture. Given this motivation and methodological constraint, Bratman is thus not himself concerned with the possibility that there are interesting cases of joint action to which his account cannot be applied.

According to Bratman, a shared intentional activity is a coordinated activity that is the outcome of a certain pattern of intentions and beliefs that are distributed among the participants. It is this pattern that Bratman identifies as the shared intention. To understand Bratman's view of shared intention, we must briefly look at his view of individual human agency first. To fully get a grip on human agency, we must according to Bratman acknowledge the important role that *planning* plays in our

agency. To do this, we must go beyond the belief-desire model of human agency, according to which our actions can be exhaustively explained by appeal to combinations of beliefs and desires that rationalise and cause them. Bratman argues that we must introduce another type of mental state that is irreducible to beliefs and desires: namely "plans" or—with another word—"intentions". What is distinctive of the mental state of intention is that it "involves a characteristic kind of commitment." (Bratman, 1987, p. 15) If I intend to *F*, then I am committed to *F*-ing. I have then settled on *F*, and thus, the issue of whether to *F* have been resolved. From then on, the *F*-ing cedes to be an object of deliberation and the commitment to *F* becomes a constraint on further deliberation. While I can change my mind and rescind my intention to *F*, intentions are relatively stable and resist reconsideration.

Intentions are thus both outputs of, and inputs to, practical reasoning. As an illustration, let us say that I, on a Thursday one week, form the intention to send the full thesis draft to my supervisor on Sunday. I am then committed to doing this, and the commitment will constrain my deliberation about how I should spend my Friday and Saturday. In view of my belief regarding how much I need to write to complete my draft and my belief about the hangover-inducing effects of alcohol, it will be irrational for me to intend to drink many cocktails at Penny's birthday party on Saturday (intending this while retaining the intention to hand in the full draft on Sunday would create an inconsistent web of beliefs and intentions; there is a *demand for consistency*). Given my commitment, I can also be criticised for being irrational if I don't start to deliberate about what to do in order to carry out my intention successfully (there is a *demand for means-end coherence*). I need to identify what sections that are missing from my draft, read up on the relevant literature, and so on. No such norms of rationality attach to the mental state of desire. I can desire to finish the thesis draft and send it to my supervisors on Sunday, while also desiring to drink many cocktails at Penny's party, even if I know that at most one of these desires can be satisfied. No one can criticise me for having such inconsistent desires. Having inconsistent desires may be unfortunate but it is not irrational (having mutually unsatisfiable desires is, alas, part of the human condition). (Of course, if I have those cocktails at Penny's party even if what I *most* desire, all things considered, is that I

hand in my thesis draft on Sunday, then I can be criticised for acting in a way that is inconsistent with what I most desire.) However, one cannot knowingly have such mutually unsatisfiable intentions without being criticizably irrational.

This difference between the norms associated with desire and the norms associated with intention flow, on Bratman's view, from the different roles that these states play in our cognitive economy. The role of desire is to influence our conduct so that we in general act so that we seek the good rather than the bad; the role of intention is to guide and coordinate our planning, and enable planning to *control* our conduct. Provided that I have not rescinded my intention to send the thesis draft to my supervisors before the end of the week, I will be irrational if I don't in fact perform the action when the time to act comes. While desires influence my conduct, they do not control it in this way. I am therefore not being inconsistent if I abstain from drinking cocktails at Penny's party, even if I have a desire to do so.

In sum, intentions control an agent's conduct and coordinate her planning over time. Now, it may seem that Bratman thinks that intentions are always directed at some future performance of an action, or some future state of affairs that one is committed to bringing about. But the plan-like commitments to act that Bratman calls intentions can according to him also arise spontaneously and be directed at the present rather than the future.[8] To distinguish this notion of intention as a plan-like commitment to act from a broader notion of intention as any kind of goal-directed state, I will sometimes use the term 'plan-intention' to indicate that the notion is the former "Bratmanian" one.

Bratman's account of joint action builds on his theory of individual planning agency in two respects. First, he identifies shared intention with a functional role that is analogous to the functional role of plan-intentions in this theory of individual agency. Secondly, the plan-intentions of individuals are the crucial building blocks

[8] Bratman does not have much to say about present-directed intentions though. For critical discussion of Bratman's treatment of present-directed intentions and spontaneous actions, see section 5.1.

of his account of a sociopsychological structure that may play the functional role of shared intention. Bratman's aim is to develop a "construction" of interrelated intentions and beliefs that can play a role in two or more agents' joint pursuit of a goal, which is analogous to the functional role that intention plays for an individual agent's pursuit of a goal:

> Our shared intention, then, performs at least three interrelated jobs: it helps coordinate our intentional actions; it helps coordinate our planning; and it can structure relevant bargaining. And it does all this in ways that track the goal of our [joint activity]. Thus does our shared intention help to organize and to unify our intentional agency in ways to some extent analogous to the ways in which the intentions of the individual organize her individual agency over time. An account of what shared intention is should explain how it does all this. (Bratman, 1993, p. 99)

Bratman then submits that the following sociopsychological structure can and does play the role identified by the functional specification:[9]

We intend to *J* 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).

---

[9] Bratman elaborates on this construction in his most recent papers on shared agency (2009a, 2009b). This elaborated construction makes some points about efficacy and stability of the participants' intentions more explicit, but these aren't important for my purposes here. The construction of (1), (2) and (3) is "View 4" in (Bratman, 1993, p. 106), although the construction is there presented as a set of individually necessary and jointly sufficient conditions. Since the paper "I intend that we *J"* (1997), Bratman has been open to the possibility that shared intention is a multiply realisable phenomenon. Note that in his most recent writings, Bratman sometimes writes as if he is merely presenting sufficient conditions for shared intention (2009a, 2009b). But this should not be taken as a further step back to an even less ambitious position. He still takes himself to be giving an account of "one important kind of shared intention" (2009a, p. 154, 2009b, p. 45). At his keynote talk at the Collective Intentionality VII conference in 2010, Bratman stated that the conditions were not merely sufficient but also necessary parts of an unnecessary (but sufficient) package of conditions. He made this point by analogy with the "inus"-clauses in John Mackie's account of causation (Mackie, 1965). I have not seen him make this analogy in print though.

(3) (1) and (2) are common knowledge between us.

For example, for you and I to be going on vacation together, this activity must be the outcome of a shared intention to go on vacation together. For this to be the case, each of us must have an intention of the form "I intend that we go on vacation together". Each must also intend this by way of ("in accordance with and because of") the other's intention. We do not intend to go on vacation together if I intend that we do this by way of me drugging you, locking you up in the trunk of my car, and driving us to the hotel. Each of us must also intend that our subplans for carrying out our intentions should be made consistent ("mesh" with each other). Suppose that I intend that we spend our time on vacation partying every night until late in the mornings. You, on the other hand, intend that we spend the early mornings on pleasant walks in the countryside instead. If we know that our subplans clash and neither of us is willing to adjust our subplans, then we don't have a shared intention to go on vacation together.

Note that condition (2) does not require that our subplans *are* in fact consistent. All that is required is that each of us *intends* them to be consistent. This makes us committed act in such a way that we can find a solution that we can agree on, possibly only after a process of bargaining or negotiation. There is no guarantee that we will agree on a solution, in which case our shared intention fails to lead to the intended shared intentional activity of going on a vacation together in the appropriate manner. On the other hand, if our subplans in fact happens to be consistent (both of us intends to go for morning walks) but one or both of us lacks the intention directed at this meshing of subplans, then there will be no shared intention to go on vacation together. This is because the successful coordination in this case would just be a lucky accident rather than an intended result. We need the pattern of attitudes that constitutes the shared intention—that is, the construction of (1), (2) and (3)—to achieve successful coordination as an intended result since that pattern will otherwise fail to play the functional role that Bratman identifies shared intention with. Of course, we can still "share" an intention in the loose sense that both intend "that we go on vacation together", but recall that this is not what Bratman means by the term

'shared intention': A shared intention is a construction that fills a certain role for interpersonal coordination in the pursuit of a common goal.

Finally, the fact that we have the appropriately related intentions specified in conditions (1) and (2) must be "common knowledge" between us. It is often suggested that for you and I to have common knowledge that *p*, it is not merely required that you know that *p* and that I know that *p*, but you must also know that I know that *p*, I must know that you know that *p*, you must know that I know that you know that *p*, I must know that you know that I know that *p*, and so on in an infinite concatenation of ever higher-order beliefs about *p* (following Schiffer's (1972) definition of what he calls "mutual knowledge*"). Bratman never explains why this third condition is needed, but merely says that "it seems reasonable to suppose that in shared intention the fact that each has the relevant attitude is itself out in the open, is public." (1993, p. 103) I will not say more here about the common knowledge condition, as chapter 2 is devoted to the role of common knowledge in joint action (as we shall see, a common knowledge clause is part of virtually all accounts of shared intention). At any rate, in the example under consideration, our intentions that we go on vacation together will presumably be common knowledge between us, since the intentions are likely to have been formed through a process of open and explicit discussion. Furthermore, this open discussion itself openly manifests that our intentions that we intend to go on vacation together are held in accordance with and by way of the other's similar intentions.

There is something puzzling about the idea that you and I are supposed to each *intend* "that *we J*". The content of my intention is not merely "to *F*", where *F* is the part of our *J*-ing that I am in direct control of (to lift my end of the sofa, say), and the content of your intention is not merely "to *G*", where *G* is the part of our *J*-ing that you are in direct control of (to lift your end of the sofa). We wouldn't have a shared intention if the following was the case: I intend to lift my end of the sofa, you intend to lift your end of the sofa, and each intends this because of the other's intention and by way of meshing subplans. Intuitively, this would merely be an unusually interdependent case of two agents acting in parallel rather than two agents intending

to do something together. It fails to account for the fact that when we lift and carry a sofa together, then the reason each of us lifts our end of the sofa is that we each intends that we lift and carry the sofa together. The intention "that we carry the sofa" is primary. So, the intentions "that we *J*" in condition (1) do appear to be needed.

The idea that both of us could intend "that we *J*" is puzzling though, because, in Velleman's words, it "involves the sharing of something that ordinarily seems indivisible." (Velleman, 1997, p. 35) Recall that an intention is a conduct-*controlling* attitude. If I am in control of our activity, if it is up to me whether *we J*, then I am free to intend that we *J*. How could you then also intend that we *J*? That would require that it would also be up to you whether we *J*, which would be incompatible with me being in control of our *J*-ing. If I delegate the intending to you, then the issue is no longer up to me, so I can no longer be part of the intending. But if I do *not* delegate the intending to you, then the issue will no longer be up to you, so you cannot take part in the intending. Or so it seems.

Bratman's planning theory of intention allows him to solve this puzzle though. While one cannot intend *to* perform someone else's action, one can intend *that* someone else performs an action. While I cannot intend to perform my guest's action of leaving before midnight, what I can intend is *that* my guest leaves at midnight. The planning conception of intention thus "allows us to be more liberal about what can be intended than we are about what can be attempted; for references to things other than our own actions can function appropriately in our plans." (Bratman, 1992, pp. 330–331) To illustrate this possibility, Bratman provides the following mundane example:

> I can predict that if I ask you for the time you will tell me the time; so I can intend [that I[10]] get the time from you. This is not an unusual form of control of your action, but just ordinary predictability of ordinary agents. (Bratman, 1999, p. 155)

[10] Bratman's wording here is actually "so I can intend *to* get the time from you" (my emphasis), but this must be a slip. It is the only place (as far as I am aware) where he writes anything that suggests that his view is that one might be able to intend *to* perform another's action.

In a similar way, I can intend "that we go on vacation together" if I can assume (with justification) that you will intend "that we go on vacation together" when you become aware of my intention that we do so (see Bratman, 1997, sec. VI). This may be the case if I know you well and I am fairly sure that you will join me in going on the vacation.[11] Alternatively, I might first intend "that we go on vacation together if you intend that we go on vacation together". When you become aware of my conditional intention, you can simply intend (unconditionally) "that we go on vacation together", and as a result, I will intend the same.

There is also a more restricted worry about intentions "that we *J*" concerning their use in an account of joint activity. If the concept of 'our *J*-ing' that is part of the content of the participants' intentions is the very concept that Bratman is trying make sense of—that is, the concept of shared intentional activity—then the account is viciously circular (see Petersson, 2007). The solution is, of course, to appeal to some other less robust concept of "our" activity, which isn't the concept of a multi-agent activity that is the outcome of a shared intention. Here, Bratman appeals to "a concept of our [e.g.] painting the house together that involves only the idea that, roughly, we are each intentionally painting that house in ways that avoid collisions." (2009b, p. 47) Pettit and Schweikard, who give an account of joint action that is similar to Bratman's, appeal to what they call a 'joint performance', which "can be conceptualized just as a pattern of behavior in which our different efforts combine to effect a certain result." (2006, p. 29) Each participant thus intends a joint performance, rather than a joint action. Thus, according to both these proposals, there is a weak notion of joint activity, "that we *J*" or "joint performance", which is supposed to figure in the content of the participants' intentions. However, this is not

[11] Perhaps I don't need to have any well-grounded expectation about whether or not you will form the intention that we go on vacation together. In the absence of evidence, I may simply decide to *rely* on the proposition "that you will intend that we go on vacation together" in my planning and practical reasoning, even if I might think it is unlikely that you will (as Alonso, 2009 suggests). My reliance might be justified by pragmatic considerations. Perhaps I could not bear not spending my vacation together with you, and I judge that relying the proposition that you will intend to spend it on me will increase the slim chance that you will so intend.

the genuine and interesting concept of joint activity that is the target of their accounts.

According to Bratman, the small-scale egalitarian joint activities that we adult human beings engage in on a daily basis are clearly cognitively and conceptually very demanding. Participants must arguably have the concept of belief and the ability to accurately attribute beliefs to others (in order to have common knowledge). They must also be able to have various plan-intentions with complex contents. This means that they must posses the concept of joint activity ("our *J*-ing") as well as the concept of "meshing subplans". Furthermore, they must be able to have higher-order plan-intentions concerning the plan-intentions of others as well as be able to recognise others' higher-order plan-intentions that concern their own plan-intentions.

1.2.2 Kutz's minimalist account of "joint action as such"

There are clearly examples that we might think deserve the label 'joint action' even if they do not have much in common with the cases that Bratman's account is applicable to. In particular, Bratman's "adult humans in a broadly modern world" participate in many joint actions that are not the outcome of interlocking intentions, where very little information about what each individual contributes is common knowledge. Think of the thousands of people participating in a petition on the World Wide Web for example. If the goal of the petition is achieved, then it seems right for a signer of the petition to proudly pronounce: "We did it!". This claim would be true even if the signer had no idea who any of the other members belonging to this "we" are. Bratman's account is not applicable to such a case.

Kutz criticizes Bratman and others (such as Margaret Gilbert and Raimo Tuomela) for focusing on a narrow range of joint activities in light of which they develop their accounts, a range that is restricted to small-scale, highly reciprocal and more or less egalitarian joint activities. According to Kutz, the accounts of Bratman, Gilbert and Tuomela are illuminating when it comes to some kinds of joint action, but they fail to capture joint action in its full generality. Kutz's ambition is explicitly to give an

account of "joint action as such" (2000, pp. 4, 31): Joint action is *one* phenomenon that can be captured by his *minimalist* account of joint action. On his view, there is one ingredient that the case of two people going for a walk together and the case of thousands of people signing an online petition have in common, namely that the participants have what Kutz calls "overlapping participatory intentions". As Kutz uses the term 'participatory intention', this is an ordinary personal intention to perform an action that frames the action (for the agent herself) as a contribution to a collective outcome. Several individuals have *overlapping* participatory intentions when the collective outcome that their actions contribute to is in fact the same collective outcome (none of them needs to be aware of this fact though). Kutz argues that this is both necessary and sufficient for the actions of several agents to count as one joint action.[12] If I have the intention to sign an online petition for an asset freeze and travel ban on the corrupt elite of the neighbouring country Atlantis, then, when I enter my name and email address on the website, I "frame" what I am doing as something that contributes to the attempt to collectively, with the other signers of the petition, bring about the outcome that Atlantis' corrupt elite gets their assets frozen and are banned from traveling into my country. If you have an intention to perform a different action, signing *your* name and email address on the website, but frame that action as a contribution to what happens to be the same collective outcome as the one I am contributing to, then we have overlapping participatory intentions. If we both perform these actions then we are both participating in the same joint action of trying to influence our leaders to make certain decisions.

Note that in this kind of case, there is no mutual responsiveness between us: our decisions to sign the petition are completely independent and the joint action would take place even if neither of us participated. Each of us is merely joining a collective

[12] He claims that another ingredient that characterises joint activities is that each participant is favourably disposed to the possibility that the others become aware of their contribution. Kutz (2000, p. 6) calls this a condition of "mutual openness" (this should not be confused with what I call "openness" in chapter 2). However, he suggests that "perhaps there is even a limiting case of collective action in which the ordinary condition of mutual openness is overridden by strategic concerns." (p. 19 n. 40). At any rate, the condition of mutual openness is not important for my concerns here.

endeavour that is already under way. This means that Kutz's participatory intentions are different from Bratmanian intentions "that we *J*". A participatory intention does not *settle* whether the joint action occurs or not, but merely settles whether or not the subject of the intention contributes to it. As Kutz puts it, neither of us has the "executive perspective" on the joint action that is required for intentions "that we *J*". Of course, some of the participants involved in the petition will probably have such an executive perspective. Perhaps the petition and website was launched by a single individual. This person is indeed settling whether or not the collective attempt to hassle Atlantis' corrupt elite will occur or not.[13]

For Kutz to succeed in showing that he has given an account of joint action *as such*—a set of individually necessary and jointly sufficient conditions for a collection of actions to constitute a joint action—he must show that there are no forms of joint action where participants lack participatory intentions. If there is a case of joint action where participants do not have overlapping participatory intentions, then his account cannot be an account of joint action as such. Kutz is not really concerned with this challenge though, as he primarily has in mind an objector that thinks his account is too liberal rather than too conservative. If he were concerned with it, then perhaps he would appeal to the idea that a joint action must be jointly intentional under some description, which implies that the content of the participants' intentions must refer to some collective outcome or activity.[14] But might it not be enough that several agents perform actions that happen to be aimed at

[13] Kutz (2000) illustrates the difference between participatory intentions and intentions "that we *J*" by considering the case of an orchestra playing *Eroica* (pp. 23–24). While the cellist arguably merely has an intention to contribute the performance, the conductor has (let us assume) the power to settle whether or not they play the *Eroica.*

[14] Kutz emphasises the importance of the idea that the characterisation of joint action mirrors the contents of the participants' intentions (see p. 27), but he does not explicitly appeal to this idea in order to defend the necessity of participatory intentions. He also draws on the team-reasoning literature, according to agents must be able to think of themselves as acting qua members of a group that acts together in order to rationally solve what appears to be trivial game-theoretical coordination problems (2000, pp. 7–9; for accounts of team reasoning, see Bacharach, 2006; Gold & Sugden, 2007b).

one and the same goal, and this goal is achieved as the combined result of their efforts? This can occur without any of the agents regarding what they are doing as a contribution to a collective outcome. Consider the following case: Pontus and Petter live in different top-floor flats in the same building. They are both tormented by the fact that the building's staircase has become full of trash. When Pontus is about to leave for work one morning, he forms the intention "that the staircase gets cleared of trash"—thus incorporating this end in his planning and practical reasoning—and he starts to pick up trash as he descends down the staircase (note that this intention is not a participatory intention since he does not *conceive* of his activity as contributing to a collective outcome). As Pontus leaves the flat, Petter comes home from his night shift at work and enters the staircase from below. He also forms the intention "that the staircase gets cleared of trash", and starts to pick up trash as he ascends up the stairs. Halfway down/up the staircase, they meet each other and each realises that his intention has been satisfied thanks to the actions of both of them.[15]

Would it not be correct here to say that "they cleared the staircase of trash" in a collective sense of "they", as opposed to the distributive sense of "they" in use when we say that "they entered the staircase"? We could say that "they *jointly* cleared the staircase of trash", even if it might be stretching everyday language to say that "they cleared the staircase of trash *together*". Maybe we should not give much weight to such linguistic intuitions, but Kutz himself does so when he discusses a case that he takes to demonstrate that overlapping participatory intentions are sufficient for joint action:

> Suppose that while we are having a picnic, it begins to rain. I jump up, grab the sandwiches and head for the car. I intend to do my part of our saving the picnic, hoping you will simultaneously grab the drinks and the blanket. If you do, then *it is reasonable to say we will have jointly saved the picnic*. (Kutz, 2000, p. 18 my emphasis)

[15] This example was inspired by a similar example in (Butterfill, 2011) involving two communists who independently of each other both set out one dark night to paint the same large bridge red. Starting from different ends, they meet each other at the middle of bridge and discover that they have both succeeded in painting the bridge red.

I agree with Kutz's linguistic intuition here. But if this shows that it is sufficient for two or more actions (performed by different agents) to constitute a joint action that they are the outcomes of overlapping participatory intentions, then a similar intuition about the case of Pontus and Petter ought to demonstrate that it is sufficient for two or more actions (performed by different agents) to constitute a joint action if they are aimed at the same goal and the achievement of this goal is the combined result of these actions. It is just as (in)appropriate to say that the picnickers saved the picnic *together*, as it is to say that Pontus and Petter cleared the staircase of trash together. My point here is not that we ought to rely heavily on such intuitions about what it is appropriate to say, but merely that if this is what we do rely on, then it seems that not even overlapping participatory intentions are necessary for joint action as such. (Since neither Pontus nor Petter conceives of their own activity of picking up trash as a contribution to a collective outcome, neither of them has a participatory intention, and thus, nor do they have overlapping participatory intentions.)

Despite presenting his account as radically different from that of Bratman, Kutz's account is actually largely in agreement with it regarding the characteristics of small-scale joint activities such as that of two people painting a house together or two people going for a walk together (see pp. 17, 21–22, 24, 27). In such cases, we must attribute intentions that concern the whole activity, intentions "that we *J*", or as Kutz calls them, "group-intentions" (2000, pp. 22, 24).[16] Furthermore, the intentions of the participants of these activities have common knowledge of each other's intentions and these intentions must be related in ways that ensure mutual responsiveness between the participants. But according to Kutz, the contributions of the participants to the joint activity (our *J*-ing) will always be directly caused by their participatory intentions and only indirectly by their group-intentions. Participatory intentions thus

[16] However, Kutz at one point express doubt about whether plan-like commitments to act in favour of some state of affairs being brought about, such as "that we paint the house together" are really the same kind of attitude as that of intentions to perform certain actions: "One might ask whether there really are such non-standard intentions, or whether they are instead figures of speech, either expressing a hope that we will do something, or standing in for an individual intention to promote our doing something." (2000, pp. 21–22)

play an ineliminable role, according to Kutz, even in cases of small-scale highly interdependent and egalitarian joint activity.

1.2.3 Discussion

I will here make two points in light of Bratman's and Kutz's approaches to joint action and their respective resulting accounts. The first point is that we should not assume that there is *one* phenomenon that is the target of various accounts of joint action. Bratman never claims that "shared intentional activity" should be taken to be joint action *as such*. First of all, he sets aside cases that involve what he could call "robust sociality", in which participants have different authorities and powers, act on behalf of organisations and social institutions, and so on.[17] Neither is he primarily trying to explicate a "folk notion" of joint action that is supposed to be in play when we in everyday discourse talk about people jointly doing something or doing it together (even if such talk often do refer to shared intentional activities). Hence, Bratman's and Kutz's accounts are not really in conflict.[18] Kutz explicitly concedes that there are kinds of joint action that are richer and more robustly joint than his minimalist account is able to fully capture. Such kinds are more fully explicated by an account such as Bratman's. As I argued in the last subsection, I do not think that Kutz actually gives an account of joint action *as such*, since there seem to be cases of joint action that do not involve overlapping participatory intentions. Indeed, I think there is little reason to think that we can find individually necessary and jointly sufficient conditions for joint action *as such*.[19] And even if we could, it is not clear why such a broad universal account of joint action would be interesting.

---

[17] In other words, Kutz is unfair when he writes that "Bratman and Tuomela and Miller have been *misled* by their reliance on […] cases of small-scale, highly interdependent, and non-hierarchical cooperation […]." (2000, p. 22 my emphasis)

[18] At least if Bratman accepts that shared intentional activities always involve participatory intentions derived from the intentions "that we *J*".

[19] Kutz briefly raises the possibility that "[c]ollective action types might simply hang together in a familiar familial fashion", but then goes on to argue that this is not the case (2000, p. 4).

The second point I want to make is that despite their divergent starting points and vocal differences, Bratman and Kutz are in agreement about the socio-psychological characteristics of a certain class of joint action, namely those that are small in scale, non-hierarchical, and where participants are in each other's presence (or at least in direct communicative contact). This will be the type of joint activities that I focus on in this thesis as well. For both Bratman and Kutz, participants in these activities have plan-intentions "that we *J*" (but see footnote 16). This is not equivalent to merely being committed to perform one's contributory actions that would be part of the joint activity, should it occur. Rather, the commitment concerns these actions as parts of the whole, where the whole is, so to speak, implicated in the "parthood" of these contributory actions. Secondly, the joint activity involves mutual responsiveness of intentions and actions between the participants. Finally, they agree that the relevant intentions and beliefs are common knowledge among the participants.

1.3 Two lacunas in the philosophy of joint action

In this thesis, I focus on two relatively unexplored lacunas in the joint action literature. On the one hand, I explore the joint activities of relatively cognitively unsophisticated agents, such as young children and non-human primates, and on the other hand, I look at how participants carry out shared intentions and achieve online coordination of actions and movements.[20] To explore the latter lacuna is important because it highlights that patterns of interlocking plan-intentions, mutual belief and so on, are only the tip of the iceberg of what enables and sustains joint activities. Exploring the first lacuna is important because it undermines a false dichotomy between purely individualistic actions on the one hand, and forms of cooperative action such as "shared intentional activity" that are the outcome of the participants'

[20] Tollefsen and Dale (2012) identify a number "problems" that face various philosophical accounts of joint action. Two of these problems are the "over-intellectualization problem" and the "execution problem". The former concerns the "myopic" nature of the philosophical accounts, and the latter concerns the their "skeletal" character. I have adopted these adjectives too to frame my project. But given motivations such as Bratman's, I think it is misleading to talk about the myopic and skeletal character of philosophical accounts as *problems*. They are more accurately taken to be *lacunas*.

shared intention. Note that the joint activities of, for example, young children or non-human primates, are typically carried out by participants who are in each other's presence and are mutually responsiveness, and who can more or less be considered as equal contributors to the activity. They are in this way similar to Bratman's cases of modest sociality (they certainly do not involve participants who are acting on the behalf of groups or who have institutionally sanctioned powers and responsibilities). Furthermore, there is good reason to think that we can learn something about the small-scale joint activities of adult humans by looking at their developmental precursors and at the joint activities of species that are closely related to us. In addition, trying to understand the cooperative activities of species with different cognitive capacities and limitations is arguably an interesting project by itself.

1.3.1 Joint activities with cognitively unsophisticated participants

As I highlighted at the beginning of the chapter, young children and many non-human animals engage in forms of joint activity. Before the end of their first year, children start to engage in coordinated joint visual attention with caregivers and peers (see papers in Moore & Dunham, 1995). Around the same time they also start to engage with others in simple joint activities such as the building a block tower together, or taking turns digging a hole in the sand (Brownell, 2011; Hay, 1978; Verba, 1993). In their second year, engagement in "shared pretence scenarios" is also common (Harris, Kavanaugh, Wellman, & Hickling, 1993; Rakoczy, 2008). For example, two toddlers might pretend that they are engaged in a telephone conversation with each other, each holding a remote control as if it was a mobile phone. Arguably, chimpanzees and various social carnivores—wolves, hyenas, lions—engage in cooperative hunting (Boesch & Boesch, 1989; Drea & Carter, 2009; Stander, 1992).

These social non-human animals are not sophisticated "mindreaders", that is, they lack an explicit understanding of others as agents whose behaviour is driven by mental states such as beliefs, desires and plan-intentions. Arguably, children under the age of 3 also lack such a robust "Theory of Mind". Following Tollefsen (2005, p.

81), I assume that a robust theory of mind includes the following: (i) an understanding of other persons in terms of their thoughts, intentions, and beliefs; (ii) an understanding that other persons' thoughts, beliefs, and intentions may differ from one's own; and (iii) an understanding that others have thoughts and beliefs that may not match with the current state of affairs (false beliefs). Before 3 to 5 years of age, children fail at so-called elicited-response false belief tasks, in which they have to take into account an agent's false belief about (for example) an object's location in order to correctly answer a question about where the agent will look for the object (see Wellman, Cross, & Watson, 2000). Despite this, developmental psychologists have adopted Bratman's account of shared intentional activity to make sense of the joint action of young children (Carpenter, 2009; Moll & Tomasello, 2007; Tomasello & Hamann, 2012). While children's spontaneous behaviour (such as their looking behaviour) is sensitive to the mental states of others, including their false beliefs, already at 15 months of age (Baillargeon, Scott, & He, 2010; see Onishi & Baillargeon, 2005), such sensitivity does not reflect the kind of understanding of others that is required for participation in a shared intentional activity. Apparently in contradiction to this, Malinda Carpenter (2009) states that 12-month-olds "arguably show evidence of understanding something about others' intentions or plans for action—the means others have chosen to use to achieve their goals." (p. 382) She interprets this as showing that they understand something about Bratmanian plan-intentions and that, thus, the mutual responsiveness problem doesn't really arise. But Carpenter seems to use the notion of a 'plan for action' in a very broad sense that encompasses low-level control of movement sequences during action. Bratman's notion of planning is much more narrow. Recall that intentions are plan-like commitments that play a role in practical reasoning that is defined by characteristic norms of consistency and means-end coherence. Plans are furthermore hierarchically structured so that plans typically need to be filled in with subplans—also intentions—as a result of deliberation. Attributing such plan-intentions to others seems to require more than an implicit sensitivity to how others' behaviour is guided by e.g. beliefs and desires. Furthermore, some developmental psychologists claim that participating in joint activities plays a role in facilitating the development of sophisticated mindreading skills (e.g. Hughes, Fujisawa, Ensor, Lecce, & Marfleet,

2006; Moll & Tomasello, 2007; quoted in Butterfill, 2012a). As Butterfill (2012a) argues, if this is true, then Bratman's account of shared intentional activity cannot applicable to the joint activities that toddlers engage in.

These examples of joint activities with participants who are relatively cognitively unsophisticated are clearly not counterexamples to Bratman's account itself. Bratman himself is simply not interested in the joint activities of non-planning agents. There is certainly room in Bratman's overall picture of agency for simpler accounts of joint action. For example, there is a place for creatures who lack plan-intentions but who are still acting purposively: "Many animals, human and nonhuman, are purposive agents – agents who pursue goals in light of their representations of the world." (Bratman, 1999, p. 5) Furthermore, there are ways in which such agents are able to coordinate their activities over time despite their lack of planning capacities. He notes that "planning is not the only mechanism that coordinates an individual's purposive activity over time. A tiger hunting her prey may exhibit wonderfully coordinated activity without being capable of such planning." (1993, p. 101) While Bratman says nothing about what these mechanisms could be—that just isn't something he is concerned with—nothing of what he says about purposive agency suggests that there might not also be mechanisms that coordinate several individuals' purposive activities with regard to a common goal. Consider the hunting of a more social cat for example, such as the lion. Lionesses frequently hunt in groups, stalking one and the same prey from different directions before they trap the prey as it tries to escape. (I am not in a position to positively claim that these lionesses really have a common goal, but suppose that they do.) There may thus be patterns of mental states or perceptual and motoric processes that fulfil a functional role similar to the role that Bratman identifies with shared intention (that role would not include the facilitating the interpersonal coordination of *plans* though).

Cases of joint activity with young children or non-human animals are not obvious counterexamples to Kutz's minimalist account either. To show that they would indeed be counterexamples, we would have to show that the young children or non-human animals do not have overlapping participatory intentions, and I am not sure

how this could be shown. Nevertheless, there are clearly mechanisms that enable children and social animals to coordinate their cooperative activities, so even if they do in fact have overlapping participatory intentions when engaging in joint activities, that does not give us much of a grip on how the cooperative hunting of the lionesses differs from the case of two adult humans who independently of each other sign one and the same online petition.

What I want to emphasize here is that the view of joint action and cooperation that one gets from reading contemporary analytic philosophy of joint action will be somewhat myopic. An account such as Bratman's may lead one to adopt a false dichotomy between purely individualistic self-regarding actions on the one hand, and full-blown shared intentional activity on the other hand. This is partly encouraged by the fact that accounts of shared intention such as Bratman's are often prefaced by the observation that there are two different senses in which *we* might be doing something: *we*, considered as a collectively, might be walking (together); or *we*, considered distributively as a set of individuals, might be walking (in parallel) (Bratman, 2009a, pp. 150–151; Gilbert, 2009, p. 168; Kutz, 2000, pp. 1–2). The accounts that are then presented as accounts of what characterises activities where we are doing something collectively rather than distributively. This suggests a false dualism that perhaps is not actually embraced by the accounts' creators, but which may mislead others.

This false dualism is visible in a debate between Michael Tomasello and Christophe Boesch concerning whether or not chimpanzees are really collaborating when they hunt in groups. Tomasello and Katharina Hamann argue that only humans engage in what they call "true collaboration" (Tomasello & Hamann, 2012). Drawing on Bratman (1992), they assume that participants only truly collaborate when the following is the case:

> "(a) the participants have a joint goal or intention in the sense that they *each have the goal or intention that we (in mutual knowledge) do X together*; and (b) the participants coordinate their roles—their plans and subplans of action—including helping the other in her role as needed. (Tomasello & Hamann, 2012, p. 2 my emphasis; see also Tomasello, 2009, p. 61)

Tomasello and Hamann then contrasts such true collaboration with what they argue goes on in the group hunting of chimpanzees:

> During the hunting, whereas *each chimpanzee is trying to get the monkey for itself*, and not helping the others in their roles at all, human hunters do such things for their partners as giving them weapons, clearing trails for them, carrying their child or weapon, repairing a weapon, instructing collaborators in best techniques, and so forth (Hill, 2002). The short story is thus that chimpanzees have *no joint goal that "we" capture it and share it*, helping the other in his role as needed, and no sense of commitment in either direction. (Tomasello & Hamann, 2012, p. 8, my emphases)

I do not want to question whether or not Tomasello and Hamann are correct in construing the difference between humans and chimpanzees in the way they do. But the stark contrast they draw between humans and chimpanzees here leaves no room for a middle ground between the coordinated behaviour individuals who are each pursuing their own self-regarding goal (e.g. "to get the prey for oneself") on the one hand, and the rich form of cooperation that is exemplified in the case humans hunting together (see also Boesch, 2005, p. 692, who makes the same stark contrast, but puts chimpanzees in the realm of true collaboration). There is arguably theoretical space for the view that a hunter could have the goal "that the prey is caught", "to catch they prey", or even "that I catch the prey" without it being part of the conditions of satisfaction of this goal that it gets this prey "for itself" even if the hunter does not have the goal "that *we* catch the prey".

What we need are accounts that better fit what we know about the socio-cognitive profiles and motivations of various non-human animals and young children. We need a richer taxonomy of accounts of joint action in light of these constraints (Rakoczy, 2006, p. 124). My contribution to this project consists of chapters 2 and 3. In chapter 3, I give an account of a kind of joint action driven by what I call a "joint goal", which is intended to occupy this middle ground between the coordinated activity of purely individualistic agents and "true collaboration" (or "shared intentional activity"). Other accounts that have been developed for similar purposes are Butterfill's (2012a) account of joint actions driven by what he calls "shared goals", and Tollefsen's (2005) account of the joint activities of young children, according to which they may have a shared intention that is composed of a pattern of interconnected intentions-in-action and perceptual states (rather than plan-intentions

and beliefs). I touch on Butterfill's account in chapter 3, and discuss and elaborate Tollefsen's account in chapter 4 (and in Blomberg, 2011).

While making sense of the agency of non-human animals and young children is a fascinating pursuit in itself, I also believe that getting to grips with less sophisticated and simpler cases of joint action may transform the way we should think about the collaboration and joint action among us "adult humans in a broadly modern world". Besides its intrinsic interest, one reason for focusing on examples involving non-human animals and young children, rather than, say, simple spontaneous joint activities with adult humans, is that when adult humans come into the picture, it is much more tempting to interpret an episode of joint activity as being within the purview of some previous plan-like shared intention to act jointly (see e.g. Bratman, 1999, p. 139). Once cognitively less demanding accounts of joint action is available, this may thus lead us to view *some* episodes where we exercise our own agency together with others in a new light.

1.3.2 Online coordination of actions

As Tollefsen and Dale (2012) point out, the leading accounts of joint action are "skeletal" in that they merely focus on the personal-level mental states that are required for joint action. Furthermore, they focus on planning rather than the execution of a joint activity as it unfolds. They have nothing to say about the subpersonal mechanisms that might enable the online coordination involved in the executing these shared intentions, nor do they have much to say about the possibility of unplanned but spontaneously coordinated executions of joint actions.[21] The skeletal character of shared intention accounts of joint action arguably invites one to

[21] While Searle does indeed think that joint action can be spontaneous, and that we often act with what he calls a "collective intention-in-action", it is still true that his account does not have much to say about what this involves. He simply states that joint actions are the outcome of individuals who are acting on a special kind of psychological attitude, a we-intention. The collective or joint aspect is thus part of the attitude rather than the content. I briefly discuss Searle's account in chapter 4.

view joint action and cooperation through the frame of what Mark Risjord calls "the standard picture":

> [I]n any joint action, a shared prior intention sets the common plan and roles, while the individuals fulfill their different parts with actions that are individually intentional. The individual actions become part of a genuine joint enterprise only insofar as the agents commit to, accept, or have common knowledge of the goals and roles of a prior joint intention. There is, therefore, an asymmetry between individual actions and joint actions. For individual action, prior intention and intention-in-action are independent. An action can be intentional without a prior intention to do that act. By contrast, an action cannot be part of a joint performance without a joint prior intention. The intentions-in-action of several individuals are bound together as a joint action only in virtue of shared prior intentions. (Risjord, 2012, pp. 4–5)

As Risjord notes, this picture of joint action does not contain any resources to make sense of the idea that there might be spontaneous joint actions if spontaneous actions are intentional actions that are performed without a prior intention. (To this we can add that the picture makes it difficult to accommodate the idea that creatures without capacities for planning and mindreading could engage in joint action). However, I will argue, in chapter 4, that intentions-in-action can in some sense also be joint and so can the execution itself of a shared prior intention.

While Bratman says that "mutual responsiveness *in action*" is a constitutive feature of shared intentional activity, he does not explicate what we should take from this abstract label (1992, pp.338-339, emphasis in original). He does not have anything to say about what it may involve or how people achieve it. Of course, some joint activities will largely be coordinated by way of the meshing of plans rather than by online coordination of actions. For example, if you and I organize a workshop on the philosophy of joint action together, then much of the coordination work will consist in ensuring that our plans and the way in which we fill in those plans (that is, commit to subplans) are consistent.[22] But when it comes to other joint activities, knowing that the activity is the outcome of shared intention in Bratman's sense will tell us virtually nothing about how it is coordinated and carried out. Consider the case of

[22] Note that this coordination work is likely to itself by an example of shared intentional activity according to Bratman, namely a conversation (1999, p. 130).

two figure skaters performing a well-rehearsed joint acrobatic manoeuvre on ice. As Bratman frequently emphasises, plans are always more or less partial, either because the time has not come yet for "filling out" the plan with commitments on particular means for fulfilling it, or because the plan has reached a level of granularity where further planning is not necessary or even contra-productive. Bratman himself notes that "my plans will typically be at a level of abstraction appropriate to my habits and skills." (p. 31) In many mundane joint activities, this level of abstraction will arguably be found at a level that is above the level where triadic coordination between participants and their common goal—"joint action-tracking mutual responsiveness", to use Bratman's phrase—must occur in order for the joint action to be successfully and reliably executed. If you and I shake hands with each other[23], what accounts for the joint achievement of having our hands meet in the appropriate place, close our grips on each other's hands, and shake our hands in an appropriate manner will not be the meshing of our plan-like commitments. Rather, perception-action feedback loops and subpersonal mechanisms for action control and monitoring will play the crucial role here. Recent research in social and cognitive neuroscience suggests that such mechanisms may guide actions toward outcomes that are represented as collective outcomes as well as individual outcomes (see chapter 5).

Furthermore, it seems plausible that some joint actions are spontaneous, that is, that they are performed without the formation of states that incorporate plan-like commitments. Consider two people on a street who rush to help a third person who has tripped and fallen to his or her feet. Both act immediately and spontaneously in response to the accident.[24] Now, Bratman thinks that plan-intentions need not be future-directed but also present-directed. This would be a limiting case of a "partial" plan: The plan merely specifies the goal of the action and simultaneously causes an

[23] At least according to Pettit and Schweikard (2006), this is a type of joint action that is explained by appeal to a shared intention.

[24] This joint action gets initiated in a way similar to the joint saving of the picnic discussed by Kutz (see page 19), but the joint helping up and steading of the fallen pedestrian, I take it, involves many features that are characteristics of the kind of activities discussed by Bratman, such as mutual responsiveness and mutual awareness among the helpers regarding what they are doing.

action directed toward that goal. However, the only reason for thinking that present-directed plan-intentions are involved here seems to be a craving for a misplaced generality of theoretical scope. Besides the fact that there are good reasons to doubt that present-directed intentions are the same kind of state as future-directed intentions according to Bratman's own functional definition of intention (see footnote 8), it seems plausible that even adult humans in a modern world like ours sometimes act merely in a capacity as purposive agents, temporarily outside a framework of planning structures.[25]

In other words, in order to understand how joint action is coordinated during its unfolding performance and in order to make sense of the spontaneous joint actions, we need to go beyond an analysis that focuses on the structures of intentional states that are the distal causes of the actions of mutually responsive individuals. Chapters 4 and 5 are dedicated to this lacuna in the joint action literature. Besides the elaboration of Tollefsen's (2005) account that I present in chapter 4 (and in Blomberg, 2011), Pacherie (2007, 2012) has also given attention to online coordination of joint action, as have Tollefsen and Dale (2012). There is of course much interesting empirical work here, and the purpose of chapter 5 is to review some of it.

1.4 Roadmap of the thesis

In **chapter 2**, I argue that, even when restricting our attention to joint activities that are small-scale, highly interdependent, and largely egalitarian, there are several kinds of such activities. Some of these kinds are richer—more robustly "joint" in some sense—than others. The richer the form of joint activity, the more it demands of participants in terms of cognitive sophistication and conceptual resources. I argue for this pluralism about joint action by focusing on what motivates the common knowledge condition (CK-condition), which is a component in almost all philosophical accounts of shared intention and joint action. I identify and explicate

[25] At one point, Bratman admits that this is a possibility: "It may be, though, that we sometimes do things that are merely purposive […]." (2000, p. 35)

two motivations. The first is to rule out cases that involve "concealment" of the participants' intentions or (higher-order) beliefs about these intentions. There is, I submit, an everyday notion of a robust and rich form joint action that is incompatible with the presence of such concealment (this parallels the motivation for a CK-condition in Gricean accounts of communication (Clark & Marshall, 1981; Schiffer, 1972)). The other (related) reason for introducing a CK-condition is that joint action, like individual action, ought to be non-accidentally coordinated. When agents act jointly, they must coordinate the activity *together*. I argue that, since concealment cases can only be constructed if participants have the concept of belief, such cases need not be ruled out when it comes to participants who arguably lack this concept (such as young children and non-human primates). Joint activities in which such agents participate is thus of a less robustly joint kind than joint activities that sophisticated mindreaders such as adult humans beings sometimes engage in. The less robustly joint kind must arguably still involve non-accidental coordination though, so the goals of the participants still need to be "out in the open". I suggest that there is form of openness that does not require common knowledge, but which does require that agents have some awareness of a mutual goal-dependency. In addition, there is a notion of an even weaker form of joint action that may be accidentally coordinated.

In **chapter 3**, I offer an account of what it takes to have what I call a 'joint goal'. The account specifies a pattern of goals and beliefs that I argue enables agents to mutually benefit from the fact that they have a common goal. Two or more individuals can according to this account have a joint goal without anyone having the goal "that we *J*", and their goals only have to be directed at (more or less) the same set of states of affairs. The goals need not also represent that set of states of affairs in the same way. This account of a joint goal allow us to identify an interesting form of joint cooperative activity without shoehorning it into either side of a dichotomy between purely individualistic behaviour and "true collaboration" (Tomasello & Hamann, 2012). At the end of the chapter, I also consider what sorts of behavioural phenomena and experimental results suggest that a multi-agent activity is coordinated by a joint goal.

While chapters 2 and 3 deals with the myopic character of much of the philosophical work on joint action, I proceed to the second lacuna in chapters 4 and 5. In **chapter 4**, I give an account of joint actions characterised by mutual responsiveness in action, but not by mutual responsiveness in planning. I argue that in some joint activities, such as in a dance performed by two figure skaters, actions may be joint in virtue of each participant having a "socially extended intention-in-action" with conditions of satisfaction that range over not just their own bodily movements but also over the bodily movements of the other(s). Following Tollefsen (2005), I suggest that the structure of Bratman's account of shared intention can be modified so that it is formulated in terms of Searlean intentions-in-action (Searle, 1980, 1983) rather than in terms of plan-intentions. But this requires something that many philosophers of action take to be impossible: that one could intend not only *that* someone else perform an action, but also intend directly *to* perform another agent's bodily movements. I argue in chapter 4 that this is in fact possible under some circumstances. A stepping-stone in my argument concerns what the structure of the content of one's intention-in-action is when one directly manipulates a tool. I argue that in the case of "fluent" tool-use, the content of one's intention-in-action range not merely over his or her bodily movement, but also over the movement of a tool. I thus claim that an agent's repertoire of basic actions may include both tool-using actions and actions that include the performance of another agent's bodily movements.

Chapter 5 is essentially a review chapter of various strands of empirical work on joint action. The chapter is motivated by an attempt to answer the following question: How do participants successfully coordinate their actions in joint activity in the absence of plans or beyond what their plans specify? The main point of the chapter is to show that in many cases, joint action is not something people need to work hard to get off the ground, but something that comes naturally to us. I review research in cognitive psychology and in social and cognitive neuroscience that suggests that there are various subpersonal mechanisms by means of which joint actions can be coordinated.

Finally, in the **conclusions**, I sum up the main points of the thesis and offer some thoughts about future work that takes off from the research on which this thesis is based.

2 Common knowledge, openness, and joint action

According to most accounts of joint activity, in order for two or more agents to be acting together, they must have "common knowledge" (or "mutual knowledge") of each other's goals or intentions concerning the activity (Alonso, 2009, p. 458; Bratman, 1992, p. 335, 1993, p. 103, 2009a, p. 160; Cohen & Levesque, 1991, p. 491; Gilbert, 1990, p. 7, 2008, p. 502; Miller, 2001, p. 59; Pettit & Schweikard, 2006, p. 24).[26] As it is often metaphorically put, the attitudes have to be "out in the open", "in the public domain" or be "above board" between the participants. On this basis alone, it may seem legitimate to claim, following Chant and Ernst (2008, p. 550), that common knowledge plays a "central role" in these accounts. However, it is not always clear what central role this is supposed to be. The common knowledge condition (henceforth CK-condition) is only rarely explicitly motivated. For example, Bratman (1993) introduces the CK-condition by simply asserting that "it seems reasonable to suppose that in shared intention the fact that each has the relevant attitudes is itself out in the open, is public." (1993, p. 103, see also 1992, pp. 334–335)[27] Similarly, Miller claims that "mutual knowledge is what distinguishes joint action from interdependent action that is not joint", but never explains what it is about mutual knowledge that gives it this status as a distinguishing feature (2001, p. 60). This lack of explicit motivation is surprising given that the "cost" of

[26] As we saw in chapter 1, Kutz argues that common knowledge is *not* constitutive of joint action *as such*. But he repeatedly suggests that common knowledge is constitutive of *a certain kind of joint action*, namely small-scale, highly reciprocal and broadly egalitarian joint activities (Kutz, 2000, pp. 6 n. 8, 17, 27). See also footnote 29.

[27] In Bratman's more recent work on shared agency (2009a, 2009b), the CK-condition has been modified somewhat and receives a more explicit motivation: see footnote 34. Gilbert requires that in order for agents to have a shared intention to do something, it must be common knowledge between them that each has expressed their personal readiness to commit them all to intend as a body to do it (see also Gilbert, 1990, p. 7, 2008, p. 502). While it is clear that this state of common knowledge is what transforms all the individual expressions of personal readiness into the joint commitment, she does not elaborate on what it is about common knowledge that allows it to play this transformative role.

incorporating a CK-condition appears to be high; the CK-condition seems to make accounts of joint activity very cognitively demanding.

It is often claimed that a proposition *p* is common knowledge among the agents in a population if and only if the following is the case:

(everyone knows that)$^n p$

for *all n* (Chant & Ernst, 2008, p. 553; Kutz, 2000, p. 6 n. 8; Miller, 2001, p. 59; see also Schiffer, 1972, pp. 30–31). This means that, if it is common knowledge between you and I that, say, each of us intends "that we *J*", then everyone (that is, you and I each) knows…

…that each intends "that we *J*" ($n = 1$)

…that everyone knows that each intends "that we *J*" ($n = 2$)

…that everyone knows that everyone knows that each intends "that we *J*" ($n = 3$)

and so on and so forth *ad infinitum*.

In light of this iterative definition of common knowledge, it may appear that common knowledge is unreachable by mere mortal adult humans. To have common knowledge, agents would have to know an infinite number of facts about their own and other agents' higher-order knowledge (I will explain in section 2.1 why this infinite iteration of higher-order knowledge states seems to be necessary to adequately characterise what it is for a fact to be "out in the open"). Furthermore, these facts include facts about higher-order knowledge of an infinitely high order! In light of this, some philosophers have suggested that the CK-condition should be replaced by something weaker.[28] Kutz argues that the condition may be too strong since common knowledge is a "cognitively demanding state" (2000, p. 6 n. 8).[29]

[28] Others appeal to an "unanalyzed" (Bratman 1992, p. 335 n. 15, 1993, p. 103 n. 20) or "intuitive" notion of common knowledge (Bratman 2009, p. 160; Alonso 2009, p. 458 n. 44). Perhaps they do so because the analyses of common knowledge that are available strike them as psychologically implausible.

[29] Kutz suggests that Sperber and Wilson's (1986) "weaker, and so more pliable, notion of mutual manifestness" is more appropriate than common knowledge (2000,

Similarly, Pacherie (2007) claims that "[t]he condition of mutual knowledge […] is difficult to satisfy and one may wonder whether such a strong condition is really necessary" (p. 166).[30] I will argue that there is at least one account of common knowledge—namely David Lewis' account—that is not susceptible to this general critique of common knowledge as being too cognitively demanding for finite beings like us.

I will primarily be concerned with a more specific worry about the CK-condition in this chapter. Even if common knowledge does not (contrary to initial appearances) require of agents that they have infinite minds with infinite time on their hands, common knowledge arguably still requires of agents that they have the concept of belief. This is not a problem for the idea of common knowledge itself or for its applicability to the joint activities of cognitively normal adult humans. But it creates a dilemma for those who take the CK-condition to be a *necessary* condition on a multi-agent activity being a joint activity.[31] As I pointed out in chapter 1, it appears on the face of it that agents who lack the concept of belief—such as young children and some non-human primates for example—do participate in joint activities. In effect, this appearance is either deceiving, or the CK-condition must be rejected as a necessary condition on what it takes for a multi-agent activity to be a genuine joint activity (Tollefsen, 2005; cf. Breheny, 2006).

---

p. 6 n. 8). The notion of mutual manifestness is very close (but not identical) to Lewis' notion of common knowledge though. As we shall see, contrary to what Kutz thinks, Lewis does not adopt the iterative definition of common knowledge (see footnote 41).

[30] Pacherie rejects the CK-condition in favour of a condition of what she calls "mutual presumption" (p. 171). This condition is satisfied if the participants act on the "implicit assumption that other human agents are sufficiently cognitively similar to us that their attitudes and intentions can be successfully simulated or inferred." (Pacherie 2007, p. 171)

[31] As I explained in chapter 1, the conditions in Bratman's account are not necessary conditions tout court. Instead, each of the three conditions is a necessary condition of a package that is sufficient for a shared intention to be in place, but this package is itself not necessary. There may be other packages of conditions that are also sufficient for a shared intention to be in place. I will henceforth set this complication aside in this chapter.

In this chapter, my aim is to accomplish the following: First, I want to defend the incorporation of a CK-condition in accounts of joint action. I claim that there is an intuitive notion of a rich and robust kind of joint action that does require that the goals or intentions of the participants are common knowledge among them. When it comes to such joint activity, in which adult human beings often participate, a CK-condition is needed to rule out cases of multi-agent activity that introduce elements of "concealment" as cases of this robust kind of joint activity.[32] But the CK-condition also accomplishes something else: it allows us to make sense of the idea that agents can non-accidentally coordinate a multi-agent activity together. Arguably, that coordination of a multi-agent activity is non-accidental is required for the activity to count as a truly joint activity. Just as a complex individual activity is composed of non-accidentally coordinated component actions directed toward a goal (think of all the steps involved in preparing and cooking a meal for example), so a joint activity should be composed of non-accidentally coordinated component actions directed toward a goal. Secondly, I will argue that the introduction of a CK-condition is not the only way of cashing out the idea that joint activity is non-accidentally coordinated (even if common knowledge is necessary for ruling out the aforementioned "concealment" cases). Common knowledge is sufficient but not necessary: There are other forms of "openness" that can rule out accidentally coordinated multi-agent activities as examples of truly joint activities. Crucially, there is a form of openness that does not require of agents that they have the concept of belief.

The upshot of this is that even when we restrict ourselves to small-scale highly reciprocal joint activities with only a few agents who are in each other's presence, joint activities are not of one and the same sociopsychological kind. We do not have to consider cases such as online petitions, flash mobs, or large-scale industrial action to show that there are kinds of joint activities that differ from, for example,

[32] These concealment cases are similar to the cases that have motivated a CK-condition in Gricean accounts of communication (Clark & Marshall, 1981; Schiffer, 1972, pp. 17–30).

Bratman's shared intentional activity. Cases of small-scale highly reciprocal joint activities such as social play and coordinated group hunting can demonstrate this too. Some such cases are of a kind that is more richly or robustly joint than others. The more robust forms require that the participants have more sophisticated socio-cognitive capacities. I particular I argue that there are at least two everyday notions of small-scale and highly reciprocal joint activity, and these correspond to two sociopsychological kinds: one that requires of participants that they have common knowledge of each others intentions, the other than requires another kind of "openness" that does not require of participants that they have the concept of belief.[33]

I will also look at two philosophical accounts of what common knowledge is. There are two reasons for doing this. First, to defend the CK-condition in an account of a particularly rich kind of genuine joint action, I need to establish that common knowledge is in fact achievable by finite beings like adult humans. Secondly, to establish that the joint activities of agents such as young children raise is a real dilemma, I have to look closer at accounts of common knowledge to conclude that common knowledge really requires agents to have the concept of belief. Hence, in the second part of the chapter, I present and critically discuss Schiffer's and Lewis's accounts of common knowledge. I argue that neither of these have the resources to make sense of the possibility of "openness" among agents who lack the concept of belief in a satisfactory way. I then go on, in the third part, to present some preliminary ideas regarding what openness (without common knowledge) might consist in. But first, I will explain why common knowledge is required for a kind of particularly rich kind of joint activity.

2.1 The need for openness

[33] Note that focusing on openness and common knowledge is not the only way in which one can show that there are many kinds of joint activity, some conceptually or cognitively more sophisticated than others. In chapter 3, for example, I argue that not all joint activities involve intentions or goals with contents "that we *J*". Instead they involve intentions or goals that, while they are not self-regarding, are not about the activity of one's group either (see page 19).

In the next two subsections, I present what I believe to be the two most convincing reasons for thinking that common knowledge is a defining feature of an intuitive notion of a particularly rich kind of joint activity. First, the CK-condition plays an important role in ruling out problematic cases that we, without the condition, would have to classify as instances of genuinely joint activity. This motivation for the CK-condition underlies the frequent use of metaphors of openness, publicity and transparency in association with the introduction of the CK-condition. Secondly, the CK-condition seems to be required to capture the fact that when agents act jointly, their coordination is non-accidental. With the advent of common knowledge of each other's goals or intentions, the coordination is supposed to become something that the participants achieve together.

While I believe these two motivations for the CK-condition are the decisive ones, another reason behind the condition is to account for the role that shared intention plays in linguistic communication and joint planning. Given my focus on the joint activities that are not planned and that are carried out by relatively unsophisticated cognitive agents, such considerations in favour of the CK-condition is not so relevant for my purposes.[34] Hence, I will not discuss them here.

[34] In his later work, Bratman takes the functional role of shared intention to provide the primary motivation for introducing a CK-condition (Bratman, 2009a, 2009b): "Since […] shared planning is part of the normal functioning of the shared intention, we need an element in our construction of shared intention whose functioning involves such thoughts and knowledge of each about our shared intention." (2009a, p. 160) Shared planning requires communication, and, according to many accounts of communication, this requires common knowledge of background information relevant for interpreting utterances, as well as of the intentions of speakers (Schiffer, 1972b; Clark, 1996; Sperber & Wilson, 1986). If these accounts are correct, then common knowledge will indeed be required for an account that takes shared planning to be part of the functional role of shared intention. The shift in Bratman's motivation for the CK-condition between his earlier and later work on shared intention is accompanied by a slight change in *what* is supposed be common knowledge among participants according to the condition. In his later work, the CK-condition requires that "there is common knowledge among the participants of the conditions cited in the construction", where the 'construction' (the account) includes the CK-condition itself (2009b; 2009a, p.160). Now, on Bratman's view, that there is common knowledge among participants that *p* implies that each participant actually believes that *p* (see 2009a, p 160). This means that for the CK-condition in his later account to be met, each participant must believe that "there is common knowledge

Note that the CK-condition is typically a condition on what must be the case for agents to have a *shared intention*, but since a joint activity is the outcome of a shared intention on most accounts, this condition must also be met for a multi-agent activity to count as a genuine joint activity.

2.1.1 Ruling out concealment cases

Consider the following case:

Hector and Celia are two 7-year-old siblings. In light of their age, they are unusually sophisticated mindreaders. They are also quite unusual (again, in light of their age) in that they both still find the activity of building block towers very enjoyable and not at all silly. Now, they are both sitting on the floor at home, each in possession of a small bag with wooden blocks. Each has come to intend "that we build a block tower together". Furthermore, each intends this in accordance with and because of the other's intention that they build a block tower together. Each also intends that their subplans for bringing about their common goal mesh. In other words, each at least represents (in the content of their own intentions) the other's intention that they build a block tower together, as well as the fact that that the other has subplans for bringing the intended goal about. These attitudes of Hector and Celia cause them (in the appropriate way) to perform actions that together lead to the building of a block tower.

It seems intuitively plausible to say that this building of the block tower by Hector and Celia is a genuine joint activity. Hector and Celia do not merely each intend to

---

among the participants that [the conditions in Bratman's account are met]". On Bratman's earlier account of shared intention, the CK-condition only required that the participants' intentions "that we *J*" and the interdependence of these intentions were common knowledge among the participants. It was not explicitly required that the fact that the CK-condition was satisfied was itself an item of common knowledge (this is the more typical form of the CK-condition, see Alonso, 2009, p. 459; Pettit & Schweikard, 2006, p. 24).

perform a part of something that accidentally makes up a joint activity in some loose sense: The case is clearly different from that of two individuals who merely perform individual actions in parallel. From a theoretical perspective, the case also goes some way toward satisfying Bratman's account of shared intention (see section 1.2.1). Conditions (1) and (2) of Bratman's account would be satisfied.[35] As described, it is not clear whether the CK-condition is also satisfied. However, our reactions to various elaborations of this case will reveal that in taking the case to be an intuitive example of a genuine joint activity, we take for granted that the CK-condition is satisfied too. (In the rest of this chapter, I will frame the discussion in terms of Bratman's account, but note I am not here voicing what I think Bratman would say about common knowledge, but I am rather offering my own view of the topic. Bratman would not, I think, make the appeal to an everyday notion of joint action that I make here.)

But consider whether the case is still an example of a genuinely joint activity in light of the following additional information:

Hector has three yellow blocks and Celia has three red blocks. Earlier in the day, Celia told Hector that their parents would get very angry if they were to see the top face of any red block on the floor. However, Hector did not believe her, and he now thinks that Celia was trying to trick him into believing that their parents had an oddly negative attitude toward red blocks (let us assume that this is what she was trying to do, although this is not yet important). However, Celia has now completely forgotten that she told Hector this lie about their parents. Now, each still intends "that we build a block tower together", and each does this in accordance with and because of the other's intention. Each also intends that their subplans for bringing about that they build a block tower mesh. Because of what Celia told him earlier

[35] We could also gloss the case without appeal to plan-intentions, and describe it merely in terms of beliefs and desires that cause goal-directed actions. In such terms, the activity of each is directed toward the goal "that we build a block tower together". The actions of each of them is sensitive to the other's goal, so each acts (or will act) on their goal because they believe that the other's activity is directed toward that same goal.

about their parents, Hector now believes (erroneously) that Celia believes that he intends "that the top face of any red block is covered" (rather than that he intends "that we build a block tower together"). These attitudes cause them (in the appropriate way) to take turns putting blocks on top of each other: let us say that Celia starts the process by putting down a red block on the floor and combined result of their actions is the building of a block tower.

Now, from Hector's point of view, what is happening is fully consistent with his mistaken belief. Note that if Celia indeed would have put down the red blocks with the intention "that we build a block tower together" not in accordance with Hector's intention that they do the same, but because she thought that she could exploit his attempt to avoid the wrath of their parents, then their turn-taking would lead to exactly same behavioural interaction as that which their actual intentions and beliefs results in. Furthermore, everything I said about the case before adding the information about Hector's false belief about Celia's belief about his own intention still holds true. But does it remain an example of a genuinely joint activity? I think not. It does not seem to be the paradigm case of joint action that it was before the introduction of Hector's false belief. Arguably, it ought to be the case that each of Hector and Celia intends "that we build a block tower together" not only because the other intends the same, but also because they believe that the other's intention is reciprocally sensitive to their own intention. As the case has just been described, Hector believes that the fact that he intends that they build a block tower together is not a consideration that in any way motivates Celia's intention that they build a block tower together. From Hector's point of view, Celia is exploiting the fact that he behaves in a certain way, in order to achieve her own personal goal. Hector will then, in turn, see himself as exploiting this alleged fact about Celia in order to achieve his own personal goal. From his mistaken perspective, Celia intends that they build a block tower together not only if he intends the same, but also if he intends that the top face of any red block is covered or anything else that would lead him to stack his blocks on top of hers. In other words, she is not responsive in the appropriate way to his intention. As long as what he intends leads to behaviour that complements what she is doing in a way that helps her reach her goal, she will maintain and act on her

intention. Now, Hector is using this alleged fact about her, that her intention is responsive to his behaviour in this way, to successfully realise his own intention that they build a block tower together. From Hector's point of view, they are each, one might say, merely using the other as a tool. This seems to jar with what constitutes "genuine jointness". To make the point more vividly, suppose that Hector thinks that Celia is a teaser who is determined *not* to do whatever he intends to do. Accordingly, Hector falsely believes that, if Celia recognised his intention that they build a block tower together then she would not intend the same. But since he thinks that she is mistaken about what his intention is, he can go on and act on his intention that they build a block tower together. Here, he would perform his part of the building of the block tower thinking that he was exploiting Celia's mistaken perspective on what is supposedly guiding his behaviour. Note that if this happened, conditions (1) and (2) could still be satisfied. There ought to be no doubt (or higher-order doubt) about the fact that no one is trying to exploit or deceive in the way illustrated by the case.

To rule this kind of case out, let us add a belief-condition to Bratman's first two conditions. In order for the *J*-ing of two agents A and B to be genuinely joint, the *J*-ing must be the outcome of the following pattern of attitudes among A and B:

(1) (a) A intends "that we *J*" and (b) B intends "that we *J*".
(2) A intends "that we *J*" in accordance with and because of (1a), (1b), and meshing subplans of (1a) and (1b); B intends the same.
(3) (a) A believes (1) and (2) and (b) B believes (1) and (2).

The third condition takes care of rules out the case where Hector has the false belief about Celia's belief about his intention. If (3) is satisfied, then Hector believes that Celia intends that they build a block tower together *in accordance with his own intention that they build a block tower together*. If he believes this then he cannot also believe that Celia believes that his intention is "that the top surface of any red block is covered".

Unfortunately, these conditions do not appear to be sufficient for ruling out cases that are similarly problematic. We can construct a case where all of (1), (2) and (3) are satisfied, but which we arguably would not want to classify as a case of a robust kind of genuine joint activity. Consider the following additional modification of the case with Hector and Celia:

Celia remembers lying to Hector about their parents' unfavourable attitude toward uncovered top faces of red blocks. However, Celia's blocks are in fact green rather than red. In light of this, Hector no longer has any reason to think that Celia is under the impression that he now intends "that every red block is covered with a yellow block" (rather than "that we build a block tower together). So each correctly believes that the other intends "that we build the block tower together" in accordance with and because of their own intention "that we build a block tower together" (hence, the above conditions (1), (2) and (3) are fulfilled). Now, Hector has recently been amusing himself by pretending to be colour blind while hanging out with Celia. In particular, he has recently been consistently calling green objects 'red'. Celia knows that there is nothing wrong with Hector's colour vision or grasp of the colour vocabulary, but she (falsely) believes that his pretence was a genuine attempt to mislead her. Furthermore, she thinks that he thinks that she has in fact been misled. In other words, Celia falsely believes that Hector thinks that she thinks that he is colour blind and under the impression that her blocks are red rather than green. As a result, Celia now believes that Hector thinks that she intends "that we build a block tower together" not in accordance with his similar intention but in light of his putative intention "that every red block is covered with a yellow box". But in fact, it has not even occurred to Hector that Celia might think that he thinks that she thinks that he is colour blind. He never intended to deceive Celia regarding his colour vision or his ordinary use of the terms 'red' and 'green'.

Now, we could add the following condition to our Bratmanian account of shared intention:

(4) (a) A belives (3) and (b) B believes (3).

Such a condition would rule out the modified case just discussed since according to (4), for block-tower building to be a genuinely joint activity, it must be the outcome of a pattern of attitudes that includes Celia's belief that Hector believes that she intends that they build a block tower together in accordance with his intention that they build a block tower together. However, it would now be possible to construct even more complicated cases that still would not strike us as being instances of genuinely joint activity. Intuitively, in a genuine joint activity, the attitudes that drive the actions of the individuals are, metaphorically speaking "above board", "completely out in the open" etc. What is problematic with the information that we have been adding to the original case with Hector and Celia is that it introduces an element of concealment, brought about by deception, manipulation, misunderstanding, or by some higher-order suspicion that such an element is present at a lower level.

Since the CK-condition is typically introduced into accounts of shared intention and joint action without explicit rationale, it is hard to say whether ruling out concealment cases accurately reflects what the condition is supposed to achieve according to various authors. As far as I know, Pettit and Schweikard (2006) are the only authors who explicitly motivate the CK-condition with an appeal to a concealment case. They consider the case where two people, you and I, have intentions "that we *J*" that are interdependent in the right way, but where common knowledge about this fails to obtain because, "for example, you might regard me, wrongly, as someone who takes you to be acting like a zombie, as if under hypnotic suggestion." (p. 23) In light of this (briefly described!) example, they suggest that "we need to introduce a […] clause that serves to silence it and any other possibility of the same kind. The clause stipulates that everything amongst the parties is above board." (Pettit & Schweikard, 2006, p. 23) The thought is presumably that if one is acting under hypnosis, then one's behaviour isn't guided by one's beliefs and goals in the appropriate way, and this undermines a type of mutual responsiveness that is characteristic of genuine joint action. If I believe that you believe that I am under hypnotic suggestion, then I believe that you are merely exploiting the fact that I

behave in a certain way in order to achieve your own goal, and I will thus, in turn, see myself as exploiting this alleged fact about you in order to achieve my own goal. Pettit and Schweikard appeals, like I do, to an everyday intuitive notion of a kind joint action, a kind that they characterise as "properly joint" and that involves "unforced cooperation" (2006, pp. 20, 22–24).

I believe that ruling out such cases is one decisive reason for having a CK-condition in an account of shared intention. As I have pointed out, this kind of motivation for a CK-condition parallels that which has led some philosophers of language of introduce a CK-condition in Gricean accounts of communication (Clark & Marshall, 1981; Schiffer, 1972, pp. 17–30). It might be objected that while there is a robust intuitive notion of genuine communication, there is no such corresponding intuitive notion of joint action. Hence, intuitions telling us that some case or other is not really a case of genuine joint action must be taken with a grain of salt. In a sense, I think this is right. One of my aims of this chapter is, after all, to argue that there are many notions of genuinely joint activity. But I want to insist that there is something missing from the concealment cases that make them less genuinely joint than cases were the relevant beliefs and intentions or goals of the participants are, as it were, "completely out in the open". Rather than thinking of multi-agent activities as simply either being properly joint or else not being joint at all, it is better to think of a continuum of increasingly joint cases of multi-agent activity. What I am claiming here is that the satisfaction of a CK-condition constitutes a step up toward increasing jointness on such a continuum.

Note that the modifications that introduce concealment into the block tower building case are only possible if Hector and Celia are mindreaders with sophisticated metarepresentational abilities. If they were slightly younger, and thus lacking the concept of belief, then these problematic cases could not arise.[36] So what is arguably

[36] A very robust finding in developmental psychology is that children fail at so-called elicited-response false belief tasks before 3 to 5 years of age (Wellman, Cross, & Watson, 2000). This suggests that younger children lack the concept of belief and do not interpret the behaviours of others in terms of a robust "Theory of Mind" (ToM). This interpretation of the false belief task data is not uncontroversial, but I will in this

the main motivation for introducing the CK-condition in accounts of joint action is actually undermined when the agents involved lack the concept of belief.[37] But this does not mean that the CK-condition is not needed when the participants *do* have the concept of belief.

If ruling out concealment was the only valid motivation for introducing the CK-condition, then we could now drop the condition in the context of an account of joint activity with participants who lack the concept of belief. This raises the question of whether there is a kind of joint action that agents who entirely lack concepts of the mental or intentional (including the concept of goal) could engage in. Of course, Bratman's account, even without the CK-condition, could not be an account of that kind of joint action, since condition (2) demands of participants that they have intentions concerning both their own intentions and the intentions of others.[38] For such creatures, should we not even require that each has a first-order belief about the other's goal for there to be genuine jointness? Consider again the case Hector and Celia building a block tower, but suppose they both have severe forms of autism. Neither of them, let us suppose, have any mental state concepts at all. Nevertheless, they are both skilled at predicting what the other will do in terms of behavioural regularities. In particular, they know that whenever they put down a block of their own colour the other will put down a block of the other colour on top of that block. Each has the goal to build a block tower, and on the basis of the knowledge of the regularity in the other's behaviour, they can each exploit the other's contribution to achieve their personal goal of building a block tower. In this case, each agent's

---

chapter assume that it is correct.

[37] Presumably, this also undermines the main motivation for insisting on a CK-condition in accounts of communication when applied to the communicative utterances of young children. But since communicative intentions, on the Gricean view, are intentions to induce *a belief* in an audience, the fact that children who arguably lack the concept of belief appear to make communicative utterances is still puzzling for Gricean accounts of communication.

[38] Note that there are good reasons to doubt that young children who lack a robust Theory of Mind are able to satisfy Bratman's second condition even if they have some metarepresentational capacities (Butterfill, 2012; Tollefsen, 2005). I set this issue aside here in order to focus on common knowledge and openness.

expectation about the other's response to their own block placement will be correct in virtue of the fact that they have a common goal, but neither of them represents the other's goal.

Perhaps such "mindblind" coordination characterises the multi-agent activities of some non-human animals, such as the pack hunting of wolves or hyenas for example. The individuals have a common goal and thanks to their ability to predict the behaviour of the other individuals in the pack, they can adapt their own goal pursuit to the others' behaviours in way that maximises their chance of achieving their goal. (I elaborate on what it means for several individuals to have a common goal in the next chapter). Again, we have a case where each agent seems to be exploiting the other to reach his or her own goal. While there isn't an element of "concealment" present in this case, there is a lack of transparency. Each participant is blind to the goal of the other. I will call such cases "opacity cases". Now, I cannot think of any good principled reason for not calling such a multi-agent activity "joint" in a weak sense of the term. This would constitute yet another kind of joint activity, which is even less rich than that which applies to most joint activities involving young children for example, who do seem to have the concept of goal if not belief.

As it happens, ruling out concealment and opacity cases is not the only role that common knowledge is supposed to play in accounts of joint action. There is another important motivation for the introduction of a CK-condition.

2.1.2 Making coordination non-accidental

When two or more agents are faced with a "coordination problem", the outcomes of the action alternatives open to each agent depend on which actions the others settle on. Here is a mundane case: Say that you and I want to meet up at noon. It does not matter for either of us where we meet, but in order to reach our goal of meeting each other, both must go to the same location (wherever that is) at noon. The best choice of where to go for each is determined by the other's choice of where to go. So, in order to successfully and reliably meet in the same place, each of us needs to have a

justified expectation about where the other is about to go. But how is this possible given that our expectations about each other's destination are interdependent? Where I ought to go depends on where you ought to go, which in turn depends on where I ought to go, which again, in turn, depends on where you ought to go, and so on and so forth.

Clearly, we can "solve" this kind of coordination problem by, for example, explicitly agreeing beforehand about where to meet. Now consider what kind of knowledge about each other that we must have in order to pull off this coordination if we have explicitly agreed to meet in Bristo Square at noon. It is not sufficient that both of us believe that the location for our meeting is Bristo Square. After all, if I do not also believe that you believe that the location for our meeting is Bristo Square, then it is not at all clear that I will meet you if I go to Bristo Square. This is because my only reason for going to Bristo Square rather than some other location is that you believe this is the meeting point. So if I go to Bristo Square lacking the belief about your belief about the location of our meeting, it will in a sense be a matter of luck if we happen to meet each other there. The same could of course be said of you. Unfortunately, the situation is not improved if each of us not only believes that the location for our meeting is Bristo Square but also believes that the other believes this. If I do not also believe that you believe that I believe the location of our meeting is Bristo Square, then it is not at all clear that I will meet you if I go to Bristo Square. After all, we have just established that if you do not believe that I believe the location for our meeting is Bristo Square, then it would be a matter of luck (from your point of view) if you happen to meet me in case you go to Bristo Square. So, if I do not believe that this is something you believe, then I have no reason to expect that you go to Bristo Square rather than to some other location. Again, it seems that if I go to Bristo Square and meet you there, I ought to be pleasantly surprised. It would be appropriate for me to exclaim, "Oh, you decided to go here too! That's great!" But if we had explicitly agreed to meet at Bristo Square, as stipulated, then this would not be an appropriate thing to say. But my (third-order) belief about your belief about my belief would not stop the exclamation from being inappropriate either. In fact, no finite level of higher-order belief would be sufficient to make sense

of the intuition that this exclamation would be highly inappropriate! What is needed to make sense of the intuition is the following: When we explicitly agree to meet at Bristo Square, the fact that the location of our meeting place is Bristo Square becomes "above board", "completely out in the open", or common knowledge between us. This suggests a desideratum that any account of common knowledge should meet. As Gilbert points out,

> in seeking an account of common knowledge one is in part seeking an account of a phenomenon such that, for a given proposition p, acts which are premised on the belief or assumption that p will surprise no one if this phenomenon is present. (Gilbert, 1989, p. 194)

In the context of joint action, when two agents have a shared intention to do something their experience of the other's actions that are performed in pursuit of the shared intention will be characterised by a lack of surprise. If we cut off the regress of higher-order beliefs at *any* finite level, all the levels below will be undermined, including the first-order belief that the location of our meeting is Bristo Square. Hence, if we go there and encounter each other, surprise will be an appropriate response.

The claim here is merely that the occurrence of surprise indicates the absence of common knowledge. I am not suggesting that lack of surprise itself is a reliable mark of common knowledge. Actions premised on the belief or assumption that *p* may not surprise for other reasons than because it is common knowledge that *p* (and even if it is not common knowledge that *p*). For example, each of us might receive an email from a third party informing us that if we go to Bristo Square at noon, the other will be there too. Given that we trust the testimony of this third party, each of us might arrive at Bristo Square to meet the other as expected. Here, our first-order expectation is the result of a common cause, rather than the result of an agreement that is the basis for common knowledge.

Some kind of "openness" (in the sense that things are entirely expected, rather than that they are not hidden or concealed) seems to be required for coordination of action to be non-accidental, and arguably, it is part of our intuitive notion of joint action that it is non-accidentally brought about. In support of this, we can draw on an

analogy with complex individual activities. The moves a climber makes as he scales a rock face are non-accidentally coordinated with each other to combine in such a way that they bring the climber to the top. If they are not coordinated in this way, then we have reason to question whether it was his agency that brought him to the top rather than mere luck (there just happened to be holds in the rock face in the right places for him get up there). In order for two agents to act jointly, their actions must be non-accidentally coordinated in a similar manner.

The problem that the CK-condition is supposed to solve is *not* that lack of common knowledge undermines the participants' confidence in successfully bringing about a collective outcome. This may be an unwelcome consequence of not having common knowledge though. Lack of common knowledge may either undermine a participant's ability to form the intentions mentioned in Bratman's conditions (1) and (2) (assuming a strong belief condition on intention) or it may make it imprudent for her to act on these intentions. Consider what happens if it would be *much* more inconvenient for me if I went to Bristo Square but failed to meet up with you than if I gave up trying to coordinate and instead stayed at home. Assume that the same is true of you.[39] If the location of our meeting were not common knowledge between us in this case, then it wouldn't be prudentially rational to go to Bristo Square for either of us. In addition, if we assume a strong belief condition on intention, then neither of us could even form an intention "that we meet at Bristo square". Hence, lack of common knowledge could thwart our attempts to act on our intentions, as well as stop us from meeting other conditions that must be met for us to share an intention. But if we assume that our problem of deciding whether or not to go to Bristo Square

[39] A more dramatic but analogous case is that of the Coordinated Attack Problem, also known as the Generals' Paradox (Campbell, 2005; Gray, 1978, pp. 465–466; Rubinstein, 1989; Wilby, 2010). In their discussion of the role of higher-order beliefs and common knowledge in joint action, Chant and Ernst (2008) use the Coordinated Attack Problem to make the point that common knowledge is sometimes required for shared intentions to be formed and joint action come about. What they want to analyse is the "role of interactive [higher-order] knowledge in *generating* collective intentions." (p. 555, my emphasis) I am concerned with a slightly different question, namely why common knowledge should be taken to be a constitutive element of a particularly strong notion of joint action.

at noon has a different pay-off structure, or if we assume a less strict belief condition on intention, then this need not be the case.[40] Thus, these considerations do not cut to the heart of why common knowledge of each other's intentions and beliefs is a necessary component in an account of an intuitive notion of genuinely joint activity.

Note that this motivation for the CK-condition is not undermined if the agents lack the concept of belief. While the problem with the case I presented here – you and I coordinating to meet at Bristo Square at noon by having merely finite levels of higher-order beliefs – assumed that we possessed the concept of belief, the intuition that the coordination ought to be non-accidental in genuinely joint action does not itself depend on agents having the concept of belief. After all, coordination problems arise not only when agents' higher-order beliefs have to be coordinated, but also in attempts by two people to perform interdependent bodily actions in a coordinated fashion. Say that Hector and Celia decide to build a platform on their tower. Above the platform, they build a tower-structure each on opposite sides of the platform (so that the tower splits up into two turrets after the platform). In order to avoid having the structure collapse, they have to simultaneously place their blocks on each side of the platform, or it will fall over. Whether t_1 (rather than t_2) is the right time for Hector to release his grip on his block to let it rest on the platform will thus depend on whether t_1 is the right time for Celia to do so, and this will, in turn, depend on whether t_1 is the right time for Hector to do so, and so on back and forth. Note that the coordination problem of getting themselves and the other to both act at t_1 does not depend on the metarepresentational abilities of Hector and Celia, but it is simply imposed by the task that they are facing together (note that my point is that the

[40] Tuomela and Miller's reason for introducing the CK-condition in (1988, pp. 55–57) seem to be that, without common knowledge, participants' confidence in success (in the other doing their part of the joint action) will be undermined, and they will therefore not be able to intend to their own part of the joint action. Ludwig (2007) argues against this claim, and submits that "it is not required that members of a group even believe that others will play their parts in a joint action which they nevertheless intend to do and perform intentionally." (p. 387) Elsewhere, Tuomela (2007, p. 113) seems to argue that a CK-condition is needed because it typically won't be prudentially rational to participate in joint action without it. Note that this is slightly different from arguing that common knowledge is required for having the requisite intentions as a matter of conceptual necessity.

problem does not depend on such abilities, not that *solutions* to the problem do not depend on such abilities).

To sum up, the two main motivations for the CK-condition are the following: (i) to rule out concealment cases that are in tension with an intuitive notion of genuine joint activity; and (ii) to make sense of the idea that coordination in genuine joint activity is achieved non-accidentally by the participants. Motivation (i) is undermined when it comes to participants who lack the concept of belief, but a similar type of case, opacity cases, reveals that there is an intuitively important difference between, on the one hand, joint activities where coordination is enabled by first-order beliefs about the goals of the other participants, and, on the other hand, joint activities where coordination is enabled by mere "smart behaviour reading". In contrast, (ii) remains important regardless of the metarepresentational capacities that the participants are equipped with. While it is clear that something like the infinite regress implied by common knowledge is required by (i), common knowledge may be just one of many ways in which the openness required by (ii) can be provided.

I have so far argued that both reasons for introducing a CK-condition in accounts of joint action with cognitive normal mindreading adults are reasons that should be taken seriously. But if the CK-condition is in general too cognitively demanding, or even completely psychologically implausible, then we might have to revise our intuitive notion of what is involved such genuine joint action. In the next section, I present and critically discuss how Schiffer (1972) and Lewis (1969) have tried to make the infinite regress associated with common knowledge compatible with a commitment to psychological realism. In the case of Lewis' account, I also consider whether common knowledge really requires that agents have the concept of belief. If it doesn't, then perhaps it can both help us rule out concealment cases and make sense of a kind of openness that agents lacking the concept of belief can participate in. However, I conclude that if Lewis' account is to be at all informative, then it will require that agents have the concept of belief.

2.2 Common knowledge as a virtual infinite regress

We have already encountered the iterative definition of common knowledge: A proposition *p* is common knowledge between the agents in a group if and only if

(everyone knows that)$^n p$

for *all n*. If this was true of Hector and Celia for example, then the concealment cases that we came up with would indeed be ruled out. In addition, the openness required for non-accidental coordination to occur would be in place. But how could common knowledge construed in this way play a role in the real-time decision-making of agents with mere bounded rationality and finite minds? The definition seems to demand from agents that they have beliefs with infinite inexhaustible content ("I believe that you believe that I believe…*ad infinitum*). Furthermore, to ascertain that no concealment is present, or that coordination is indeed non-accidental in the requisite way, the agents must presumably make an infinite number of inferences to an infinite number of beliefs (unless they could somehow all appear simultaneously and non-inferentially?). Thus, the iterative definition seems to make common knowledge into a mere ideal state of affairs that will never realised in the actual world where joint action happens. So by introducing a CK-condition to rule out problematic concealment cases and capture the sense in which coordination in join action is non-accidental, we face the problem of articulating a psychologically realistic account of common knowledge.

I will here explore two ways in which the infinite regress of the iterative definition can be made harmless (in Schiffer's account) or avoided entirely (in Lewis' account).

2.2.1 Virtualising the infinite regress with dispositional beliefs (Schiffer)

The iterative definition becomes outrageously unrealistic only if we assume that each proposition in the definiens must be the content of an occurrent belief that each agent actively considers. This would obviously make common knowledge unattainable given the iterative definition of common knowledge. If one also thinks that an

occurrent belief is something like a tokened sentence in a language of thought, then the notion of occurrent beliefs of an infinitely high order becomes absurd. But those who subscribe to the iterative definition of common knowledge clearly do not embrace these assumptions. According to Schiffer (1972) – as far as I know the first to propose this definition[41] – the beliefs (or knowledge-states) referred to in the definiens may merely be dispositional beliefs (or "standing beliefs" as they are also called). Consider Schiffer's following defence of his iterative account in response to two potential objections:[42]

> First, it is no objection to the claim that S knows that p that the thought that p never once entered S's head. For example, I trust that it is true of each philosophy don in Oxford that he knows that his maternal grandmother was never married to Benito Mussolini. Second, it is no objection to the claim that S knows A knows S knows A knows S knows that p that S may have to be "convinced" or "brought to see" that he is entitled to claim to know this. I doubt that many a non-philosopher would agree right off the bat that he knows that he knows that he knows that 843 + 2 = 845. (Schiffer, 1972, p. 36)

This makes the infinite regress into a merely "virtual" one. There is no obvious reason why the number of dispositional beliefs should be finite, nor why the contents of the beliefs couldn't be explicated without appeal to the notion of infinity. Now, I am not saying that Schiffer's response to the two objections he considers do not lead to some complications, because they do. Presumably, the point of the last sentence in the quote above is that while the non-philosopher wouldn't agree "off the bat" that he knows that he knows that he knows that 843 + 2 = 845, the non-philosopher would nevertheless know this since he could be brought to see that he should accept it. Even if Schiffer is correct about the general point that one believes many things which has never crossed one's mind and which one would have to work hard to infer, it seems

---

[41] The iterative definition is very often erroneously attributed to Lewis's earlier work *Convention* from 1969 (see e.g. Campbell, 2002, p. 160; Kutz, 2000, p. 6 n. 8; Skyrms, 2009, pp. 137–138). However, the iterative definition is nowhere to be found in (Lewis 1969).

[42] Note that Schiffer's account also includes a characterisation of how the infinite regress of higher-order beliefs can follow from a finite basis (similar to Lewis' basis). But Schiffer, unlike Lewis, identifies what it is to have common knowledge with having the beliefs in the infinite regress rather with being in a situation that provides the finite basis. See Wilby (2010) for a lucid discussion of this aspect of Schiffer's account (see also Peacocke, 2005).

implausible to say that every proposition that one has reason to believe is the content of a dispositional belief that one has. This would imply that one knows all truths about mathematics if one accepts certain mathematical axioms and principles. This is at least in conflict with everyday notions of belief and knowledge. Perhaps this is not something we need to worry about, but there is a more serious problem with relying on dispositional beliefs to virtualise the infinite regress. The problem is that it is hard to see how common knowledge could make a difference for experience and action if the infinite regress is merely one of dispositional beliefs.[43]

Note that it isn't required that the agents themselves grasp that they are in a situation of common knowledge for the purpose of ruling out concealment cases. That is not something that the agents themselves are concerned with. The CK-condition rules out concealment cases for us analysts who are trying to account for what needs to be the case for them, the agents, to be involved in a genuine joint action. However, it is different when it comes to the second motivation for the CK-condition. It looks like participants need to have an occurrent grasp of the fact that they have common knowledge (rather than some finite level of higher-order belief) in order to be able to together coordinate a multi-agent activity in a non-accidental manner. If you and I cannot distinguish between the case of having common knowledge that the location for our meeting is Bristo Square on the one hand, and the case of merely having finite higher-order beliefs about the location of our meeting, then neither of us would be able to make a reliable judgement about whether it would make sense for me to go to Bristo Square at noon or not. I would then not be able to judge whether or not I would be surprised in case I went to Bristo Square and met you there.

So, on the one hand, it looks like the state of common knowledge must be graspable somehow, in order to make an impact of the experience and judgements of the

[43] The higher-order dispositional beliefs do reflect a kind of awareness of their contents only in the following sense: If you and I have common knowledge that *p*, then *if* someone would query either of us about whether you believe that *p*, whether I believe that you believe that I believe that *p*, or whether you believe that I believe *p* (for example), then we would both answer affirmatively (see Pettit & Schweikard, 2006, pp. 23–24).

participants (Gilbert, 1989, p. 193; Heal, 1978). But if common knowledge is identified by the infinite set of higher-order beliefs as in the iterative definition, then such an occurrent grasp seems to be psychologically impossible.

2.2.2 Virtualising the infinite regress with reasons to believe (Lewis)

I will look in more detail at what is arguably the seminal philosophical account of common knowledge: namely the account presented in Lewis' *Convention* (1969). There are good reasons for focusing on this account in particular. Like Schiffer's account, it avoids the most straightforward objections to the effect that common knowledge is unattainable in the real world. Furthermore, it is the account that is most prominent in the philosophical literature on shared intention and joint action. Bratman frequently refers to Lewis' account as an example from "a large literature on common knowledge" (1999, p. 102 n. 15, p. 117 n. 20; see also 1999 p. 111, n. 8, where he refers to Lewis' notion of a 'basis for common knowledge'). When Bratman does say something more substantial about common knowledge, it is always consistent with Lewis' account (see e.g. Bratman 1999, pp.102, 139, 2009a, p.160). Pettit and Schweikard (2006, pp. 23-24) straightforwardly refer to Lewis (1969) as the source of their notion of common knowledge. Gilbert (1990) notes that "[e]xactly how to define [common knowledge] is somewhat moot" (p. 13 n. 4) and then refers to Lewis (1969), Schiffer (1972), Heal (1978), and her own account (Gilbert, 1989, pp. 186–197). In light of its prominent position, an interesting question is whether Lewis' account can provide what is requested from it within the context of theories of shared intentionality and agency.

Like Schiffer, Lewis sidesteps the objections concerning psychological realism by virtualising the infinite regress that is associated with the notion of common knowledge. But Lewis does this by putting the regress entirely outside the heads of the individuals involved. The infinite regress is not even a regress of dispositional beliefs, but a regress of *reasons to believe*.[44] Furthermore, unlike Schiffer, Lewis

[44] As Vanderschraaf & Sillari (2009) notes, "it would be more appropriate to speak

doesn't *define* common knowledge in terms of this regress, but in terms of a state of affairs called a *basis for common knowledge*, from which the infinite regress can be derived. For example, when we agree to meet at Bristo Square at noon, our making the agreement constitutes a basis for common knowledge about where we will meet at noon. And on this basis, we can each expect the other to be at Bristo Square at noon, and if we set cognitive limitations aside for a moment, each of us would on this basis also be *justified* in having higher-order expectations about these expectations of a potentially infinitely high order.[45] A basis for common knowledge need not be an explicit agreement of course, it could also be shared habit or practice (we always meet at Bristo Square), or an environment that we both inhabit and perceive and both perceive that the other perceives (if we sit facing each other at the floor with half-built block tower between us, then this situation constitutes a basis for our common knowledge that there is a half-built block tower between us).[46]

Lewis characterises the conditions under which common knowledge obtains in the following way:

> Let us say that it is common knowledge in a population P that [q] if and only if some state of affairs A holds such that:
>
> (1) Everyone in P has reason to believe that A holds.
>
> (2) A indicates to everyone in P that everyone in P has reason to believe that A holds.

---

of 'common reason to believe'" instead of 'common knowledge' as the target phenomenon of Lewis' account.

[45] Lewis interchangeably refers to higher-order 'expectations' and higher-order 'beliefs' (see p. 55). I therefore assume that he takes an expectation to simply be a belief about some future event. Note that there is nothing in Lewis's account that necessarily restricts the content of common knowledge to concern the future though.

[46] Bratman gives the following example of a case where two agents get common knowledge of each other's intentions and beliefs in the way required for them to have a shared intention, without recourse to any explicit communication: Two strangers arrive at a public basketball court and starts to take turns shooting even if they do not explicitly agree on taking turns shooting (Bratman, 1999, p. 139). Here, the idea is that the social setting makes it common knowledge that each intends that they do this, and that each intends to do it by way of the other's intention to do the same.

(3) A indicates to everyone in P that [q].

We can call any such state of affairs A a basis for common knowledge in P that [q]. (Lewis, 1969, p. 56)

He explains what it is for a state of affairs to indicate something to someone in the following way:

Let us say that A indicates to someone x that [q] if and only if, if x had reason to believe that A held, x would thereby have reason to believe that [q]. What A indicates to x will depend, therefore, on x's inductive standards and background information. (Lewis, 1969, p. 53)

The account builds on the idea that the potentially infinite chain of higher-order expectations can be generated from a basis that is finite. Since we agree to meet at Bristo Square, we both have reason to believe that the agreement has been made. After all, we speak the same language and each understands what the other says. Our exchange of speech acts also indicates to each of us that both of us have reason to believe that the agreement has been made. In addition, the state of affairs that we are part of indicates to both of us that we will meet at Bristo Square at noon.

With a basis for common knowledge, something very much like the infinite list of justified higher-order beliefs in the iterative definition follows:

I have reason to believe that [q] [from (1) and (3)]
You have reason to believe that [q] [from (1) and (3)
I have reason to believe that you have reason to believe that [q] [from (1), (2) and (3)]
You have reason to believe that I have reason to believe that [q] [from (1), (2) and (3)] and so on into infinity [from (1), (2) and (3)]

While Lewis distinguishes 'reason to believe' from 'actual belief', he does not say much about what having a reason to believe involves. That an agent has a reason to believe that p at least does not imply that the agent actually believes that p. Lewis' "actual beliefs" seem to include all one's occurrent beliefs, but perhaps they also

include some subset of his dispositional beliefs. He talks about the actual (higher-order) beliefs as being "formed" or "generated" (see e.g. p. 56) from a basis of common knowledge, and their formation is the result of "actual reasoning" (p. 55). This surely sounds like he is talking about the formation of occurrent or consciously entertained beliefs. But at same time, it seems like Lewis takes what one believes (simpliciter) to include more than what one occurrently entertains or consciously endorses. Otherwise it is hard to make sense of his claim that "[a]nyone who has reason to believe something will come to believe it, provided he has a sufficient degree of rationality." (p. 55)[47]

Lewis' account has the advantage that the infinite number of propositions about beliefs that is contained in the definiens of the iterative definition, each of which one has reason to believe is true, need not be part of what one believes or knows. Hence, Lewis's account is not open to any simple charges of lack of psychological realism. Participants need not actually *make* any inferences to any justified or true higher-order beliefs (or even first-order occurrent beliefs), they only need to be able to have higher-order reasons to believe (as Paternotte, 2010, p. 260 emphasises).[48] So Kutz's charge that (Lewisian) common knowledge is a "cognitively demanding state" is wrong, at least if he meant that it is too demanding in terms of cognitive load or processing (2000, p. 6 n. 8). Neither does it seem to be the case that a demand for common knowledge is so "difficult to satisfy" as Pacherie takes it to be (2007, p.

[47] Perhaps Lewis uses the term 'rationality' to refer not only to the agent's logic of reasoning, but also to his degree of cognitive control, perceptual capacities, attention, and so on. In order for an agent to come to have an occurrent belief that *q* when he has reason to believe that *q*, he must presumably direct some degree of attention to the issue of whether *q* is the case or not.

[48] When common knowledge serves to solve coordination problems—when it grounds rational action—then Lewis thinks that, typically at least, the members of *P* do not just have reason to believe that others have reason to believe (etc.) but also have actual higher order beliefs—that is, they *actually* believe propositions—of some complexity. When Lewis comments on a simple case where two people agree to meet again the next day to carry on a conversation, he describes them as having higher-order beliefs of an order between three and five (see p. 52, also p. 32). According to Lewis, the cut-off point beyond which no actual higher-order beliefs are generated is determined by the rationality of the members of *P*, their background information, and their standards of inductive inference.

166). In addition, Lewis is not (unlike Schiffer) pressed into having to claim that we believe, say, all the true propositions of mathematics that follow from axioms that we accept.

Recall that the iterative definition is problematic because it seems to make it impossible for agents to grasp that they have common knowledge. In other words, given this definition it is hard to see how common knowledge could make a difference to an agent's experience or action. How does Lewis's account fare in light of this problem? Since Lewis does not define common knowledge in terms of a set of infinite higher-order beliefs or reasons to believe, it is not obviously the case that agents may not be able to grasp the fact that they have common knowledge that *q* in contrast to merely having finite higher-order belief that *q*. In order for common knowledge of something to be a reason that motivates participants into action, they must have an occurrent grasp of the fact there is common knowledge. This is not impossible in light of Lewis' account.

For a member *x* of population *P* to have an occurrent belief "that there is common knowledge in *P* that *q*"*, x* must be able to grasp that she is in a state of affairs *A* such that (1) everyone in *P* has reason to be believe that *A* holds, that (2) *A* indicates to everyone in *P* that everyone in *P* has reason to believe that *A* holds, and that (3) *A* indicates to everyone in *P* that *q*. There is no obvious need for *x* to infer an infinite number of higher-order beliefs of an infinitely high order to entertain this belief. Admittedly, the content of this belief looks quite complex, but we should not demand that participants must be able to *articulate* the content of a belief that they entertain.

In order to say more about what cognitive and conceptual demands that Lewis' account places on participants, we need to get a better idea of what it is to have a 'reason to believe' something. This also matters when we consider whether Lewis' account might provide an account of openness that is applicable to the joint actions of young children (and perhaps also non-human animals) who lack the concept of belief. Tollefsen (2005) claims that "no matter what account of common knowledge you present, it presupposes that participants have knowledge (either tacit or explicit)

of the mental states of others." (Tollefsen, 2005, p. 82, n. 12, also p. 92) And having such knowledge requires "an understanding of mental representation" (ibid.). But must agents really have metarepresentational capacities, and in particular the concept of belief, to have common knowledge given Lewis' account?

Besides saying that what one has reason to believe somehow depends on one's inductive standards and background information, Lewis does not give us much guidance on this issue. In an exposition and analysis of Lewis's account of common knowledge, Cubitt and Sugden's (2003) provides the following definition of 'reason to believe':

> To say that some individual i has reason to believe some proposition x is to say that x is true within some logic of reasoning that is *endorsed* by (that is, accepted as a normative standards by) person i. For x to be true within such a logic of reasoning, it must *either* be treated as self-evident *or* be derivable from propositions that are treated as self-evident using the inference rules of the logic. (Cubitt & Sugden, 2003, p. 184)

Given this definition, for two or more agents to have common knowledge of *q*, they must each have the concept of belief. 'Ought' implies 'can' here. What an agent has reason to believe, on this reading, is dictated by the agent's evidence and the logic of reasoning *that the agent has endorsed* (whether this logic is also endorsed by us as analysts is irrelevant). It would clearly be against the spirit of this to allow an agent to have a reason to believe a proposition that he cannot *actually* believe. If an agent's reasons to believe are relative to the logic of reasoning he endorses, they should also be relative to the concepts that he can entertain. In order to be common knowledge between two or more agents, *q* must either be treated as self-evident *by the agents* or be derivable *by them* from propositions that *they* treat as self-evident. So, on this reading, for an agent to have a 'reason to believe' something, the reason must be accessible to the agent in a fairly strong sense. The agent must have the conceptual resources to believe what he has reason to believe. In line with this, Glüer and Pagin (2003) claim, regarding Lewis' account of common knowledge, that "if [an agent] is unable to form beliefs about beliefs, then, it seems, no state of affairs *A* will give him

reason to believe, and thus will not *indicate* to him, that everyone in the population *P* has reason to believe that *A* holds." (p. 42 n. 27)[49]

However, it is not obvious that this is the reading Lewis intended, and it is certainly not the only possible reading. There are notions of having a reason to believe *p* according to which I can have a reason to believe *p* even if I lack a concept that I need in order to believe *p*. Perhaps all that is required for me to have a reason to believe *p* is that some evidence for *p* is available to me. I might then have reason to believe, for example, that the animal that I see swimming over there in the lake is a beaver, even if I don't have the concept of a beaver (so I instead believe that it is a 'giant water rat'). Evidence that the animal is a beaver is available to me: I can see the animal's broad tail slapping in the water, and I have noticed that there are gnawed-off tree trunks here and there along the waterfront. For all that, I may not be able to form the demonstrative thought "*that's* a beaver" because I do not possess the concept of a beaver.

By these lights, I could have evidence that you have evidence that my goal is that we build a block tower together even if I cannot believe that you believe that this is my goal. (The evidence here could be, for example, that we are facing each other, that we have a few blocks each, that I make eye contact with you and that I grab and raise a block suggestively.) But given such a weak notion of what it is to have a 'reason to believe', we face the problem of how the agents would be able to grasp that they have common knowledge of their goals, rather then merely second- but not third-order evidence of their goals, as well as the problem of explaining how common knowledge about each other's goals should lead to agents experiencing no surprise when actions are performed in pursuit of these goals. Furthermore, we will get no guidance from the account regarding what the cognitive or conceptual demands are of being able to stand in relations of common knowledge. For the purposes of understanding what openness consists in among agents who lack the concept of

[49] Glüer and Pagin make this claim in the context of a discussion about whether it would be a problem for Lewis' theory of linguistic meaning (of which his account of common knowledge is a part) if speakers without the concept of belief exist.

belief, Lewis' account either ends up construing openness as unachievable (since it requires the concept of belief), or it tells us nothing about *how* openness enables coordination of joint action to be non-accidental. As a consequence, the account does not give us any clues about what cognitive capacities and concepts that agents must possess in order to act on common knowledge.

To sum up this second section, we have seen that there are accounts of common knowledge that avoid some of the problems associated with a CK-condition. The condition is not necessarily too cognitively demanding in general. The infinite regress of higher-order beliefs need not be arrived at through an infinite number of steps in someone's reasoning, and the beliefs need not be actively entertained. Instead, the infinite regress can be "virtualised", either by making the beliefs dispositional, or by making identifying common knowledge with a certain type of situation—a basis for common knowledge—that provides agents with reasons to believe the propositions in the regress. When it comes to the joint activities of mindreading adult humans, both these accounts can be plugged into an account of joint action to successfully rule out concealment cases. However, I have argued that Schiffer's dispositional belief account fails to explain how common knowledge enables coordination to be non-accidental. Lewis' account seems to be more congenial when it comes to this second desideratum, since it avoids making the idea that participants can grasp the fact that they have common knowledge of some proposition into a mysterious phenomenon that requires appreciation of an infinite regress. My main concern here, however, is whether the CK-condition implies that participants must have the concept of belief. While this is not obviously the case given Lewis' account, Schiffer's account clearly implies this. The problem with Lewis' account, on the other hand, is that it is entirely uninformative regarding what conceptual capacities that are required for agent to be able to have common knowledge with others.

In the next section, I consider what "openness" might be if it doesn't require the concept of belief, given that openness is something that can make sense of the idea that coordination in joint action is non-accidental.

2.3 Openness without common knowledge

Young children engage in sustained episodes of joint attentional activity already in their second year of life, long before they seem to acquire the concept of belief. Now, joint attention is typically taken to involve a kind of mutual awareness or openness between the co-attenders concerning their direction and focus of attention (Eilan, Hoerl, McCormack, & Roessler, 2005; Moore & Dunham, 1995; Seemann, 2012). This openness enables young children to coordinate genuine joint *attention* together in a non-accidental way, despite their lack of the concept of belief. Perhaps this kind of openness can play the role that common knowledge normally plays in accounts of joint action. This is essentially what Tollefsen (2005) proposes. She claims that joint attention can replace common knowledge and "introduce the openness that needs to be present in cases of joint action." (p. 92) Recall that the concern to rule out concealment cases never arises in the case of agents who lack the concept of belief – since the element of "higher-order concealment" depends on them possessing the concept of belief – so this kind of openness only needs to account for how coordination can be non-accidental.

There are two desiderata on an account of the kind of openness that we need. First, it must make sense of how coordination of joint *action* can be non-accidental. Secondly, it must not require that participants have the concept of belief. With regard to the second desideratum, we should consider whether openness could occur between agents who have no concepts of the mental or the intentional at all (including the concept of goal).

Hutto (2012) proposes that "basic joint attention" (p. 314) is best understood "in terms of interactive unprincipled embodied engagements" rather than in terms of reciprocal mindreading (p. 329).[50] To be fair, Hutto doesn't discuss openness or

[50] Hutto does not merely think that pre-linguistic children cannot recognise or attribute propositional attitudes to others, he think that they themselves don't even *have* propositional attitudes. They merely have what he calls "intentional attitudes" with "nonsemantic directedness". This claim of his is not relevant for my purposes

common knowledge, but if we take as a given that joint attention involves openness, then Hutto must be claiming that openness can be achieved by pre-linguisitic infants who according to Hutto have no mental state concepts.[51] In the case of joint attention, we could say something more about what the embodied interactive engagements that results in openness are: if A and B are jointly attending to x, then A's attention is directed toward x partly because A's attention is controlled by B's attention and B's attention is directed toward x partly because A's attention is controlled by B's attention. There seem to be no reason why this couldn't occur without A or B *representing* the attention of the other. Furthermore, this triadic interaction could be subserved by reliable subpersonal mechanisms as well as the overt behaviour of the co-attenders (bodily orientation, gaze-direction, and so on). I do not see why such interaction could not in principle deliver the kind of non-accidental joint coordination that we take to be an important feature of joint attention. Perhaps this triadic coordination, even in the absence of any metarepresentation, could even give rise to a distinctive experiential signature so that the character of A's and B's perceptual experiences of x would be different in the context of joint attention than in the context of an individual perceptual attending (as argued by Campbell, 2005). And perhaps the best way of making sense of the mutual dependency of attentional control involved in joint attention is that the triadic coordination is the realiser of what we might call a "*dual-subject* psychological

---

here.

[51] Hutto (2012) writes that what "distinguishes joint attention is that it involves a meeting of minds by means of adopting a common point of focus." (p. 309) But arguably, his most explicitly described example of a basic joint attentional episode does not involve such a meeting of minds: "Imagine two footballers battling furiously over possession of a ball. This is an activity that requires each player not only to monitor the ball's position and progress but also to keep tabs on the other's monitoring of the same and their monitoring of such monitoring. Through these means, skillful players swiftly pick up on their opponent's possible moves and strategies as they emerge during a game, allowing the exploitation of crucial opportunities." (2012, pp. 309–310) Such a competitive encounter does not seem to involve openness. Perhaps this is why Hutto does not discuss this puzzling aspect of joint attention.

state" of mutual awareness that *q*, say, that there is a block tower between us (Wilby, 2010, p. 93).[52]

While this may be a good treatment of the kind of openness involved in joint *attention*, it isn't applicable to the case of genuine joint *action*. In the case of basic joint attention, what is "out in the open" is a concretely manifested fact that is available to perception, such as "that there is a block tower between us", or "that there is a candle between us". In contrast, in the case of joint action, what must be out in the open is that I have a certain goal, and that you have the same goal, say, "that we build a block tower together" or "that a block tower is built". In a joint action, each participant is exercising her agency, that is, performing goal-directed actions. The emergent triadic coordination of embodied behaviours and objects in the world—Hutto's embodied interactive engagements—is not sufficient for agents to engage in a genuine kind of joint action.[53] The interactants in the triadic interaction of basic joint attention are concrete and relatively easy to track and monitor (e.g., my gaze direction, your gaze direction, the location of the block tower on the floor). But in the case of joint action, what must be recognised, tracked and monitored are each of our goals or intentions, and the actions that we perform in pursuit of them. The kind of openness that we need must therefore involve the agents representing each other's goal.

To account for non-accidental coordination, it must also be the case that each agent's own goal is formed partly because they represent the other has having this goal too. Moreover, they each need to somehow be aware of the fact that the other's goal

[52] In light of this radical-sounding proposal, it is worth noting that Bratman remarks, about his more recent account of shared intention, that it is "conceptually conservative in the sense that the concepts it uses—*with the possible exception of the concept of common knowledge*—are available within the theory of individual planning agency." (2009a, pp. 162–163, my emphasis, see also 2009b, p. 58) He also says: "I am not in a position to claim that [common knowledge] can itself be understood solely in terms of structures of individual agency." (2009b, p. 51 n. 17).

[53] I am not denying that some episodes of joint attentional activity qualify as this genuine kind of joint activity, but there are some joint attentional episodes that, while giving rise to openness, don't qualify as instances of this kind of joint activity.

depends on their own goal in a reciprocal fashion, so that if either didn't have this goal, then the other wouldn't have the goal either (alternatively, they need to be aware of the fact that the other wouldn't *act on* their goal if they themselves didn't act on their goal). But for an agent A to appreciate this dependency, she doesn't have to be able to attribute a belief to B that has A's own goal represented in its content, or any other belief for that matter. In other words, while an awareness of mutual goal-dependency is needed, this awareness need not be underpinned by representation of the dependency *as mediated by belief*. Unfortunately, I do not have much to say about how this mutual dependency between the goals of two individuals may arise, but perhaps the non-contentful embodied interactive engagements that Hutto refers to have a role to play here. The kind of openness that involves agents representing each other's goal may depend on a more primitive kind of openness that emerges from embodied triadic interactions.

A requirement of such an awareness of mutual goal-dependency, instead of a CK-condition, would allow agents who have the concept of goal or desire, but not that of belief, to interact with each other in such a way that their goals are "out in the open". Young children as well as non-human primates seem to be agents who fit this socio-cognitive profile. There is plenty of evidence demonstrating that even before their first birthday, young children are able to recognise the goal that the behaviour of others is directed toward (Behne, Carpenter, & Call, 2005; Carpenter, Akhtar, & Tomasello, 1998; Gergely, Bekkering, & Király, 2002; Woodward, 1998). For example, at merely 9 months of age, infants respond in appropriately different ways to superficially similar behaviours that have different goals. They express more impatience when an adult fail to give them a toy because of unwillingness than if they fail because they are unable to (Behne et al., 2005). Studies of imitation in infants show that their imitation is not modelled on the precise behavioural performance of the other agent, but that it is sensitive the goal toward which the behaviour was directed, as well as to what the most effective means of achieving that goal is (Carpenter et al., 1998; Gergely et al., 2002). It has also been shown that some non-human primates are sensitive to the goal-directness of others' actions (e.g. Call, Hare, Carpenter, & Tomasello, 2004; Phillips, Barnes, Mahajan, Yamaguchi, &

Santos, 2009). However, young children (before 3 to 5 years of age) as well as non-human primates appear to lack a proper concept of belief. This suggests that they may explain and predict the actions of others not in terms of beliefs, desires and intentions, but in terms of facts about the world (that is, what they believe to be the case) and desires or intentions that they attribute to others (Perner and Roessler (2010, p. 205) call this "the *hybrid account* of children's conception of intentional action").

One way of buttressing the idea that participants at least need to represent goals in order to stand in the relation of openness needed for genuine joint activity, is to consider the "opacity case" with the incarnations of Hector and Celia who have severe autism, presented at the end of section 2.1.1. If Hector exploits his knowledge of Celia's behavioural patterns and Celia exploits her knowledge of Hector's behavioural patterns back, then this does not seem on a par with a case where their goal-formation processes are sensitive to the goal of the other and they are aware of this goal-sensitivity. Openness involving awareness of mutual goal-dependency helps rule out opacity cases, but a completely non-representational type of openness does not rule them out. This shows that awareness of mutual goal-dependency is an ingredient in one important kind of joint activity. This kind of joint activity is less rich than the kind of joint activity that involves common knowledge, but it is richer than a joint activity that merely involves completely non-representational openness (if such openness is possible). This suggests that there is an important intuitive difference between joint activity that involves openness in the sense of awareness of mutual goal-dependency, and joint activity that involves some even weaker type of openness.

2.4 Varieties of joint activity

What I have argued is that while a CK-condition is indeed required to capture a strong notion of joint action that agents who have the concept of belief can engage in, there is a form of openness that allows agents who only have the concept of a goal, and who can be aware of the fact that their own and the other's goal stand in a

relation of mutual dependency: either both goals should be present or neither of them should. My third aim in this chapter was to substantiate the claim, made in chapter 1, that there are many kinds of joint activity, of which some require more cognitive sophistication than others.

In light of concealment and opacity cases, there seems to be at least three notions of joint activity that can be distinguished. When there is a possibility of concealment, which comes with the concept of belief, then the CK-condition is needed. While concealment is automatically ruled out when we are dealing with joint activities with agents who lack the concept of belief, openness that involves awareness of mutual goal-dependency is still required to account for non-accidental coordination and rule out opacity cases. Perhaps there is also a kind of joint activity, exemplified by basic cases of two people jointly attending to something for example, that involves a kind of openness that doesn't even involve them representing the behaviour of each other as goal-directed (I will leave this possibility open). Thus, there are different kinds of joint activity, all of which require different forms of openness. This is perfectly consistent with Tollefsen's claim that the joint activities of young children (which are characterised by whatever kind of openness that is provided by joint attention) are *of the same kind* as those activities that are captured by Bratman's account (see Tollefsen, 2005, pp. 83–84), even though Bratman requires common knowledge.[54] There are several kinds of joint activity, all of which belong to the same super-kind of joint activity that involve openness.

Now, I think it would be a mistake to identify this joint action super-kind that incorporates an overarching "openness"-condition with joint action *as such*. As I mentioned in the introduction of this chapter, Kutz and Pacherie have both proposed accounts of joint action that do not include a CK-condition, and arguably, the conditions that they replace it with do not make a demand for openness of any kind.[55] For example, Kutz claims that in order for participants to be acting together, "each

[54] For more on Tollefsen's account, see chapter 4.

[55] See footnotes 29 and 30.

must not only act in light of beliefs about the other's plan, but each must also be favorably disposed towards the other's possible knowledge of this strategic sensitivity." (2000, p. 6) Such favourable dispositions can clearly be in place without openness. According to Kutz, joint action can accordingly be coordinated in a purely accidental manner. Consider the following example, due to Ludwig (2007), who also claims that joint action can occur without common knowledge:[56]

> Suppose that country X launches a pre-emptive nuclear strike against country Y. After the initial strike, some missile silos in country Y are still operative. However, country Y has established an elaborate procedure for firing its missiles as a safeguard, which requires two on-site operators, who are physically isolated from one another, and one remote operator, all to punch in a secret code and turn a firing key at their locations in order to launch a missile. Consider the team charged with this for surviving silo 451. After the strike, which interrupts communications between them, none of them knows whether the others have survived, and have some reason, perhaps even preponderant reason, to think that they have not. Nonetheless, they intend to launch the missile. Each of them intends that they do it, and so each of them intends to do his part in launching the missile. Each punches in his code, and then turns his key, hoping that there are still others who are doing their parts, however unlikely it may seem; and so they launch the missile in silo 451 together, and they do so intentionally. (Ludwig, 2007, pp. 387–388)

If the three operators perform a joint action in this case, then this joint action is not of the rich kind that is non-accidentally coordinated, in which the goals and beliefs of participants are "completely out in the open" (the operators surely have the dispositions required by Kutz though). But I see no reason to deny that there is some notion of weakly joint action according to which it is correct to say that the two operators did indeed launch the missile together, and did so intentionally.

This is consistent with my remark about the case of the severely autistic incarnations of Hector and Celia. Even if their mutual exploitation of each other's behavioural patterns is completely devoid of openness, their activity might still qualify as a form of weakly joint activity. Arguably, keeping a broad open view of different varieties of joint action will be important if we what to understand, for example, the

[56] See Kutz (2000, p. 18) for another example used to make the same point.

development and evolution of capacities for, and forms of, more advanced kinds of joint activity.

2.5 Conclusion

Most accounts of shared intention and joint action include a CK-condition, but the condition is typically included without any explicit motivation. In this chapter I have tried to make the motivation for a CK-condition explicit. I have argued that there are indeed good reasons for including a CK-condition: First, there is the need to rule out certain problematic counterintuitive cases that involve elements of concealment. Without the CK-condition, we would have to classify such cases as instantiations of genuinely joint activity. Secondly, genuinely joint activity should be non-accidentally coordinated. A CK-condition is one condition that rules out joint activities that are accidentally coordinated. Furthermore, I have defended the condition from objections that common knowledge is too cognitively demanding in general (as common knowledge seems to require of agents that they make an infinite number of inferences to higher-order beliefs of an infinitely high order). While I do not want to claim that I have decisively put these objections to rest, I showed ways in which one can defend accounts of common knowledge from these objections. In effect, there are good reasons for including a CK-condition in accounts of the kind of joint action that adult humans engage in when they, say, go for a walk together or wash dishes together.

I have also argued that common knowledge is merely one among many (or two, at least) forms of openness. Agents who lack the concept of belief, such as young children, can still engage in joint activities that are characterised by openness concerning the goals of the participants, since there are forms of openness that do not require that the participant have the concept of belief. I have suggested that what is needed in order to make sense of the idea that participants can coordinate a joint activity in a non-accidental manner is that each is aware of the fact that their goals stand in a relation of mutual dependency. However, they need not represent this dependency as being mediated by the other's belief about their own goal. Hence,

young children who have the concept of a goal but not that of belief can appreciate this form of openness. Since they do not have the concept of belief, one of the motivations behind the CK-condition, that concealment cases must be ruled out, is undermined. Perhaps there are also forms of openness that doesn't even require agents to have the concept of a goal. Such openness might be characteristic of joint attentional episodes that very young children engage in for example. However, I have argued that such openness is not adequate for an account joint *action* (as opposed to joint *attention*).

In the final section of the chapter, I suggested that there is no good reason to think that the presence of some form of openness is a condition that is constitutive of joint action *as such*. Some everyday notions of joint action require some form of openness, but reflection on cases suggest that there are also weaker notions that don't require any form of openness.

There is one feature that all the cases of joint action that I have described or referred to so far seem to have in common, namely that they all involve actions that are performed by different agents but which in some sense are all directed toward the same goal. In some cases, this converging directedness is present in virtue of a Bratmanian shared intention. In other cases, it is present in virtue of sociopsychological structure that I will call a "joint goal". In the next chapter, I look at what it means for actions to be directed at the same goal—for agents to have a "common goal"—and what it takes for two or more agents to instantiate the sociopsychological structure that I identify with a joint goal.

3 What it takes to have a joint goal

In this chapter, I develop an account of cooperation based on the notion of having what I will call a 'joint goal'.[57] As in the previous chapter, the aim is to develop conceptual and theoretical resources for understanding the joint activities of agents that are cognitively unsophisticated compared to normal human adults. According to the account of a form of joint activity that I propose here, participants in such an activity do not need to represent what they are doing (or about to do) as part of a joint activity.[58] Furthermore, while they must have a common goal, they do not need to represent this goal *in the same way*; they may each construe the goal in a different "aspectual shape".[59] Both these features contrast my account with most philosophical accounts of joint activity. Most accounts require that participants intend "that we *J*" or intend "to perform my part of our *J*-ing" (Alonso, 2009; Bratman, 1992, 1993; Kutz, 2000; Pettit & Schweikard, 2006). For our carrying of the sofa to be a genuine joint activity, each must intend "that we carry the sofa". On the proposal that will follow, our carrying of the sofa could be a form of joint activity if each had the goal "that the sofa is moved downstairs" or "to move the sofa downstairs" (that is, without either of us representing this as the result of a joint or cooperative activity). Furthermore, on most accounts, the participants need to represent their common goal in the same way (so that the goals are specified intensionally in the same way). This requirement is typically not explicitly spelled out, but it is, I believe, widely embraced (Miller, 1995, p. 53 is explicit about this). In order for us to have a shared intention to kill Batman, it is not sufficient that you intend that we kill Batman and I intend that we kill Bruce Wayne. If we are not aware of the fact that "Batman" and "Bruce Wayne" are co-referring terms, then we will not be able to coordinate a joint

[57] For the sake of simplicity, I will from now on restrict myself to talk about two agents and I will only discuss examples of joint cooperative action involving two agents. But I see no reason why my account may not be scaled up to cases involving more than two agents.

[58] Pacherie and Dokic (2006, p. 110) and Butterfill (2012a) also allow for this possibility.

[59] To use Searle's term (1992, p. 155).

assassination. Such a requirement is appropriate if we are giving an account of the joint activities of planning creatures such as adult human beings who are equipped with capacities for linguistic communication. When we engage in complex joint activities involving the coordination of plans, breakdowns in communication and coordination is to be expected if we represent our goals and surroundings in different ways. Since I will here be focused on joint activities in the here and now, where goals can be anchored to perceptually available objects in the participants' immediate environment, such a 'same aspectual shape'-requirement is less apt.

Among cognitive scientists who are interested in developmental and comparative aspects of cooperation, cooperation is typically defined in terms of two or more agents pursuing a "shared", "joint", "common", or perhaps "collaborative" goal (e.g. Boesch & Boesch, 1989; Brinck & Gärdenfors, 2003; Brownell & Carriger, 1990; Chalmeau & Gallo, 1995; Henderson & Woodward, 2011). Unfortunately, these researchers rarely (if ever) explicate what they mean when they use these terms. In the animal cognition literature, the notion of a common goal sometimes seems to simply refer to a situation where the same token object is the *target* of several agents' goal-directed states or actions. In an influential paper on cooperative hunting among chimpanzees, Boesch and Boesch (1989, p. 550) operationally define cooperation as "two or more individuals acting together to achieve a common goal". Chalmeau and Gallo (1995) talk about the common goal of the chimpanzees studied by the Boesches as being "the prey" itself (p. 103). Similarly, Brinck and Gärdenfors talk of goals as actual objects that are present in the agents' environment, such as "water to drink, food to be had, or an antagonist to fight" (2003, p. 485). It may be that these are just elliptical ways of referring to full propositional contents such as "that we catch the prey" or "that we fight the antagonist". After all, the content of a goal determines under what conditions the goal is satisfied, it does not merely determine which object an agent's bodily movement is directed toward (see Jacob, 2012, p. 209). Nevertheless, the way that goals are talked about in the animal cognition literature suggests that a common goal perhaps do not need to have any collective content (or "'we'-content"). Joint cooperative action directed toward such a common goal wouldn't demand of the participating agents that they have a concept

of cooperation or joint action—of "our *J*-ing". Given our interest in the cooperative activities of relatively cognitively unsophisticated agents, an account that did not require of agents that they have goals with collective content would arguably be preferable.

My aim in this chapter is to explicate a notion of joint action that is broadly in line the Boesches' definition of cooperation that I referred to in the previous paragraph.[60] I use the term 'common goal', 'same goal' and 'shared goal' as synonyms. Thus, if two agents have a shared goal or the same goal, then they have a goal in common. Besides specifying what exactly this entails, I will give an account of what must be the case in order for a common goal to enable cooperation. The mere fact that two agents happen to have the same goal does not itself facilitate cooperation after all. In this chapter, the aim is thus to develop an account of what I will call a 'joint goal', such that when two or more agents have a joint goal, they are cognitively situated vis-à-vis each other and their surroundings in such a way that the fact that they have a common goal can play a role in facilitating and coordinating joint action in pursuit of the goal. I reserve talk about *one* agent "sharing *in* a goal" (with one or more other agents), to refer to how one party of a joint goal is cognitively situated in relation to the other parties and the surrounding environment (see page 110).

I am using the term 'joint goal' in a way that is analogous to the way that Bratman uses the term 'shared intention': it is a label for a type of socio-psychological structure that is appropriately situated. Coordinated activity that is the causal outcome of this socio-psychological structure is, I claim, an interesting kind of joint activity. My account is thus a kind proto-version of Bratman's account of shared intentional activity. It is a "proto-version" because a joint goal cannot play the full functional role that Bratman takes a shared intention to play. Recall from section 1.1 that a Bratmanian shared intention is a socio-psychological pattern of mental states

[60] The Boesches' definition of cooperation appears to have been influential among those who study cooperation among young children and non-human animals. See, for example, Chalmeau and Gallo (1995), Brownell et al. (2006) and Naderi et al. (2001).

that facilitates the coordination of the participants' plans and intentional actions in ways that track their common goal, as well as structures their bargaining with regard to this coordination. The role of a joint goal is narrower: A joint goal helps coordinate the actions of two or more agents with regard to the agents' common goal. The account is thereby applicable to the joint activities of what Bratman calls "merely purposive agents" who lack plan-intentions and who may have no linguistic abilities at all (and thus have no means to engage in bargaining).

In the next section, I articulate some terminological choices and assumptions that I have made. In section 3.2, I use an example of apparent cooperative problem-solving from the animal behaviour literature to explain why the notion of having a common goal should be part of a useful notion of cooperation. I then go on to distinguish, in section 3.3, between three different senses in which two or more agents can have a goal in common, and argue that the sense that is crucial for understanding cooperation among relatively cognitively unsophisticated agents is that of two agents having an "extensionally common goal". When two agents have a goal in common extensionally, then the goal of each agent will be satisfied by (more or less) the same sets of states of affairs, but the agents need not represent these states of affairs in the same way, under the same aspectual shape. Merely *having* a common goal is not sufficient for this commonality to play a role in facilitating cooperation though, and in sections 3.4 and 3.5 I discuss what else needs to be the case for a common goal to be able to play the wanted facilitating role. I summarize and present the account of a joint goal in section 3.6. In section 3.7, I try to show how this notion could be empirically tractable by looking at some behavioural phenomena that count as evidence for the presence of a joint goal.

3.1 Terminological preliminaries: outcomes, goals and goal-directed states

The goal of an action, say, "that the mug is grasped", picks out an outcome among many that defines what must be the case for the action to be successful. If I reach out to grasp a mug, my action might have all of the following outcomes: I push the

handle on the mug so that it rotates counter-clockwise, I cast a shadow on the table, I make ripples in the coffee, and also, I grasp the mug with my hand. Only the last outcome is the goal of my action; the other outcomes are side effects of my action. So the term 'goal' refers to the outcome that an *action* (or activity) is directed toward. But as I will use the term, it also refers to the content of a goal-directed (mental) state of an *agent*.[61] I assume that there is a close connection between these two senses of 'goal': an action is directed toward its goal in virtue of the content of a goal-directed state that the agent of the action is in. Despite this assumption, it is important to keep the senses apart since an observer may recognise what goal an action is directed toward without attributing a goal-directed state to the agent that is performing the action. Recognising the goal-directedness of an action is not the same as recognising an agent as guided by a goal-directed state. Furthermore, an agent may have a goal without concurrently performing any overt bodily movements, while mentally preparing to perform an action for example.

I will not defend my assumption that actions are goal-directed in virtue of the content of a goal-directed state of the agent performing the action. But I take it that is a plausible assumption. This does not mean that behaviours may not be purposeful in a weaker evolutionary sense, without having this purposefulness in virtue of a goal-directed state that is the cause of the behaviour. Perhaps the behaviour of insects is purposeful only in this weak sense.[62] Arguably, the behaviour of a collection of agents may in this weaker sense be collectively purposeful too. For example, consider the feeding behaviour of a family of *Stegodyphus* spiders. When a large prey such as a fly lands in the family's web each spider independently approaches the prey in response to the vibrations in the web caused by the prey (Ward & Enders,

[61] Note that the term 'goal' is sometimes used to refer to the goal-directed state itself (e.g. Tomasello, Carpenter, Call, Behne, & Moll, 2005, p. 676) and the term 'goal state' is in turn sometimes used to refer to the content of a goal-directed state (e.g. Gallese, 2010, p. 207; Millikan, 2004, Chapter 16)—that is, to what I refer to as a 'goal'. The term 'goal-*directed* state' has the advantage of avoiding this ambiguity.

[62] The egg-laying behaviour of the female *Sphex* digger wasp repeatedly glossed by Daniel Dennett as the outcome of an inflexible automatic mechanism comes to mind (e.g. Dennett, 1998, pp. 191–192 n. 11). For critical discussion about the empirical details and philosophical use of the *Sphex* example, see Keijzer (2012).

1985; quoted in Brosnan et al., 2010, p. 2701). Each then (again, independently of each other) starts to pull the prey toward their communal nest where the prey is digested and consumed by all. No spider could on its own catch and transport the large prey. Even if there are no goal-directed states involved in the guidance of this behaviour, the function of the behaviour of each spider may well be to bring it about that the prey is collectively brought to the nest, and this can only occur if several spiders simultaneously pull at the prey. If this is actually the evolutionary function of spiders' behaviours, then it is arguably an interesting case of a form of collective behaviour (see Butterfill (2011) for a deflationary definition of joint action that would capture this and similar cases). However, as I will use the term 'action', the spiders are not performing actions, and thus, the spiders' collective behaviour do not constitute a form of joint action either. This is not a substantive philosophical point, but a circumscription of the phenomenon that I am interested in. The point of controversy in debates about whether or not creatures such as young children, non-human primates or social carnivores are really engaging in joint cooperative action concerns what kinds of cognitive processes and representations (if any) that enable them to engage in the apparently cooperative activity; whether their actions are directed (in the evolutionary sense) toward a collective outcome is beside that point.[63]

The goal of an agent is a non-existent intentional object that is determined by the content of the agent's goal-directed state. This state successfully or unsuccessfully guides the performance of the action. The state of affairs that is brought about can either satisfy or frustrate the goal. The conditions of satisfaction of a goal are determined by the content of the goal-directed state, as well as by its context. Some philosophers take the content of a mental state to be its complete conditions of satisfaction (e.g. Searle, 1983), but I will instead take the conditions of satisfaction to

[63] I do not want to deny that questions about distal and proximate causes of behaviour are related. As Noë (2006, p. 13) points out, if there are opportunities to be had in interactions that is cooperative in the biologist's sense (interactions that result, on average, in a fitness benefit for all participants involved), then it is likely that (possibly psychological) mechanisms that improve individuals' ability to partake and coordinate their actions with each other will be selected for.

be determined by both a state's content and its context (following Recanati, 2007). The conditions of satisfaction for my mug-grasping action are not merely determined by the content of my goal-directed state but also, for example, by where and when 'here' and 'now' are indices for. If I now have the goal "to grasp the mug", then that goal will not be satisfied if I grasp the mug tomorrow. On this view, an agent need not represent everything that determines the conditions of satisfaction.

An agent can have many goals at the same time, and more than one goal can be satisfied by the bringing about of the same state of affairs. If I have the goal to make soup for dinner, the goal to eat something with plenty of iron in it, and the goal to use that last celery stalk that is left in the fridge, then the state of affairs of me making and eating lentil soup with celery in it will satisfy all three goals. Goals are individuated by their contents, so the goal of making soup and the goal of making lentil soup are two different goals, even if they are had in the same context. The sets of states of affairs that satisfy these goals are different but overlapping (assuming that the specification of the content of the first goal is not merely an elliptical specification of the content of the second goal).

Since I am assuming that a goal is the content of a goal-directed state, I will have to clarify what the relevant notion of "sameness" should be when agents are said to have the same goal. Do the goals have to have the same content as it is intensionally specified or is extensional sameness sufficient? This makes the presentation of my account different from a closely related account, namely Butterfill's (2012a) account of what he calls "shared goals" (he uses the term 'shared goals' to refer to a pattern of goals and expectations about goals-directed actions that can play a certain functional role, rather than to refer to two agents simply having *the same* goal). Butterfill takes actions to be directed toward goals, but he does not assume that they are goal-directed in virtue of any goal-directed states (2012a, pp. 36–37). An effect of this is that Butterfill doesn't have to worry about what the criteria should be for agents to have the same goal.

The claim that an action is goal-directed in virtue of the content of a goal-directed state of the agent performing the action needs to be qualified slightly. The claim is that her behaviour (or thought) is caused either by (i) a mental state with a world-to-mind direction of fit such as a desire or an intention (meaning that if there is a discrepancy between world and goal, then this calls for the world rather than the goal to be changed), or by (ii) the combination of some desire or intention and a means-end belief, where the means is the goal toward which the agent's action is directed. An example of (ii) is the following: I have the desire to eat and the means-end belief that "if I cook dinner, then I will get to eat". This combination may cause me to go into the kitchen, look into the fridge, put some ingredients in a pan, and so on, a combination of actions—an activity—that are all directed toward the goal of cooking dinner. Note that in this example the goal of cooking dinner may not be the content of a state with a world-to-mind direction of fit. While there is a state with world-to-mind direction of fit involved (the desire to eat), the content of this state is not the goal of cooking dinner. Of course, it is conceivable that, in the process of deliberating about what to do, I form or acquire an "instrumental desire" or intention to cook dinner (see Sober & Wilson, 1998, p. 217). But this need not be what happens. As Sterelny (2003) points out, "[w]e can trade talk of instrumental goals for talk of beliefs", and thus "convert intentional explanations that mention instrumental goals into intentional explanations that mention only ultimate goals." (p. 88) We can imagine a creature with only one desire—say, the desire to survive—but who still performs various actions and engages in various activities directed toward many different goals.

3.2 Why cooperation requires a common goal

It will be useful to start off by considering why it is that having a common goal should be a defining feature of cooperation. I will do this by examining a case drawn from research on of cooperative problem solving among non-human animals. Two possible interpretations of this case will be considered. According to the first interpretation, the two agents who are involved in the case do *not* have a common goal. According the second interpretation, they do.

In a study of cooperation in a species of large parrots called keas, Tebbich et al. (1996) tested whether kea dyads could together make food available for retrieval using a seesaw. The food reward was placed in a box under a Plexiglas lid that was attached to one end of the seesaw (see Figure 1). If a kea stepped onto the seesaw's other end, the lid was lifted, making the food available for the other kea to retrieve.

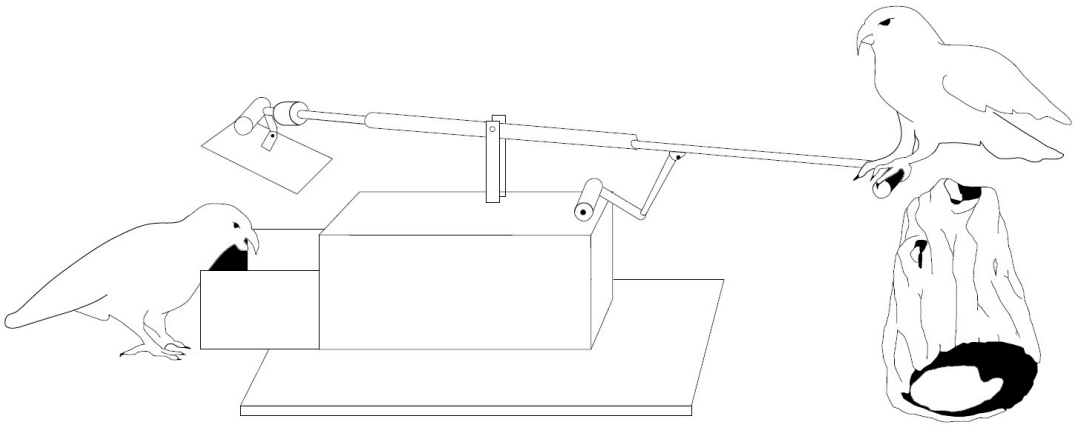

Figure 1 Testing apparatus used by Tebbich, Taborsky, and Winkler (1996).

Three of the four dyads that were tested succeeded in bringing about this state of affairs on most trials. The kea that retrieved and ate the food—the "recipient"—was always the dominant member in the dyad, and the "donor" who stepped onto the seesaw was always the submissive member. The dominance relationships between the keas were determined through observations of the keas interacting in a context independent of the setup of the experiment. Tebbich et al. showed that the dominant kea is more likely to approach the submissive kea in an aggressive manner the further the submissive kea is from the seesaw (this correlation was absent in control trials in which no food was placed under the lid).

On one plausible interpretation, the successful retrieval of the food by the dominant kea is an outcome of the submissive kea's goal to avoid aggressive attacks ("that I avoid aggressive attacks") and the dominant kea's goal to retrieve the food ("that I retrieve the food"). Assume for the moment that these are the only goals that are relevant for explaining the keas' behaviour. Given the way the keas' environment

was set up by the researchers, the opportunities for satisfying these goals are interdependent in such a way that if one goal is to be satisfied, then the other goal will have to be satisfied too. The only way in which the donor can avoid aggressive attacks is for the donor to step onto the seesaw, and the only way in which the recipient can retrieve the food is to make the donor step onto the seesaw. In the experiment, the goal of each kea is thus only satisfied by the outcome of the donor stepping onto the seesaw. Let us call the case when given this interpretation *The Individualistic Keas* case. If this interpretation is correct, then the keas do not have a common goal.

Perhaps the dominant kea's actions are also directed at the goal "that the submissive kea gets onto the seesaw", since it has—through fairly advanced causal reasoning—figured out what will happen to the lid covering the food if that state of affairs is brought about. There is no need to assume that the submissive kea has this goal though, and given the right learning history, the dominant kea's threats and attacks may not be directed at this goal either. He may simply have been conditioned through reinforcement learning to perform the threats and attacks in such a way that the likelihood of the food becoming available for retrieval is maximised.[64] Naderi et al. here suggest such an interpretation:

> Dominant keas might have learnt that a kea on the handle 'means' they can get food, and their aggressive behaviour toward their companion was aimed at forcing them to sit on the handle. The submissive kea might have learnt that the handle is a safe place from the attacks of their dominant companions, etc. and this explanation leaves little room for an interpretation as being a complex co-operative interaction. (Naderi, Miklósi, Dóka, & Csányi, 2001, p. 61)

The point that Naderi et al. make is that individual activities may converge in such a way that they "mimic cooperation" (Chalmeau & Gallo, 1995, p. 102). In light of similar considerations, Chalmeau et al. (1995) claim that in order to establish that the

[64] However, such a behaviourist explanation would have a hard time explaining that a "role switch" immediately occurred when a formerly submissive kea ended up in a new dyad where it was the dominant member. Instead of then behaving like a donor again, it switched to behaving like a recipient.

observed behaviours of two or more agents (in their case, two chimpanzees) really amount to cooperative action in an interesting sense, we must rule out that the behaviours are not just the result of brute reinforcement learning during pursuits of purely individual rewards. In the case of the keas, since avoidance of bullying will be a reward that is contiguous with the submissive kea's stepping onto the seesaw, this behaviour will be reinforced. Likewise with the dominant kea's aggressive behaviour against the submissive kea whenever it isn't near or on the seesaw: the food-reward will be contiguous with not threatening or attacking the submissive kea when it stands on the seesaw, so this behaviour will be reinforced.

This contingent environmental interdependence of the two keas' pursuits of their respective goals doesn't seem to be sufficient for their behaviour to be cooperative in any interesting sense. While it is true that the goals drive the keas' behaviours to converge on the only state of affairs that happens to satisfy their goals, this is too contingent on the specifics of the keas' environment. We want the notion of cooperation to pick out a type of situation where agents with the right sort of psychological enabling mechanisms can reap the benefits of coordinated action in flexible and novel ways. In *Individualistic Keas*, the fact that the behaviours of the keas converge on a state of affairs that satisfy both goals isn't contingent enough on their goals and beliefs. The notion of cooperation should therefore be tied to the content of goal-directed states and to how these contents are related.

Consider the following alternative interpretation of the keas' behaviour: The dominant kea realises that in order to reach the goal of retrieving the food, it must come about "that the submissive kea gets onto the seesaw". As a result of his[65] practical reasoning, he thus adopts this as his goal. Similarly, the submissive kea realises that in order to avoid aggressive threats or attacks from the dominant kea, he should try to bring it about "that I get onto the end of the seesaw". If we assume that the submissive kea's goal also can be glossed as "that the submissive kea gets onto the seesaw", then the two keas have the same goal in the sense that the content of

[65] Both members of all dyads that "solved" the task successfully were male.

their goal-directed states can be described in the same way. Furthermore, the same set of states of affairs would satisfy both their goals. Now, assuming that the keas can detect that they have the same goal here—or at least, that their behaviour is sensitive this fact—then the keas can achieve the result that the dominant kea retrieves the food and the submissive kea avoids all threats and attacks, without any of them having to spend any energy on enforcing or resisting coercion. Let us call this *The Goal-Aligned Keas* case. Contrasting this case with the *Individualistic Keas* thus shows how agents can potentially benefit from having the same goal, allowing smoother and more efficient coordination of behaviour. This is why cooperation is a form of multi-agent activity that is directed toward a common goal of the participating agents. I will postpone the question of whether the keas are cooperating here (the coercion involved may give one doubts) and maintain focus on what it is for two agents to have a common goal. This question is more complex than it may seem.

3.3 Having a common goal

There is one sense of having a common goal according to which, roughly, two agents have a common goal if the content of their goal-directed states can be specified using the same sentence. Consider Searle's example of several individuals running to take shelter from the rain:

> Imagine that a group of people are sitting on the grass in various places in a park. Imagine that it suddenly starts to rain and they all get up and run to a common, centrally located shelter. Each person has the intention expressed by the sentence "I am running to the shelter." But for each person, we may suppose that his or her intention is entirely independent of the intentions and behavior of others. In this case there is no collective behavior; there is just a sequence of individual acts that happen to converge on *a common goal*. (Searle, 1990, pp. 402–403, my emphasis)

This is not, as Searle points out, a case of collective behaviour or joint activity in any interesting sense, even though all the individuals involved have a type of goal in common (and represent that type of goal in the same way: each represents the running as performed by "I"). Each agent has a goal that is directed toward the same *type* of state of affairs. The term 'common goal' is sometimes used by Searle and

other philosophers as a label for this relationship between goals (Cohen, Levesque, & Smith, 1997, p. 96; Miller, 1986, p. 133; see also Kutz, 2000, p. 5 n. 7 on "same goal"). However, this is clearly not sufficient (nor is it, perhaps, necessary) for two agents to have a common goal in the sense intended by the Boesches and other researchers who have adopted their definition of cooperative action. I will therefore refer to this relationship between agents as them having *a common goal-type*.

Whether having a common goal-type facilitates the cooperation between two or more agents will depend on if and how the satisfaction of one agent's goal is connected to the other's agent's goal satisfaction, and this will be up to what the environment is like in which the agents are embedded. If each of us have a goal "that I make lentil soup for dinner", then this may develop into a situation where we all make one big pot of lentil soup that we together have for dinner. But if we each have the goal "that I win the race" in one and the same hundred meter sprint race, then the fact that we have a common goal-type will not help us coordinate our actions to achieve some collective outcome.

3.3.1 Intensionally common goal

So-called "Frege cases" suggest that, for the purpose of making sense of an agent's cognition and behaviour, we cannot individuate beliefs and goals extensionally. In Frege cases, an agent holds seemingly inconsistent beliefs or other attitudes but is unaware of this because she does not know that the attitudes are representing one and the same object under different aspectual shapes (or "modes of presentation"). As far as the agent knows, she is representing two different objects. Thus, I may both intend to kill Batman at the same time as I intend to save Bruce Wayne's life. If I am unaware of the fact that Batman and Bruce Wayne are two names (aspectual shapes) that refer to one and the same person, then I can have both these goals without being criticisably irrational for having inconsistent intentions. We can easily make sense of this if we individuate goals intensionally. In contrast, it is far from obvious how those who take mental states to be individuated extensionally can make sense of these cases.

For the purpose of explaining the *joint* action of *several* agents, similar considerations arguably recommend that beliefs and goals should be individuated intensionally. After all, in many circumstances, the mere fact that two agents have goals directed toward the same states of affairs will not help them coordinate their actions unless they also represent those states of affairs in the same way. If I have the intention that we kill Batman and you have the intention that we kill Bruce Wayne, then this will not help us coordinate a joint assassination even if the same states of affairs will satisfy the *de re* intention of each to see the man variously referred to as Bruce Wayne or Batman dead. In order for our common goal to play a role in making us join forces in such a case, we thus need what I will call an *intensionally common goal*. In order for several agents to have an intensionally common goal, it is not sufficient that the same states of affairs satisfy their goals; they must also represent these states of affairs under the same aspectual shape.

How goals should be individuated across agents has not been a topic of discussion or debate in the joint action literature, although the question of how mental state contents should be individuated is big topic in philosophy of mind in general. However, Seamus Miller explicitly states that, in order for several agents to have what he calls a "collective end", the goal that they are directed toward, must be "aimed at under the same description" by them all (Miller, 1995, p. 53).[66] The phrase "under the same description" invites the thought that the requirement is only applicable to agents with linguistic capacities, but we can express the same underlying idea by saying that agents have to aim at the same state of affairs with (more or less) *the same aspectual shape* in order to have the same goal. Presumably, the reason Miller commits to this same-description requirement is that merely having

[66] Miller elsewhere formulates this same-description requirement in a somewhat less strict way, so that the state of affairs only needs to be "aimed at under more or less the same description by each agent." (Miller, 2001, p. 58). Kutz also seem to require several agents to have an intensionally common goal in order to have what he calls "shared goals". He submits that, "[b]ecause the objects of intention are intensional […], the question of whether agents' [participatory] intentions overlap depends upon the way their joint activity is described." (2000, p. 20)

an extensionally common goal will not be able to facilitate cooperation in many cases. Indeed, when it comes to more complex forms of joint activity that we "adult humans in a broadly modern world" engage in, which involve verbal communication and long-term planning, the ways in which participants represent their goals and subgoals will have to be more or less aligned. For language-using creatures like adult human beings, such cases may be quite common.

Besides Frege cases, another reason for thinking that propositional attitudes cannot simply be individuated extensionally concerns how indexical thoughts function. In the *Goal-Aligned Keas* case, the keas have a common goal because the goal of each kea concerned the same individual. As I characterised the content of their goals, they represent this individual in the same way, namely as "the submissive kea". But one might think that it is more plausible that the submissive kea represents this goal as "that *I* get onto the seesaw", rather than as "that *the submissive kea* gets onto the seesaw". (I will later consider whether the best way to gloss the goal is simply as "to get onto the seesaw", without any explicit representation of the agent herself at all.) These two ways of representing one and the same desired set of states of affairs are not equivalent since they may have very different cognitive and behavioural consequences. For the purpose of explaining an individual's behaviour, we cannot specify the content of the indexical "I" in merely extensional terms. Perry (1979) vividly illustrates this point with a story about himself following someone in a supermarket who is leaving a trail of sugar on the floor, presumably because the shopper's sack of sugar is torn. At this point, Perry in the story truly believes "that the shopper is making a mess" and he is trying to catch up with the shopper to inform him or her about this. He then looks down into his cart to suddenly realise that it is *he* who is the shopper making a mess. In the story, Perry thus forms the true belief "that *I* am making a mess", stops the cart and rearranges the sack so that the sugar is no longer spilling out of it. Note that when Perry realises that *he* is the shopper who is making a mess, he acquires a new piece of information that makes a difference to his behaviour. This shows that we cannot, for the purpose of explaining behaviour, individuate all beliefs simply according to their extensionally specified contents, that is, the sets of states of affairs that would make the beliefs true. Suppose that we do

individuate beliefs in this way. Then the belief that Perry in the story has "that the shopper is making a mess" will be the same belief as the belief "that I am making a mess". Given such a principle of individuation, we cannot explain Perry's apparent insight and the behaviour that follows from it by appeal to a change in his web of beliefs (after all, no such change has occurred: his web of beliefs after his insight will be exactly the same as it was before the moment of insight). The same point can be made about goals. If the submissive kea had the goal "that the submissive kea gets onto the seesaw", but did not know that *he* was the submissive kea, then this goal would not be able to explain his action of stepping onto the seesaw. As the submissive kea realised that he was the submissive kea, he would form the goal "that *I* get onto the seesaw". This adds weight to the idea that, for the purpose of explaining behaviour, we cannot merely individuate goals in terms of their extensionally specified contents. On the other hand, if we want to preserve the idea that the two keas do have a common goal in the *Goal-Aligned Keas* case, then we cannot simply think of the indexical 'I' as contributing a Fregean sense to the goal. If we did that, then the dominant kea's goal "that the submissive keas gets onto the seesaw" would not be the same goal as the submissive kea's goal "that I get onto the seesaw". I will return to this in the next subsection.

While a 'same aspectual shape' requirement may chime with our intuitions, as well as be appropriate in the context of an account of the robustly joint activities of linguistic creatures such as adult human beings, it is not appropriate for my current purpose. My purpose in this chapter is to construct an account of socio-psychological structure that helps agents who are relatively cognitively unsophisticated to reap the mutual benefits that opportunities for joint cooperative action provide them with. In light of this purpose, the 'same aspectual shape' requirement is too strong. It is primarily when language-using creatures engage in complex planned joint activities that breakdown of coordination may be expected due to a mismatch of *how* the world is represented (in language) by the different agents. When we are dealing with joint activities in the here and now, with goal-contents that are anchored to objects that are perceptually available to all participants, it is sufficient for agents to aim at the same states of affairs without necessarily representing them in the same way. I submit that

an extensionally common goal can play a role in facilitating cooperation in many circumstances. Agents who presumably have very different outlooks on their surroundings can thus engage in the kind of joint cooperative activity that I want to give an account of. Consider the joint cooperative action of young children and their caregivers, or the joint activity of a guide dog and its blind owner as they navigate to a familiar destination such as, say, the grocery store (Naderi et al., 2001). Arguably, this destination will be represented in very different ways by the guide dog and its human owner, and there is no reason to think that such an extensionally common goal could not be the basis for a socio-psychological structure that coordinates their activity with regard to their (extensionally) common goal.

3.3.2 Extensionally common goal

By taking the relevant notion of having a common goal to be the notion of having an extensionally common goal, the keas in the *Goal-Aligned Keas* case will count as having a common goal. This is as it should be, even if it introduces an asymmetry between how goals should be individuated intrapersonally and how they should be individuated interpersonally. As Perry shopping story illustrated, for the purpose of explaining individual behaviour, the contents of beliefs and other attitudes cannot simply be individuated extensionally.

Given a fairly well-known view of what the contents of intentions—one important kind of goal-directed state—are, fixing the notion of a common goal to that of an extensionally common goal seems to create a problem though. According to this view, the self, the indexical 'I' figures in the content of an intention (without it, the state carrying the content in question cannot be an intention but must be some other type of mental state). This seems to be Searle's (1980, 1983) view for example. Björn Petersson (2011) calls this the "agent-reference thesis" regarding the content of intentions. According to Searle, the content of a prior intention to raise my arm that I have prior to actually raising it is "[that] *I* perform the action of raising my arm by way of carrying out this intention" (Searle, 1983, p. 92, my emphasis).[67] On

[67] The content of the intention-in-action that finally causes the bodily movement is

Searle's view, goals that two agents have in virtue of their intentions will never be the same goal in the extensional sense. Given commitment to the idea that agents engaged in joint cooperative action must have a common goal, this view leads to some counterintuitive consequences.

Consider the following cooperative problem-solving task used by Brownell et al. (2006) to test young children's capacities for cooperation. Dyads of young children faced an apparatus with two handles separated by a transparent screen that controlled a toy animal, a dog-puppet that could move ("dance") to music emitted from its inbuilt speakers ("sing") (see Figure 2). To activate the dog-puppet, both handles had to be pulled (more or less) simultaneously.

Figure 2 Apparatus used in the cooperation task (from Brownell et al. 2006, p. 808).

Due to the distance between the handles and the placement of the plastic screen, one child couldn't simultaneously pull both handles on their own. In other words, due to

---

self-referential in a somewhat similar way, but the self does not figure in the content *as an agent*. Searle specifies the content as "[m]y arm goes up as a result of this intention in action" (Searle, 1983, p. 93). The object of an intention-in-action is not an action, but a bodily movement or some other event that is not an action.

the structure of the environment, the contributions of both children were required for either child to reach their goal that the dog-puppet danced and sang. Now, if each child has the goal "that *I* make the dog-puppet dance and sing", then the activity of each child is directed to separate sets of states of affairs. The child on the left will be trying to bring about that *he* makes the dog-puppet dance and sing, while the child on the right will be trying to bring about that *she* makes the dog-puppet dance and sing. This would thus be similar to Searle's case of people running for shelter in that both children have the same goal in the thin sense that their goals can be specified using the same sentence. But unlike in Searle's case, each agent can only achieve his or her goal if the other agent also achieves his or her goal. Even so, if we individuate their goals according to their extensionally specified content, then the children do not have a common goal. But while I think Searle is correct in saying that we often represent ourselves as agents in the content of our *prior* intentions (future-directed plans), it strikes me as implausible that we explicitly represent ourselves as agents with respect to our more immediate goals. As Petersson (2011) points or, both everyday talk about what people intend and desire and the phenomenology of agency suggest that the agent-reference thesis is wrong when it comes to our present-directed agency.

To gloss the goals of the children in Brownell et al.'s experiment as "that I make the dog-puppet dance and sing" seems odd. We typically express what we intend to do using infinitival clauses, so each child's goal should arguably be described as "to make the dog-puppet dance and sing". The phenomenology of agency suggests something similar. As I type these words on my laptop for example, or as I reach out and grasp the coffee mug on my desk, my focus of attention is not on my own involvement in achieving certain outcomes—the appearance of certain words on the screen, taking a sip of coffee—but on the outcomes themselves. Plausibly, this is also true of the children facing the apparatus with handles and the dog-puppet: They are focused on bringing about the outcome that the dog-puppet starts to dance and sing, not on that they themselves play a role in realising this outcome. The focus of each child is presumably to make the dog-puppet dance and sing, rather than that "I" make it so.

If each child simply has the goal "that the dog-puppet dances and sings", then their goals will be directed toward the same states of affairs and they also have the same content. But this goal could be satisfied even if they do not participate in making the dog-puppet dance and sing. It is certainly possible that this accurately reflects what their goals are, but intuitively, it seems plausible that part of what the children want is to activate the dog-puppet. The children do not merely want to *observe* the dog-puppet dance and sing, in which case a child would be equally satisfied if they observed the other child and a third agent make the dog-puppet dance and sing together. What each child wants is arguably still to *bring it about* that the dog-puppet became active. An experiment by Watson and Ramey (1987) showed that 3-month-old infants respond with much more pleasure in response to the movements of mobile if the movements occurred in response to their head movements (the head movements activated the mobile through a pressure-sensitive pillow against which the infant's head was resting), than if the movements occurred in response to the whims of an experimenter. A plausible interpretation of this result, suggested by Bermúdez (1995), is that the infants take more pleasure from watching the mobile move in response to their own movements because they appreciate the fact that they make the mobile move, that they have exercised their own agency.

While the children in Brownell's experiment were 1- or 2-year-olds, it is a plausible that their preferences in this respect are similar to those of the 3-month-old infants: They prefer to observe changes brought about by their own agency than to observe the same changes occurring independently of their actions. But this does not mean that the infants or the children explicitly represent themselves as agents in their goals. Rather, the role of their agency may merely be implicitly represented in the cognitive system's architecture (Recanati, 2007 makes a similar point of about the causal self-reference of perceptual states).[68] The function of the goal-directed states that guide the children's handle pulling (or the infants' head movements) to bring

[68] This is not how Bermúdez formulates his take on Watson and Ramey's experiment. He argues that the infants represent themselves explicitly when they take pleasure in exercising their own agency, but that this explicit representation is nonconceptual. This difference is not important for my purposes here though.

about their goals would in some sense fail to fulfil their function if the states weren't involved in bringing about the state of affairs that satisfies the goals—that the dog-puppet dances and sings, or that the mobile moves. This implies that the states would also fail to fulfil their function if the child or infant herself was not involved in making the state of affairs in question come about. However, there is no reason to think that the states would fail to fulfil their function if other agents were also playing a causal role in bringing about the satisfaction of their goal.

Note that if each child has the goal "to make the dog-puppet dance and sing *on my own*", then the goals are directed at states of affairs that are mutually exclusive. Some goals thus exclude contributions from others in achieving the goal, independently of what the world is like. Following Miller and Tuomela (2001, p. 6), I will say that a "[g]oal P of an agent X is *dividable* if and only if (X believes that) there will, or at least can, be parts or shares for at least one other agent to bring about (or sustain) P." As long as the goals are dividable, the fact that agents have the same goal in a thin sense may thus play a role in facilitating joint cooperative action. Crucially for my purposes, a goal can be dividable without concerning a joint activity. If I set out in my prison cell to dig a tunnel that emerges outside the walls of the prison, my goal would not be (normally) not be frustrated if I discover, as I have dug half the distance, that you have been digging a tunnel from the outside toward my cell. As we meet in the middle, our tunnel-digging goals are satisfied, even though neither of our goals was about a joint activity. Dividable goals may of course also be goals "that we *J*" or "that I perform my part of our *J*-ing", but the important point here is that they need not be. Dividable goals may also be only implicitly concerned with one's own agency or they may be "agent-neutral" (Butterfill, 2012b; see also Gold & Sugden, 2007a, p. 130).

On the view I advocate, the goal of each child is thus most plausibly characterised as "to make the dog-puppet dance and sing" or "that the dog-puppet dances and sings", in which case they are both directed at the same state of affairs. This allows us to hold on to the idea that for the purpose of understanding cooperation among the cognitively disadvantaged (relatively speaking), the useful sense of sameness of

goal-contents have to do with the represented state of affairs, not with the way in which the state of affairs is represented. In other words, the pullings on the handles in Brownell et al.'s experiment might be directed toward a common goal, even if one child represents the dog-puppet as, say, "Pluto", while the other child instead represents it as "Snoopy".

As I have formulated the notion of an extensionally common goal, two (or more) agents have a common goal if and only if their goals are satisfied by *the same set of state of affairs*. But this is actually a simplification, because as it stands, it is too hard and fast: there need not be *perfect* overlap between the states of affairs that satisfy the goals. Or rather, whether two agents have a common goal is not an all or nothing affair. To make this point vivid, consider the following case: Suppose that Fritz and Tom are two cats who are unusually social hunters and playmates.[69] Each of them has a goal concerning the same mouse. Fritz's goal is "that we play with the mouse" and Tom's goal is "that we hunt the mouse". Now, suppose that Fritz mistakenly ascribes his own goal, "that we play with the mouse", to Tom, and that Tom mistakenly ascribes his own goal, "that we hunt the mouse", to Fritz. As a result of these goals and goal-ascriptions, the two cats end up chasing the mouse in a coordinated fashion. After some chasing, the mouse is caught and Tom feeds on it. This outcome satisfies both Fritz's goal and Tom's goal.

Intuitively, it seems right to say about this case that Fritz and Tom jointly chased the mouse in a cooperative fashion and that they did this partly in virtue of the partial overlap of their goals. The overlap is partial because there are states of affairs that will satisfy Fritz's goal but not Tom's, and there are states of affairs that will satisfy Tom's goal but not Fritz's. If Fritz simply lets the mouse go after they have been chasing it around for a while (to Fritz's delight), then only Fritz's goal will be satisfied. Alternatively, if the two cats start by approaching the mouse together and the mouse does not notice them before Tom knocks the mouse unconscious with his

[69] Thanks to Till Vierkant for putting forward a less elaborate version of this example as a challenge the claim that having a common goal is necessary for joint cooperative activity.

paw and immediately proceeds to eat it. In this case, there has not been any chase (that is, no play), so only Tom's goal will have been satisfied. Still, despite these counterfactual possibilities, it is not clear that we should exclude this as a case of joint cooperative action.

Note that this case is similar to *The Individualistic Keas* case that I described in section 3.2. In both cases, there are two agents who seem to lack an extensionally common goal but given the environment they find themselves in, their goals are partly overlapping. However, we concluded that the dominant and the submissive kea did not have a common goal at all (recall that on *The Individualistic Keas* interpretation, the dominant's goal was "that I retrieve the food" and the submissive's goal was "that I avoid aggressive attacks"). This conclusion is correct because there are so many counterfactual but plausible circumstances —nearby possible worlds—in which only the goal of one kea is satisfied. It is unlikely that their goal-pursuits will lead to an outcome that satisfies both their goals, but as it happens, within the confines of the particular experiment, this is exactly what one would expect to happen. If we bracket the role of the researchers' design of the experiment, it is merely an accident that the behaviours of the keas converge in a way that makes both their goals satisfied. The play/hunt of Fritz and Tom is quite different, at least given some plausible assumptions about cats and mice. Assume, for example, that in most circumstances in which two cats and a mouse meet, the cats will end up chasing the mouse both if the cats play with the mouse and if they hunt the mouse. Since, let us assume, a mouse is very good at detecting and tracking cats, it is very unlikely that a cat would be able to sneak up on a mouse and immediately kill it. Furthermore, if two cats chase a mouse, the mouse will become very tired and eventually become an easy prey for a cat that intends to catch and eat it. If this is all true, then in most circumstances in which the play/hunt were to occur, Fritz and Tom would mutually benefit from coordinating their goal-pursuits even if they do not, strictly speaking, have an extensionally common goal. [70] On the other hand, if these

[70] Compare with Kutz's discussion of what it means for two agents to share goals (see 2000, pp. 20–21): "[A]gents' intentions overlap – they *share goals* – when [(a)] the collective end component of their participatory intentions refers to the same [type of] activity or outcome and [(b)] when there is a nonempty intersection of the sets of

assumptions are false, and the cats' joint chase of the mouse was more of an accident, then we would arguably be less inclined to say that the joint chase was a case of cooperative action after all.

The lesson to draw from this is that having a common goal is not an all or nothing affair. While it is clear that the individualistic keas do not have a common goal, Fritz and Tom have a common goal to some considerable degree (in all nearby possible worlds where Fritz and Tom have these goals, their goal-pursuits will lead to a coordinated chase of the mouse). The category of common goals is thus a fuzzy one, and so, by implication, is the category of what I will call "joint goals". However, to simplify the discussion, I will continue to assume that for two or more agents to have a common goal, their goals must be such that they are be satisfied by the same set of states of affairs.

To sum up, I have argued that the notion of common goal that is useful for the purpose of understanding joint cooperative activities of relatively cognitively unsophisticated agents such as preverbal infants and the non-human animals for example, is the notion of an extensionally common goal rather than that of an intensionally common goal. This marks a difference to what is otherwise a useful notion of sameness when we are considering the behaviour of one and the same agent. Note that I am not denying that a stronger notion of sameness is required for agents to count as participating in some more robust intuitive form of joint action,

---

states of affairs satisfying those collective ends." (p. 20) I take it that Kutz is here giving necessary and sufficient conditions (labeled (a) and (b) by me) for several agents to have a goal in common. In order for several agents to have a common goal, Kutz thus only requires that there is *one* possible state of affairs that satisfies their goals. But this does not mean that he would say that the dominant and the submissive kea share a goal in *The Individualistic Keas* case since the keas do not have participatory intentions (there is no "collective end component" that is part of the content of their intentions). Note though, that Kutz elsewhere in the same paper says that "when we share a goal, the intentions of each are *only* satisfied by the performance or realization of the same token activity or outcome." (p. 5 n. 7, my emphasis) This is in line with my much more stringent initial formulation of what is required for having an extensionally common goal, according to which two agents have a common goal only if the overlap between the sets of states of affairs that satisfy their goals is perfect.

according to which agents need to represent the goal in the same way. However, the purpose here is not to capture an intuitive everyday notion of cooperation or joint action. Instead, our aim is to explicate a notion of a having a common goal (or having the same goal) that is useful for theorizing about cooperation among, for example, non-human animals and very young children.

3.4 The recognised action interdependence condition

The fact that two or more agents have an extensionally or intensionally common goal is not by itself a fact that is conducive to cooperation. Whether it is or not will depend, among other things, on the structure of the agents' environment and the tasks that the agents are facing. In the context of Brownell et al.'s cooperative task presented in the last subsection, having a common goal to make the dog-puppet dance and sing *is* potentially conducive to joint action. But suppose that the apparatus was designed differently, so that the following was the case: If both handles are pulled simultaneously, then the apparatus jams and the dog-puppet breaks. The dog-puppet is only activated if only one of the handles is pulled out. In such circumstances, the fact that the children have the goal "to make the dog-puppet dance and sing" in common is more likely to hinder than to help that their common goal becomes satisfied. If it was the case that each handle when pulled activated the dog-puppet, irrespectively of whether or not (and when) the other handle is pulled, then the fact that the children have their common goal will make no significant difference with regard to how each of them pursue their goal. Furthermore, even if the tasks and the environment of the participants are such that the common goal is potentially conducive to joint action, the participants' behaviour must also be sensitive to this goal commonality for it to be actually conducive to joint action. This is why having a common goal is not enough for two or more agents to have a *joint goal*. There must also be *recognised interdependence* with regard to the pursuits of the common goal from the perspective of each participant.[71] Recall that the Boesches

[71] I am borrowing the term "recognized interdependence" from Bratman (2006, p. 2). Bratman is concerned with recognised interdependence of intentions rather than actions though.

(1989) operationally define cooperative behaviour as "two or more or more individuals *acting together* to achieve a common goal" (p. 550, my emphasis). Perhaps the inclusion of "acting together" reflects a recognition that it is not sufficient for two or more individuals to all be trying to achieve a common goal for this attempt to be joint or cooperative in any interesting sense.

Note that if the participants have intentions "that we *J*" as required by Bratman's account, or perhaps intentions "to perform my part in our *J*-ing", then there will, irrespectively of what the participants' environment is like, be actual interdependence between the contributions of the participants. For me to succeed in performing my part of our *J*-ing *as my part*, you must successfully perform your part of our *J*-ing, and vice versa. The same goes, of course, for each of our intentions "that we *J*" However, since we do not want our account of joint cooperative action to demand of participants that they have intentions or goals with such collective content, a recognised interdependence condition needs to be explicitly incorporated into our account.

Other accounts do not require that agents have intentions or goals with collective content, such as Seamus Miller's "collective end"-account and Stephen Butterfill's "shared goals"-account of joint action. Butterfill outlines a set of relations between goals and expectations that he suggests may play an important coordinative function in the joint activities of young children. The expectations include "expectations about a common effect", which means that "on balance each agent expects this goal to occur as a common effect of all of their actions directed to the goal, her own and others'." (Butterfill, 2012a, p. 40) Seamus Miller also incorporates a condition of recognised interdependence in his account of what he calls a "collective end", which he takes to be a necessary ingredient in joint action. In the following, Miller's 'end' is what I refer to as 'goal':

> [Collective ends] are ends that are necessarily shared; they are not ends that are shared only as a matter of contingent fact. Suppose you and I are soldiers being shot at by a single sniper. Suppose further that we both happen to see the sniper at the same time and both fire at the sniper in order to kill him. I have an individual end,

> and you have an individual end. Moreover, my individual end is one that I share with you; for the death of this one sniper will not only realise my individual end, it will simultaneously realise your individual end. However, this is not yet a collective end, for the fact that I have as an end the death of the sniper—as do you—does not generate independent [*sic, recte* interdependent] acts of shooting. For neither of us needs the other to realise the individual end we share. Accordingly, I may well retain my individual end, when I know you have abandoned yours. As it happens we have shared individual ends; but they are not necessarily shared individual ends. (Miller, 2001, pp. 58–59, also 1995, pp. 52–53)

Put into my terminology, Miller's requirement would demand that, for two agents to have a joint goal, their common goal must be such that it can only be satisfied if they both pursue the common goal, and furthermore, each agent must be in a position to be moved by this fact to act in pursuit of the goal. To be positioned in this way, they must have certain expectations about the likely effects of various means they can take to achieve their goal and about how these effects may combine with the likely behaviour of the other agent. It is not merely required that there is actual interdependence between the agents' goal-pursuits but that this interdependence is *recognised* by the agents themselves (if it is not recognised, then the fact that they have a common goal cannot play a role in generating their interdependent actions). Suppose that the sniper in Miller's example is actually a supervillain who cannot be killed by normal bullets, but only by injections of two particular substances that mixes into a deadly poison in his body. While we are completely unaware of this, our rifles are unbeknownst to us loaded with bullets filled with these particular substances; your bullets with one substance, mine with the other. In this case, the satisfaction of our goals happens to necessarily intertwined since we both need to fire and hit the sniper in order to get her killed. This mere fact wouldn't generate interdependent *actions*, even if our goals in the given context happen to be such that the satisfaction of one of them is necessarily tied up with the satisfaction of the other.

It is unclear why Miller's formulation of the interdependence condition is so strong. Instead of requiring that the goal is such that, by each agent's own lights, it *couldn't* be achieved merely by the performance of the agent's own action, we should just require of each agent they recognise that the combined effect of their own and the others' actions increases the likelihood of goal achievement, and that this recognition causes them to perform their own action in pursuit of the goal. Furthermore, this

increase in likelihood of goal achievement may merely be a summative increase.[72] Miller seems to require that the recognised interdependence must be stronger kind, so that the recognised increase in likelihood cannot be merely summative. On such a stricter view, it may be that you and I firing at the sniper would involve a joint goal if the actions we are about to perform complemented each other in such a way that they combined to increase the likelihood of the sniper being shot dead beyond that of merely giving us two chances of hitting the sniper. For example, if you fire your rifle first and I fire mine in quick succession, then if your shot doesn't kill the sniper, it will get her off-balance and thereby make it easier for me to hit and kill with the second shot. You might aim slightly more to the left than you would otherwise do, in the hope that the sniper will take cover toward the right where I will have a better line of sight than I do now. I might anticipate this, waiting to fire when the sniper goes for cover in the way expected. If we both fire our shots in light of such expectations, then this stricter recognised interdependence requirement would have been satisfied. However, it is not clear to me why we should formulate the requirement in this strict way. It should be sufficient if the recognised increase in likelihood of goal achievement causes each agent to perform his or her action. If the soldiers in Miller's example would have shot anyway, irrespectively of what they expect the other soldier to do, then the fact that they have a common goal will play no role in generating their action. However, if each soldier judges that the chance that just one shot hits the sniper is too low for him or her to bother taking a shot and each soldier also judges that given the summed chance of one of their two shots hitting the sniper is high enough for them to each take a shot, then the fact that they have a common goal will have played an role in generating their actions of shooting at the sniper (even if the interdependence is involved here is "merely" summative).

Note that this recognised interdependence requirement rules out the *Goal-Aligned Keas* case as an example of the kind of cooperation we are trying give an account of here. Given the setup of the experiment, there is no sense in which the fact that the

[72] This doesn't rule out that there are *other* actions that could be performed the agents could perform on their own which would increase the likelihood of goal achievement even more.

dominant kea has the same goal as the submissive kea, namely "that the submissive kea gets onto the seesaw", plays a role in making it more likely that this goal is achieved. Once the submissive kea has this goal (possibly as a result of being bullied by the dominant kea), the dominant kea is not in a position to contribute to the achievement of this goal. While there is dependency between the goals—the submissive kea wouldn't have the goal to get onto the seesaw unless the dominant kea had that goal first—there is no dependencies between the actions the keas perform in the pursuit of their common goal. (Of course, this is true given the circumstances of the experiment. With a different setup and apparatus, the recognised interdependence requirement might have been met.)

3.5 Representational and metarepresentational "glue"

For two or more agents to be able to take cooperative advantage of the fact that they have a common goal, it is not sufficient that they each recognises the interdependence between their own future action and some action that the other agent(s) may perform. An agent must also expect, at least to some minimal degree, that the other agent(s) will in fact perform this action. But this does not mean that the agents must be able to attribute goal-directed states to the other agents. As I have previously speculated, the hunting performed by packs of wolves, hyena or lions may be facilitated by the fact that the pack members have a common goal, even if none of them is aware of this fact. They may be able to successfully predict and coordinate with each other in virtue of this fact that the animals have a common goal that, even if they do not represent the others' behaviours as goal-directed. Instead, they may merely be "smart behavioural readers".

Smart behaviour reading will not enable agents to engage in flexible forms cooperation in which novel goals are adopted and pursued though. Without the ability to recognise the goals toward which the actions of others are directed, agents will not be able to reason about various possible means by which these goals may be achieved. In other words, being able to predict the future behaviour of others is not all that goal-recognition enables an agent to do. It also enables her to help others by,

for example, directing the other's attention to means of achieving her goal that she isn't currently using. Such skills may clearly be useful in facilitating more effective and robust joint cooperative action.

At any rate, insofar as we are interested in the joint cooperative behaviour of young children and non-human primates, there is plenty of evidence demonstrating that young children—from before their first birthday—and some non-human primates are indeed skilled goal-readers (Behne et al., 2005; Call et al., 2004; Carpenter et al., 1998; Phillips et al., 2009; Woodward, 1998).[73] However, note that the fact that they can recognise the goals that another's action is directed toward does not mean that they attribute *goal-directed states* to the other agents. Instead, they may simply represent the goals as states of affairs toward which the actions are pulled, or represent them as the function of the bodily movements in question (Csibra & Gergely, 2007; Gergely & Csibra, 2003). In other words, recognising the goals that actions are directed toward does not necessarily require metarepresentational capacities. One can recognise a goal that another agent's action is directed toward without representing that goal as something that is represented by the other agent. But perhaps young children do have a mentalistic understanding of goal-directedness, even if they do not yet have a similar understanding of beliefs. With such a partial understanding of other minds, they could make sense of actions in terms of partially subjective motivating reasons for action. Before the age of about 3 to 4 years, they would do this not in terms of beliefs and desires, in thought goes, but in terms of facts about the world (thus, they draw on their own beliefs about the world) and desires that they attribute to others (Perner and Roessler call this "the *hybrid account* of children's conception of intentional action" (2010, p. 205)).

[73] Can they also recognise whether a goal is dividable or whether it about the agent herself bringing about some state of affairs "on her own"? Research by Becchio et al. (e.g. 2010) demonstrates that at least adults seem to be able to discriminate between very similar actions (directed toward the same object) that are performed with competitive or cooperative intent, based solely on the kinematics of the bodily movements. In light of this, it doesn't seem far-fetched that young children and some non-human animals may be able to discriminate between dividable goals and purely "individualistic" goals.

Note that if we required that agents had an intensionally common goal rather than merely an extensionally common goal, then an agent could only reliably detect that it had a goal in common with another agent if it was able not merely to recognise the outcome toward which the other's activity was directed but also recognise the way in which the other represented this outcome. An understanding of goals as outcomes represented by goal-directed (mental) states such as intentions or desires will not provide an agent with this ability. To be able to do this, the agent must also be able to attribute beliefs different from their own to other agents (say, the belief that the dog-puppet is "Pluto" rather than "Snoopy"). An agent that lacks this ability will still, of course, represent the outcome that the other's activity is directed toward *in some way* (with some aspectual shape), but the agent does not necessarily attribute a perspective on that outcome *that belongs to the other agent*. There is no need for "Level-2"-type perspective taking (Masangkay et al., 1974).

To summarise, it may be that agents who are merely smart behaviour-readers are able to gain mutual advantage from the fact that they have the same goal and the opportunity to perform interdependent actions to achieve this goal. But if the agents are able to recognise the goal that the others' activity is directed toward, and able to detect that this is also their own goal, then this would allow them to cooperate in a much more flexible manner, and direct their activity toward a variety of goals. In the case of young children and non-human animals, it is rarely (if ever) necessary that the agents discriminate between different ways in which goals are represented. Typically, these agents can simply (and safely) assume that goals of others are represented in the same way that they represent them.

3.6 An account of joint goals

I will now summarise and make more explicit what I have said so far regarding the conditions under which two or more agents having a common goal can play a role in facilitating cooperation between them. I will say that two agents have *a joint goal* when they are related to each other and to their environment in such a way that the

following conditions are met. For two agents *A* and *B* to have a joint goal *g*, the following conditions must be met:

1. *Common goal condition:* Agents *A* and *B* have an extensionally common goal *g*. For this to be the case, their goals must be directed toward (more or less) the same set of states of affairs. It is not necessary that *A* and *B* represent *g* in the same way, that is, they need not have an intensionally common goal.
2. *Action-readiness condition:* Each of agents *A* and *B* is ready and able to initiate an action *a / b* directed toward *g*.
3. *Recognised action interdependence condition:* In virtue of their common goal *g*, *A* and *B* each believes (i) that the other agent will perform the action *b / a*, and (ii) that *g* is more likely to be satisfied if both *a* and *b* are performed than if either *a* or *b* is performed on its own.

If the beliefs mentioned in the recognised action interdependence condition result in *A* actually performing *a* and *B* actually performing *b*, then the joint goal of *A* and *B* glues *a* and *b* together into a joint action aimed at *g*. (The beliefs may of course turn out to be false, but I assume that they are typically true and true in virtue of the fact that the actions *a* and *b* are directed toward the common goal *g*.) Note that to have the beliefs required by the recognised action interdependence condition, *A* and *B* need not (but may) attribute goal-directed states to each other, it is sufficient that they expect that the other will exhibit goal-directed behaviour directed toward *g* (in other words, I here equate 'action' with 'goal-directed behaviour').

It will be useful to have a way of talking about the beliefs and goal-directed states that *one* agent must have in order to take part in a joint goal with another agent. I will reserve the verb phrase "to share in a goal with" to refer to this. For agent *A* to share in a goal *g* with agent *B*, the following must be the case (assuming that *A* has the concept of a goal):

1. *A* has goal *g*;
2. *A* is ready and able to perform action *a* in pursuit of *g*;

3. *A* believes that *B* will perform action *b* in pursuit of *g*;
4. *A* believes that *g* is more likely to be achieved if *a* and *b* is performed than if either *a* or *b* is performed on its own.

This account of joint action as the result of a joint goal has some counterintuitive consequences. Cases can be constructed in which all three conditions are satisfied but where the "joint action" would not ordinarily be characterised as cooperative. Recall Bratman's account of shared intentional activity and the crucial role that the second condition in making such shared intention activity cooperative. In order for the coordinated activity of two agents to count as a shared intentional activity, each agent must not only intend "that we *J*", but also intend the efficacy of the other's intention "that we *J*"—each intends that they *J* in accordance with and because of the other's intention that they do the same—as well as intend that their subplans for *J*-ing mesh. The requirement that the agent intend that their intentions are held in accordance with and because of the other's intention rules out cases where one agent tries to entirely bypass the intentional agency of the other agent. Consider Bratman's example of two people who each intends that they go to New York together by throwing the other into the trunk of the car and driving there (1992, p. 333). Even if the two end of fighting each other at the threshold of the trunk, each trying to lock up the other in the compartment, it will still be true that it is more likely that they end up going to New York if both agents perform their respective actions than if only one of them does (they won't end up going to New York if one of them simply stays at home and never even approaches the car). In this case, all three conditions that I have presented are arguably fulfilled. There are also cases that involve a non-cooperative element without anyone acting in conflict with the other's goals or chosen means. Suppose that one member of a dyad faced with Brownell et al.'s cooperative problem-solving task, agent *A*, believes that the other dyad member, agent *B*, will grab hold of the other handle, but *A* is not sure whether *B* will actually pull it out (perhaps he correctly believes that she thinks to activate the dog-puppet, it is sufficient if both handles are simply simultaneously held by them). If *A* then, in addition to performing the action of pulling out the handle on his side (action *a*), also

intends to, at the same time, grab hold of *B*'s free arm and yank at it hard in order to mechanically pull out the other handle while *B* is holding on to it (*B*'s action *b* is thus merely to hold the handle bar). Thus, the dog-puppet is activated (*g* is satisfied) partly as a result of *A* using *B* as a mechanical tool to pull out the other handle. This case does not seem to be ruled out by Bratman's second condition. However, Bratman also submits that a shared intentional activity must involve mutual responsiveness in action, and that this mutual responsiveness must be an outcome of the shared intention (Bratman, 1992, pp. 338–339). I take it that this requirement rules out that this dyad's activation of the dog-puppet would qualify as a shared intentional activity.

Neither of these two cases looks like a case of what we would intuitively call "cooperation". Should we then not include a condition similar to Bratman's second condition in order to filter out counterintuitive cases like these? For the purpose of constructing an account of what facilitates cooperation among relatively cognitively unsophisticated agents, the answer must be no. As I have pointed out earlier, Bratman's condition demand of participants not only that they have the conceptual capacities to have plan-intentions concerning the plan-intentions of themselves and the other participating agents, it also requires that they have the conceptual capacities for having plan-intentions concerning the meshing of their own and others' subplans. It seems unlikely that this falls within what, say, three-years-old children, let alone non-human primates and social carnivores have capacities for. This may not actually be a problem for Bratman himself since he is primarily interested in the shared agency of planning agents, such as "adult humans in a broadly modern world" (2009a, p. 153). However, others have adopted his account to understand what is involved in children's participation in joint activity (e.g. Tomasello et al., 2005, p. 680; Carpenter, 2009, p. 381). So it seems to at least be a problem for certain ways in which Bratman's account has been put to use.

I will bite the bullet here. I do not claim that the fact that two or more agents have a joint goal directed at *g* invariably lead to joint action that is cooperative. All I am claiming is that, under most circumstances, if two agents *A* and *B* have a joint goal

directed at extensionally common goal *g*, then this reliably facilitates coordinated cooperation directed toward *g*. I am here following Butterfill (2012), who submits that the pattern of expectations and goal-relations that he presents as able to supports an important role in coordinating joint activity, can only do so "in favourable circumstances", where the presence of coercion is one type of unfavourable circumstance (p. 42). Note though, that the presence of coercion will in many cases provide the material that is needed for a simple reinforcement learning story about how individualistic goal-pursuits may be shaped to mimic cooperation, as I pointed out in section 3.2. The involvement of coercion will thus often be a factor that favours an individualistic interpretation of what guides the behaviours that are observed.

Like Butterfill's account, which my account closely resembles, my account is not overly cognitively demanding. When it comes to the mindreading capacities required of the participants, the account at most requires that agents are able to recognise the goals of others. Furthermore, agents must have general capacities for reasoning about how several causal contributions may generate a combined effect. The condition of recognised action interdependence arguably implies non-trivial cognitive demands on such reasoning. Exactly what is required will of course depend on the task that the participants are facing.

The account I have given fits what the cognitive scientists quoted in this chapter's introduction gloss the goals that primates must have in common in order to be engaged in cooperation for example. Importantly, the account allows for, but does not require that agents represent goals with collective content or "we"-content. Thus, agents need not have any concept of cooperation or joint activity in order to participate in cooperative or joint activity on this account. All that is required is that their goals are "dividable" (Miller & Tuomela, 2001). As I have pointed out already, dividable goals may be "agent-neutral" without having any collective content (Butterfill, 2012b; Gold & Sugden, 2007a, p. 130). A virtue of this is that we can side-step problems or challenges of analytical circularity that accounts of joint action that require participants to have intentions "that we *J*" face (see Petersson, 2007).

The agents merely have to represent the goal of their action and that the other's contribution has a role to play in increasing the chance of goal achievement.

In the next and final section, I consider what kind behavioural phenomena that can be taken to support an inference to the claim that an activity is coordinated by a joint goal.

3.7 How to recognise 'joint goal'-driven joint action

My aim in this section is to show that the notion of joint goal is empirically tractable. Whether or not we should claim that a multi-agent activity is coordinated by a joint goal will depend on whether a joint goal figures in our best psychological explanation of how that activity came about. Since the issue here is one of inference to the best explanation, we should not demand a fixed set of behavioural criteria that can be used in every case to decisively determine whether a joint goal is present or not. The best one can hope for is a set of relevant constraints and factors that can be used to judge whether an inference to a joint goal is justified or not.

3.7.1 Accidental and coordinated collective achievements

One way of testing for whether two or more individuals have a joint goal is to see whether they reliably achieve an outcome that is desired by each but which can only be brought about if they coordinate their actions. In such a case, the performance of the subtask of each comes to nothing unless the other also performs his or her subtask. If one can observe that an individual only performs their subtask (or is more likely to perform it, or performs it at higher rate) when the other individual is likely to carry out his or her subtask—only when the other individual is present and facing in the relevant direction for example—then we can infer that this individual has some understanding of the role the other participant plays for successful task completion.

On the basis of this line of reasoning, primatologists, comparative psychologists, and developmental psychologists have tested whether various types of agents are able to coordinate their actions to achieve a common goal. I will consider some such studies

in light of what they can tell us about various creatures' capacities to have a joint goal. The following experimental setup from Melis et al. (2006) is representative. Melis et al. tested whether chimpanzees were able to rationally judge when to recruit a collaborator in order to retain food.[74] Each subject was presented with a platform baited with food that was placed behind a railing. A rope was threaded through metal loops on the platform, with both ends of the rope extending through the railing into the test room that the subject was released into. If a subject only pulled on one rope end in order to drag the food-baited platform toward the railing, then the rope would unthread through the metal loops and come loose from the platform (and thus make it impossible to retrieve the food). In order to get the platform to get closer to the railing, either both ends had to be pulled at the same time or one end had to be pulled while the other was held steady.

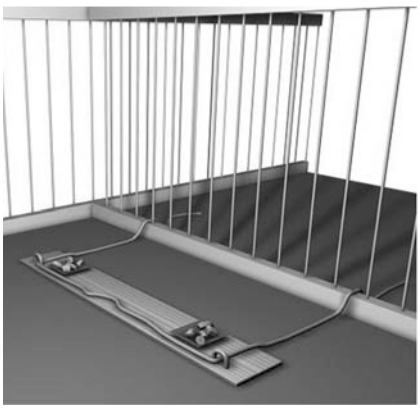

Figure 3 The baited food platform used in Melis et al's (2006) study of cooperation in chimpanzees. A rope is threaded through metal loops on the platform, with both ends of the rope extending into the test room behind the railing.

In an adjacent locked room another chimpanzee, a potential co-operator, was waiting. By removing a wooden peg, a subject could release the other chimpanzee into the test room. What Melis et al. were interested in was how the decision of a

[74] In a second experiment, Melis et al. (2006) tested whether chimpanzees could make an informed choice between two potential collaborators, which they had previously collaborated with at the same task in an introductory session. In the test session, chimpanzees preferred to recruit the collaborator who had been a more effective partner in the previous introductory session.

subject to release or not to release the other chimpanzee depended on whether it was physically possible for the subject to retrieve the food from the platform or not.

There were two conditions in Melis et al.'s experiment: the "collaboration condition" and the "solo condition". In the collaboration condition, the ends of the rope were placed three metres apart, so that it was impossible for the subject to hold or pull both ends of the rope at the same time. Hence, in order to retrieve the food on the platform in the collaboration condition, the subject would have to release the collaborator into the test room, so that each could pull one rope end and together drag the platform toward the railing. In the solo condition, then ends of the rope were placed fifty-five centimetres apart, so that it was possible for the subject to pull both rope ends at the same time. The subject could thus acquire all the food for herself without having to share half of it with the collaborator. The result was that seven of the eight chimpanzee subjects released the collaborator significantly more often in the collaboration condition than in the solo condition.[75] Melis et al. took this to show that the chimpanzees were sensitive to the fact that both their own contribution and the contribution of the collaborator were required for successfully food retrieval.

This experiment gives an either/or measure of subjects' understanding of whether or not a situation calls for cooperation: the chimpanzee can either release the potential partner or choose not to. In other experiments, subjects are repeatedly trying to do something either alone or together with a partner. This allows one to compare *the rate* at which subjects try to achieve a goal (such as retrieving food) without and with a partner. If a task requires the contribution of a partner, and there is a significant difference in the rate of attempts between when the partner is absent and when he or she is present, then this suggests that the individuals are monitoring each other's activity and modulating their activity in light of what the other is doing (see e.g. Chalmeau & Gallo, 1995; Chalmeau, 1994; Visalberghi, Quarantotti, & Tranchida, 2000 for studies of nonhuman primates).[76]

[75] Seed et al. (2008) used a very similar setup as that of Melis et al. to test rooks' cognitive capacities for cooperation, but with negative result.

[76] Alternatively, one can compare the rate when partners have different degree of

If a subject clearly shows sensitivity to the fact that a task requires another's contribution, then any interpretation that construes the behaviour of these agents as rational will involve attribution of a dividable goal to at least one agent (the subject in Melis et al.'s task). Purely individualistic goals, goals "that I do *A on my own*" cannot be satisfied if the agent knows that doing *A* is impossible without the contribution of another agent. It may of course be the case that the goal that the agent has in common with another, say "to bring the platform to the railing", is a goal he has only in virtue of an instrumental desire that he has formed in light of his ultimate desire "to eat all food myself". The goal the agent has in virtue of his ultimate desire will not be one that he has in common with another agent then.

Note that the experiments confound the situation where subjects have a common goal but fail to grasp that there is an opportunity to benefit from coordinated action, and the situation where they do not have a common goal but where they have the causal understanding required to benefit from coordinated action if they would have a common goal. The capacity to have dividable goals and to recognise that one has a goal in common with another agent can thus be masked by failure to meet other performance requirements of the task though. In particular, the task may require a quite sophisticated understanding of causal relationships for an agent to understand that the contribution of another agent can help him or her to achieve the goal. But if a subject only acts in concert with a partner, or only acts when a partner is likely to contribute, then inferring that the subject has the capacity to *share in a goal* with another seems to be justified (that is, that the subject is able to appreciate that he and another agent have a common goal and that it is more likely that the goal is achieved if both perform their contributions).

However, apparently purposive coordination may be the result of the fact that the agents are embedded in the same environment at the same time, and thus presented with similar constraints and opportunities for action. This may accidentally lead to

---

perceptual access to each other's activity, as in Mendres and de Waal (2000).

similar actions being performed roughly at the same time toward the same target object. But we cannot conclude from the fact that coordination is merely accidental that the agents do not have a common goal. Plausibly, some multi-agent activities may be coordinated in virtue of agents typically having agent-neutral goals and high social tolerance, so that they are able to act in parallel in close proximity to each other, and thereby be exposed to the same action constraints and opportunities, potentially leading to them have the same goal (see Petit et al., 1992 on "coproduction"). But this would not amount to joint activity in the sense of a multi-agent activity coordinated by what I call a joint goal.

3.7.2 Ruling out "merely local" non-accidental coordination

Recall that that the recognised action interdependence condition requires that there be recognised action interdependence *in virtue of* the fact that the agents have the common goal *g*. This means that the fact that coordination is non-accidental within the local context of a psychological or behavioural experiment may potentially mislead us. The participants' actions may be non-accidentally coordinated as a result of the wrong cause. Consider Brownell et al.'s (2006) study that I presented in section 3.3.2. Their experiment was designed to test of the cooperative abilities of 1- and 2-years-old children. Recall that the dog-puppet mounted on top of the box was activated if and only if the two handles were pulled simultaneously, which in this case meant that one handle had to be pulled within three seconds of the other handle (see Figure 2 on page 92).[77] A key issue for Brownell et al. was whether any observed coordinated behaviour, that is, any successful activation of the dog-puppet, "was coincidental or cooperative" (p. 811). They found that the successful activations of the 1-year-old dyads appeared to be accidental, but the 2-year-olds

[77] The cooperation task was actually administered in two versions. Besides the one where the children had to pull the handles simultaneously, there was a version in which they had to be pulled *sequentially* so that one of the handles (either of them) was pulled out after the other handle had been pulled out and left extended (but within three seconds after this happened). I will ignore the sequential version, since it is not important for my purposes.

successful activations were not. Hence, the activations of the 2-year-olds seemed to be the outcome of cooperative non-accidental interaction.

Brownell et al. determined whether the successful activations were accidental or not in the following way: Video recordings of the dyads' activity at the apparatus were coded according to whether their actions were "coordinated", "communicative" or "uncoordinated". For example, if a child pulled her handle when her peer was facing his own handle and had it within the reach of his arm, then this was coded as a type of coordinated behaviour that was called a "peer-proximal pull". If the participants pulled within 3 seconds of each other, then this was coded as a "coordinated pull". Types of uncoordinated behaviours included actions that interfered with the peer's use of their handle as well as solitary pulls. The coding data was then used to compute three different ratios that indicated to what extent coordinated pulling was accidental or cooperative (for example, one ratio was computed by dividing all the coordinate pulls of the dyad with the sum of all the solitary pulls performed by the dyad members).[78] The higher these ratios were, the more likely it was that the observed coordination wasn't merely the outcome of each child trying to independently make the dog-puppet move and emit music, but that it was shaped by the children's understanding of their partner's contribution to the realisation of their supposedly common goal.

---

[78] These three ratios were the following, each answering a slightly different question that were pertinent to whether the successful activations of the dog-puppet were accidental or the result of cooperative coordination:

(i) Q: How likely was it that a dyad pulled together compared to that the participants pulled individually?
A: Dyad's coordinated pulls ÷ Sum of solitary pulls in dyad;

(ii) Q: Of all of a child's handle pulls, did more of them occur when the peer was available as a partner than when the peer was unavailable?
A: Child's peer-proximal pulls (incl. coord. pulls) ÷ Child's total sum of pulls (all kinds);

(iii) Q: Were a child's individual pulls more probable when a partner was available than when the child was alone at the task?
A: Child's peer-proximal pulls (incl. coord. pulls) ÷ Child's solitary pulls when peer is off-task.

According to Brownell et al., if coordinated pulling were non-accidental, then the participants must have been exercising an ability to engage in joint action, that is, have a common goal.[79] However, I think there is a problem with this line of reasoning. At least, an inference to a joint goal is not justified. The reason why has to do with something that Brownell et al. themselves highlight. In light of what was known about children's social understanding, Brownell et al. expected (incorrectly) that even the 1-year-olds would succeed above chance level at this cooperation task. The prediction was motivated by the fact that infants at a very early age spontaneously engage in imitation, and Brownell et al. thought that imitation would probably be a sufficient mechanism for enabling successful simultaneous pulling.

Let us suppose that contrary Brownell et al.'s actual results, the 1-year-olds did succeed in non-accidentally coordinate their pullings, and that it was imitation that enabled the coordination. If this was the case, then one agent's handle pulling might merely have been an imitative response to the similar action performed by the other agent. Now, if the evolutionary function of imitation were to enable interpersonal coordination of actions toward a common goal, then the fact that one agent imitates another would at least be circumstantial evidence in support of the inference to the hypothesis that the agents have a joint goal. But mechanisms for imitation arguably evolved not to enable coordinated activity directed toward a common goal but to enable advanced forms of social learning that are exploitative rather than cooperative (Tomasello, 2009, p. xiv). Furthermore, bodily movements produced by imitation will typically not be directed at the same goal as the bodily movements of the model were directed toward (even if the model and the imitating agent may have a common goal-type). If you perform the action of raising your arm, and I imitate this action, then my action will be directed toward the goal of raising *my* arm, not of raising *your*

[79] Brownell et al. talk of "a shared goal" (p. 804) or "a joint goal" (p. 807). Those who have done similar studies of non-human primates adhere to the same logic (e.g. Chalmeau & Gallo, 1995; Chalmeau, 1994; Visalberghi, Quarantotti, & Tranchida, 2000).

arm. In other words, imitation often occurs in contexts where agents do not have a common goal.

Even if the performance of the 1-year-olds in Brownell's study had been non-accidentally coordinated, this would not have implied that this non-accidental coordination was driven by joint goal. They could each have acted on the goal "that I make the dog-puppet move and sing on my own" if their understanding of how the apparatus works was deficient. Or they could have had a common goal ("that we make the dog-puppet move and sing" or "that the dog-puppet move and sing"), and then automatic imitation would have led to a non-accidentally coordinated performance (in other words, that they had a common goal would not then have played any role in initiating or driving the coordinated pulling). Note that Brownell et al. explicitly designed the experiment in such a way that mere imitation would be sufficient to facilitate coordination for the 1-year-olds. So, in one sense, coordination would not have been accidental, but designed. But in the wild, outside the psychology laboratory, the children would have to rely on luck (rather than benevolent experimenters) to face tasks that require coordination that can be facilitated by imitation. What counts as "accidental" thus depends on what one takes the boundaries of the setting to be. Brownell's experiment was designed to enable an automatic imitation mechanism to facilitate coordination.

What I have argued is thus that what the evolutionary functions of the mechanisms are that facilitate successful coordination is sometimes of relevance in assessing whether success was the outcome of a joint goal or not. If an experiment is designed so that a mechanism such as automatic imitation can facilitate coordination, then the coordination may only be non-accidental within the local context of the experiment. In such cases, the coordination is non-accidental due to the intentions of the researchers designing the experiment, rather than intentions or goals or the participants. Note that subpersonal coordination mechanisms other than that which underpins automatic imitation may have evolved or developed specifically to enable interpersonal coordination of an activity directed toward a common goal though (Knoblich, Butterfill, & Sebanz, 2011; Sebanz, Bekkering, & Knoblich, 2006). If it

can be shown that such mechanisms are involved in enabling non-accidental coordination, then that is defeasible evidence for the presence of joint goal.

However, non-accidentally coordinated collective achievements of two or more agents certainly constitute defeasible evidence for a joint goal at work. It suggests that agents have an adequate understanding of the necessary contribution of the other agent, and that the agents have a goal that is at least extensionally in common between them. My sceptical conclusion regarding the (counterfactual!) results of Brownell et al. depends, of course, on the fact that there is an automatic imitation mechanism that an alternative explanation can appeal to. In species other than humans, this may not be the case. Note that it is fairly easy to rule out the imitation-based explanation though. All one needs to do is to construct a task that requires different but complementary contributions from the participants, rather than similar and more or less simultaneous contributions.

3.7.3 Reengagement and joint goals

Before concluding this chapter, I want to consider an experiment that tests capacities that go beyond those required for sharing in a goal. The experiment is interesting because its result is one of the key findings that Michael Tomasello appeals to in support of the claim that young children but not chimpanzees engage in "true collaboration" (which is basically "shared intentional activity" in Bratman's sense, see section 1.3.1). Felix Warneken, Frances Chen and Tomasello (2006) used subjects' responses to interruptions in a joint activity as a litmus test of whether they could engage in genuine joint activity with others. If a subject understands that the co-participant's contribution is required for successful task completion—and the subject wants to complete the task—then one should expect that the subject would try to reengage a co-participant who disengages from joint activity.

Warneken et al. had 16-months-old and 24-months-old children, as well as three human-reared juvenile chimpanzees, face four different tasks together with an experimenter. Two of the four tasks were problem-solving tasks in which the child/chimpanzee and the experimenter had to cooperate in order to retrieve a

toy/piece of food. The other two tasks were social games without any concrete toy/food reward. Figure 4 shows the objects used in the tasks that the children engaged in together with the experimenter (the tasks for the chimpanzees were very similar but not identical). The upper row shows the objects used in the problem-solving tasks, and the lower row the objects associated with the social game tasks. Both the problem-solving tasks and the social games required the participation of both the child/chimpanzee and the experimenter. For example, in the *Elevator* task, the task is to retrieve a toy from a container that can be pushed up and dropped down like an elevator (see the upper left photo in Figure 4). To succeed with this, one participant must push up the "elevator" from below, so that the other participant can reach the toy placed inside it. A participant cannot retrieve the toy on her own in this way. (Note that non-accidentally successful performance on this task cannot be the outcome of automatic imitative responses.)

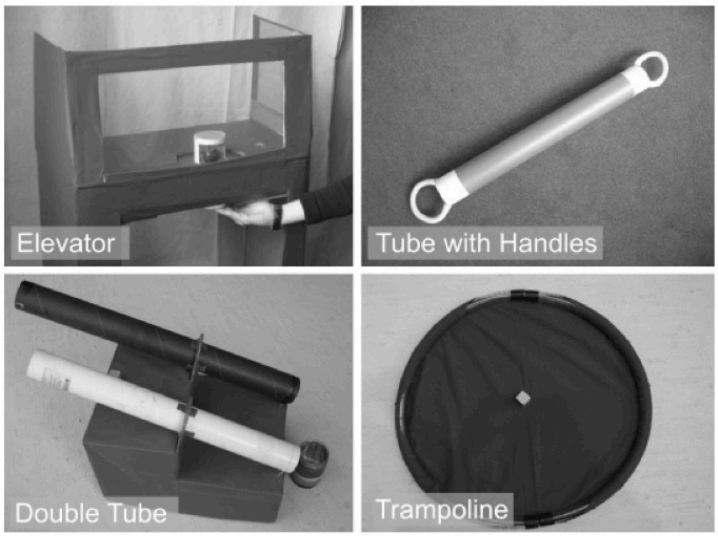

Figure 4 The tasks used for the children in Warneken et al. (2006). The chimpanzees had similar but not identical tasks. The tasks on the top row are problem-solving tasks where the participants could retrieve a toy or a piece of food if they coordinated their actions with the experimenter's. The bottom row tasks were social games, where the participant together with the experimenter could make an object slide down a tube to make a rattling sound if it is caught with the tin can (left), or make a wooden block bounce on the trampoline (right).

In some trials, the experimenter stopped performing his role within the joint activity for 15 seconds. For example, in the *Elevator* task, the experimenter let the elevator back down before the child had reached the toy or, if the roles were reversed, he

started to reach for the toy when the child has pushed up the elevator but then withdrew his hand. The researchers then looked at whether the child tried to reengage the experimenter and make him perform his role in the activity.

Warneken et al. found the two following differences between young children and chimpanzees with regard to their responses to this type of interruption. First, (1) unlike the chimpanzees, the young children were motivated to engage in cooperative social games with the experimenter/partner for their own sake, irrespectively of whether this allowed them to take part in bringing about a "concrete goal" (p. 660), such as retrieving a toy or a piece of food.[80] Secondly, (2) the children but not the chimpanzees tried to reengage the experimenter/partner if she interrupted and disengaged from an on-going joint activity, irrespectively of whether this was a social game or a problem-solving task (by, e.g., trying to push and move the experimenter or vocally instructing her to perform her role in the activity). The chimpanzees never made any such attempts. Instead, they either abandoned the activity or tried to perform the task on their own. Unlike the children, the chimpanzees never attempted to communicate with the highly familiar experimenter after the interruption (who they otherwise routinely directed gestures toward in order to get access to food). However, the chimpanzees were able to perform the cooperative problem-solving tasks successfully together with the experimenter as long as she was engaged in the activity.[81]

[80] Some studies suggest that other non-human primates (bonobos and gorillas) engage in social play for its own sake (on bonobos, see Pika & Zuberbühler, 2008; on gorillas, see Tanner & Byrne, 2010).

[81] In a follow-up study, Warneken, Gräfenhain and Tomasello (2012) used—on the one hand—tasks that children (21- and 27-month-olds) could perform individually or with the experimenter, and—on the other hand—tasks where the experimenter's contribution was necessary. However, the children did not respond to interruptions differently in these two types of task, but made reengagement attempts in equal measure. This suggests, according to the researchers themselves, that the first difference that was found in Warneken et al. (2006) is genuine, and that the children are motivated to engage in joint activity for its own sake, and not just as a means to achieving their own individualistic goals.

Why didn't the chimpanzees attempt to reengage a disengaged experimenter, despite the fact that they can reliably coordinate their activity with the experimenter in order to reach what is arguably a common goal? Tomasello (2009) argues that the differences between the human and chimpanzee subjects here boils down the fact that the children have the capacities to engage in "shared intentional activity", whereas the chimpanzees do not. Children's reengagement attempts are explained by the fact that when they are engaged in a joint cooperative activity, their activity is controlled by intentions or other goal-directed states that concerns not just their own activity but the activity of the experimenter as well ("that we *J*"). These goals will be frustrated if the other disengages from the activity, so we should expect that children try to reengage the experimenter after they disengage from the activity. According to Tomasello, that chimpanzees do not even attempt to reengage a partner when their contribution is necessary for task completion suggests that they don't have the capacity to have goals with such we-content.

However, note that if a chimpanzee has the to goal "to retrieve food" or "that I eat the food", then this goal will also be frustrated if the other disengages from the joint activity. If the chimpanzee understands that the experimenter's contribution is required for successful food retrieval, then it may seem that one should expect that the chimpanzee try to reengage the experimenter, even if it does not have the goal "that *we* retrieve the food". When the experimenter disengages, then the chimpanzee (or the child) and the experimenter no longer have a joint goal. After all, the experimenter's goal is to do nothing for 15 seconds, not to do his part of the retrieving the food/toy for the subject. This observation suggests to me that there is an alternative explanation of the difference between the children's and the chimpanzees' responses to disengagement. Perhaps the children are just more skilled than the chimpanzees are at attributing a larger variety of goals, including novel ones, to others. While chimpanzees understands that an experimenter can be manipulated to behave in a certain way, they might not understand that he can be manipulated by way of influencing his goals. Unlike chimpanzees, children may have the capacity to attribute goal-directed states such as desires to other agents (not merely the capacity that the chimpanzees share, to recognise the goal-directedness of

activities). Even when the experimenter disengages and stops acting entirely, the child can thus still think of the experimenter as a "carrier" of a wide variety of potential goals. For the chimpanzees, on the other hand, the goal that was previously held in common and which the chimpanzee recognised in the experimenter's activity simply disappears when the experimenter disengages and sits still.

3.8 Conclusions

I have constructed an account that should be useful for understanding a kind of joint activity in which relatively cognitively unsophisticated agents can participate. I defined this kind of joint activity in terms of a sociopsychological structure that I called a "joint goal". Two or more agents have a joint goal when they have a common goal and are cognitively positioned vis-à-vis each other so that they are able to benefit from this fact. I argued that agents do not need to represent a goal in the same way in order for it to be their common goal. Furthermore, a common goal need not have collective content ("we"-content): it need not have contents such as "that we catch the prey" or "that I do my part in our hunting of the prey", but may be agent-neutral ("that the prey is caught") or only implicitly identify the agent of the action ("to catch the prey"). These features of my account set it apart from e.g. Bratman's account, and arguably, most accounts of joint activity that appeal to the notion of shared, joint or collective intention. Both features also help make my account less cognitively and conceptually demanding than Bratman's account. Participants need not engage in Level-2 perspective taking in order to take advantage of the fact that they have a common goal, and they do not need to have the concept of "that we *J*". Being an agent that has a joint goal with another agent does not seem to be too cognitively or conceptually demanding for, for example, young children or various kinds of non-human primates then.

4 Socially extended intentions-in-action

Recall that according to Bratman's account of shared intentional activity, if you and I are engaged in such an activity *J*, then our *J*-ing must be the outcome of our shared intention to *J*. For us to have a shared intention to *J,* each of us must intend that we (continue to) *J*. If you and I are tangoing together, each of us must thus have had an intention of the form "I intend that we tango." Following Christopher Kutz, we may call such ordinary intentions with collective content or we-content *group-intentions* (2000, p. 21). A group-intention is an intention that belongs to an individual participant, but whose content refers to the joint activity of all participants. Sometimes the presence of overlapping group-intentions is claimed to be a necessary condition for an activity to count as a genuinely joint one (Bratman, 1992, 1993; Pettit & Schweikard, 2006; Alonso, 2009). At other times, it is presented as at least present in a core class of cases (Bratman 1997, 2009a, 2009b; Kutz 2000).[82] However, appeals to group-intentions are problematic since the content of such intentions – "that we *J*" – appears to violate a widely accepted constraint on what can be intended by an agent. The constraint, which I call *the exclusivity constraint*, says that the performance of another agent's action cannot be part of the content of one's intention to do something (Searle 1983; Kutz 2000; Bardsley 2007; Bratman 2009a, 2009b). This constraint appears to be logically implied by what Bratman calls the *own-action condition* on something being an intention, which says that one can only intend to perform one's own actions (1997, p. 53, 2009b, p. 156). Assuming that an action can only belong to one agent, the exclusivity constraint follows from the own-action condition.[83]

---

[82] Recall that Bratman is open to the possibility that shared intention may be multiply realizable, perhaps intentions "that we J" is not part of all realisations of shared intention (see footnote 9). Kutz (2000) writes: "Group-intentions are ordinary, instrumental individual intentions whose subject is the individual agent and whose object is a collective act or outcome: I intend that we will dance the tango. Clearly some paradigmatic forms collective action incorporate our action as the direct aim." (p. 21)

[83] Arguably, the combination of the idea that genuine joint activity is characterised

As we saw in section 1.2.1, Bratman argues that his planning conception of intention allows him to be more liberal about what it is intelligible to intend (see also Pettit and Schweikard 2006, p. 21). In Bratman's words, "the planning conception of intention supports the legitimacy of the appeal to my intention that *we J*." (1992, p. 331) The intentions of others can function appropriately in our plans since we are able to reliably predict that another agent will form a certain intention in many situations. This allows you, for example, to intend that I tell you the time by means of intending to ask me what time it is (Bratman 1997, p. 58). Simple examples like this show, convincingly I think, that the exclusivity constraint is not a constraint on the contents of plan-intentions.

But some joint activities that adult participants spontaneously engage in do not involve the kind of deliberation and planning that is arguably required for a shared intention (in Bratman's sense) to be formed. For example, you and I may be strangers who show up at a milonga where we spontaneously start dancing tango with each other upon merely making eye contact. Or we might be acquaintances who bump into each other on the street and fall into an informal conversation that moves from topic to topic in an improvised manner.[84] While it is no doubt possible to assimilate such cases into Bratman's planning framework (see e.g. 1999, p. 139), one might suspect that a more plausible account of such cases is possible outside it. In addition, consider joint activities in which participants are highly skilled at performing the activity and does not rely on their own skilful performance for success, but also rely on the skilful performance of the other participants with which

---

by participants having group-intentions and acceptance of the exclusivity constraint is what motivates what Risjord calls "the standard picture" of joint action (see section 1.3.2). According to this picture, the jointness of a joint activity has its sole source in shared prior intentions that define roles and plans into which the ordinary individual non-joint actions can be fitted.

[84] Bratman (2009a, p.150; also 1999, p.130) takes "our having a conversation together" as an example of the kind of "small scale shared intentional agency" that he is interested in. Kutz (2000) also mentions "conversing" as an example of a joint activity (p. 2). So does Gilbert (2006, p. 4).

they are interacting. Examples that come to mind include that of two professional dancers or ice skaters performing a dance together, or acrobats performing a well-rehearsed coordinated stunt. In such cases, the coordination of action does not appear to go by way of the coordination of plan-intentions. There is simply no time for planning and practical reasoning to do any work. Note also that it is partly reliance on a planning conception of intention that makes Bratman's account too conceptually or cognitively demanding to make sense of the joint activities of relatively cognitively unsophisticated agents such as young children and non-human animals (Tollefsen 2005; Pacherie 2007, p. 166; Pacherie and Dokic 2006, p. 110). Participants must be capable of both *having* group-intentions and *recognising* them in other participants.

In this chapter, I will focus on how to best account for these skilful tightly coupled joint activities. To accommodate such activities, joint action theorists must either provide an account of joint activity that does not appeal to group-intentions at all, or argue that group-intentions and recognition of group-intentions are possible outside Bratman's planning framework. By relying on John Searle's account of intention and action (1983, ch. 3, see also 1980), I will show that the latter option is plausible, and argue that we can have what I call *socially extended intentions-in-action*. In some circumstances, agents cannot merely intend *that* we *J*, but can also perform an action which includes the performance of another agent's action in its Intentional content.[85] That is, one can intend (us) *to J* in such a way that the own-action condition is met even if the content of the intention violates the exclusivity constraint.

In the next section, I outline Deborah Tollefsen's (2005) analysis of joint activity, which is primarily developed to deal with the joint activities of infants and toddlers,

[85] I will follow John Searle (1983) in writing 'Intentional' with a capital 'I' when I refer to Intentional content in the technical philosophical sense of the satisfaction conditions of mental states. In other words, not only intentions or actions have Intentional content, but so do beliefs, desires, hopes, etc. The Intentional content of a belief is the conditions under which the belief is true. The Intentional content of an intention is the conditions under which the action counts as having been successfully performed.

although she also suggest that it is applicable to many joint activities that adults engage in (see pp. 93–94). Her analysis follows the structure of Bratman's account closely, but the participants' intentions are Searlean intentions-in-action (Searle, 1983) rather than Bratmanian plan-intentions.[86] She thus takes something very controversial for granted, namely that an agent can have an intention-in-action with the content "something like 'that we *J*'" (p. 93). In this chapter, I thus argue that what she here takes for granted is indeed possible. In section 3, I present the exclusivity constraint and the own-action condition in more detail. Sections 4 and 5 contain the meat of the chapter. In section 4, I argue that we can have basic intentions-in-action that are "technologically extended", and whose conditions of satisfaction thus range beyond movements of the agent's body. In section 5, I argue that intentions-in-action in an analogous way can be "socially extended", and that, as a consequence, the exclusivity constraint should not be accepted as an unconditional constraint on our intentions-in-action. While I raise some doubts concerning whether Tollefsen's account is a plausible account of young children's participation in joint activities, I argue that it is nevertheless fruitful for understanding joint activities where there is little room for deliberation and planning. Such activities will be ones where the participants are highly skilled at performing the joint activity. Finally, in section 6, I consider and respond to various objections to the idea that intentions-in-action may be socially extended.

4.1 Tollefsen's account

Let us again rehearse Bratman's account of shared intentional activity. According to Bratman, a shared intentional activity is the result of a shared intention, and a shared intention is simply a pattern of "interlocking" plan-intentions of the participants about which they have common knowledge. For you and I to have a shared intention, the following conditions must be fulfilled (Bratman, 1993, p. 106):

[86] Searle (1990) has an account of what he calls "collective intentions-in-action" that is very different from Tollefsen's account. I touch briefly on this account at the end of section 4.4.

We intend to *J* 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.

For our going to see a film at the Cameo together to be a shared intentional activity, our going must be preceded (and appropriately caused) by the intention of each "that we go see a film at the Cameo together". Each must also intend this by way of—"in accordance with and because of"—the other's intention (the intentional agency of each must be involved in our activity). Each of us must also intend that our subplans for carrying out our intentions should mesh. Suppose that I intend that we go see the new Woody Allen film, while you intend that we go see the new Wes Anderson film. If we know that our subplans clash and neither intends that we should try to agree on a film, then we clearly do not have a shared intention. Neither do we have a shared intention if our subplans happen to mesh with each other without us intending that they mesh. This is because a shared intention should fulfil the function of coordinating our activities and plans that are directed toward our common goal. The sociopsychological structure should thus be able to deal with nearby counterfactual cases. Our going to see a film together is not a shared intentional activity if we see a film together at the Cameo as a result of me deceiving you into thinking that the Wes Anderson film is screened in cinema "1" where the Woody Allen film is actually about to start.

As Tollefsen (2005) points out, condition (2) can only be fulfilled if you and I are proficient mindreaders. We must not only be able to intend to mesh our plan-intentions, but each must also be able to recognise that the other intends likewise. However, children do not appear to develop a robust representational "theory of mind" until between 3 to 5 years of age (Wellman et al., 2001). Given that what 1-to-3-year-olds are doing together is appropriately described as joint activity, then this

seems to be of kind of joint activity in which condition (2) is not satisfied. Tollefsen (2005) calls this *the mutual responsiveness problem.*

Arguably, the same problem will be encountered if Bratman's account is applied to spontaneous joint actions that are performed on the fly, either within the framework of a joint activity (such as when a football player suddenly makes a pass play to a team mate within a game) or outside any such framework (such as when two strangers suddenly rush to help steady a third pedestrian that has just tripped and fallen between them). The mutual responsiveness problem arise in such cases not because the participants lack the concepts to have intentions with the kind of contents required by Bratman's account, but because there is arguably no time for the participants to make the attributions of plan-intentions required in light of Bratman's second condition. First of all, mental state attributions are holistic and underdetermined by behaviour, so there are a priori reasons for thinking that reasoning about the mental states of others is too computationally costly and time-consuming to be useful for coordinating spontaneous or improvised joint actions (Apperly, 2011, Chapter 6; Morton, 1996; Zawidzki, 2008). Secondly, empirical research suggests that people do not engage in belief reasoning automatically (Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006; Back & Apperly, 2010). Presumably, this is also true regarding reasoning about other mental states such as plan-intentions.

According to Tollefsen, the mutual responsiveness problem can be avoided if conditions (1) and (2) are reinterpreted so that the intentions they refer to are *intentions-in-action* rather than plan-intentions. 'Intention-in-action' is a technical concept introduced by Searle (1980, 1983).[87] It refers to the Intentional component of an action, which specifies the (typically) detailed and fine-grained conditions that the action's movement component must meet in order for the action to be successful.

[87] Searle simply writes "intention in action", without the dashes. I have added the dashes to emphasise that this is a technical concept of Searle's. Note that I have also inserted dashes wherever Searle is quoted using this concept, so that it always reads "intention-in-action" in this thesis.

These conditions will typically be relatively detailed and fine-grained in comparison with the conditions of satisfaction of a future-directed plan to perform an action, according to Searle: "[I]n any real-life situation the intention-in-action will be much more determinate than the prior intention, it will include not only that the arm goes up but that it goes up in a certain way and at a certain speed, etc." (1983, p. 93) One of the conditions of satisfaction is always that the intention-in-action itself is the cause of the action's movement component. In other words, intentions are causally self-referential. Crucially for the argument in this chapter, the content that one accesses in one's experience of acting is the content of the action's intention-in-action. But according to Searle, one can have an intention-in-action without having a *conscious* experience of acting (1983, pp. 91–92). What Searle has in mind here, I think, is that one may automatically perform an intentional action while one is focused on some other primary task or is daydreaming. For example, while thinking about how to formulate this sentence, I may scratch my head or take a sip of coffee without paying any attention to performing these actions. Typically, intentions-in-action are expressed verbally in the form "I am *A*-ing" or "I am doing *A*" (1983, p. 84). Deliberate actions also have what Searle calls a *prior intention*, but for actions that are not premeditated, this is the whole story. Prior intentions are expressed in the form "I intend to *A*" or "I will *A*" and have more coarse-grained contents than intentions-in-action (1983, p. 93).

Tollefsen (2005) argues that intentions-in-action are perceptually overt by means of various behavioural cues such as "facial expression, extended hands, [or] expressions of cooperativeness." (p. 93) The conditions of satisfaction of an intention-in-action are manifest in such cues, and according to Tollefsen children can "literally see these conditions." (p. 91) If this is true, then an intention-in-action version of condition (2) can be met, even when the participants are children who lack sophisticated mindreading capacities (p. 93).[88] This does not necessarily imply that they, or adult

[88] There is at least ample evidence showing that young children can distinguish animate agency from other movements, as well as that they can recognise what specific goal an agent's behaviour is directed toward (Tomasello et al., 2005; Gergely & Csibra 2003).

human beings, can perceive Searlean intentions-in-action, but the evidence is at least consistent with this hypothesis. Furthermore, Tollefsen is not alone in interpreting the evidence in this way (see e.g. Pacherie, 2000, 2007; Tomasello, Call, & Hare, 2003). I tentatively accept this interpretation of the evidence. In order for one participant's intention-in-action "that we *J*" to be sensitive to the particular content of the other's intention-in-action, namely "that we *J*", they must also be able to recognise that an action is directed toward this particular type of goal. There is evidence that at least adults are able to distinguish between cooperative and non-cooperative goal-directed movements. Becchio et al. (2010) demonstrate that subjects are able to discriminate between actions performed with cooperative intent and actions performed with competitive intent even if the actions are behaviourally very similar and directed toward the same target object. These discriminations appear to be solely based on the kinematics of the observed agent's movement.

However, as I have pointed out, Tollefsen's modification puts the account in a real (rather than merely apparent) conflict with the exclusivity constraint.[89] While agents may be able to *recognise* intentions-in-action with a content "something like 'that we *J*'" (Tollefsen 2005, p. 93), how can they *have* intentions-in-action with such contents?[90] Such contents are ruled out by the exclusivity constraint. While Tollefsen loosely refers to the content as "something like '*that* we *J*'", intentions-in-action are intentions *to* perform an action, not intentions *that* some state of affairs comes about.[91] From now on, I will therefore refer to the content of these intentions-in-

[89] Insofar as one closely links Searlean intentions-in-action with mechanisms of "motor imagery", then Elizabeth Pacherie's proposed account of joint action in (2007) seems to be very similar to Tollefsen's. Pacherie (2000) herself argues that empirical work on motor imagery throws light on Searle's notion of intention-in-action. However, Pacherie does not explicitly claims that participants must have intentions-in-action with a collective content like "that we *J*".

[90] There are two brief references on to the possibility that such contents may be problematic: "Although I think there are troubling issues about this type of intention […]." (Tollefsen, 2005, p. 83 n. 14) And later: "Aside from the difficulties with the notion of an individual intending that we *J* […]." (p. 93). But she does not spell out what these difficulties are, let alone how they could be overcome.

[91] Perhaps this is why Tollefsen qualifies the content specification with "something like".

action as "to *J*" rather than as "that we *J*". At least on some conceptions of what it is to have an intention "that we *J*", this does not amount a substantive change in content. For example, recall that according to Pettit and Schweikard (2006), a participant's group-intention concerns a "joint performance", and this "can be conceptualized just as a pattern of behavior in which our different efforts combine to effect a certain result." (p. 29) The content is thus "conceptualized at the behavioral level" (ibid.). On such a conception of "that we *J*", the content can equally accurately be represented as "to *J*".

4.2 The own-action condition and two constraints on the content of intentions

Nicholas Bardsley refers to what I call the exclusivity constraint as "an uncontroversial constraint [...] that an individual's intentions cannot be said to range over others' actions"[92], and he uses it to rule out analyses of shared (or collective) intention as inadequate (2007, p. 144). This constraint is indeed widely accepted and uncontroversial when it comes to what one can intend *to* perform. While one can intend *that* one's guest leaves before midnight, one cannot intend to perform her act of leaving (Bratman, 1993, pp. 101–102, 2009a, p. 157). Bratman accepts that "what one attempts are [only] one's own actions." (1992, p. 330). In agreement with Bratman, Philip Pettit and David Schweikard state that "[i]t is true that I cannot intend *to* X, where X-ing is a joint performance. But I may still be able to intend *that* we X together." (2006, p. 21) According to Bratman, the contents of what one can intend *to do* are constrained by the own-action condition on something being an intention, according to which "the *subject* of an intending is always the intended *agent* of the intended activity." (2009a, p. 156) Group-intentions, then, should be glossed as "I intend that [we bring about some state of affairs]", rather than as "I intend us to [perform some action]". Accordingly, I take it that Bratman accepts the exclusivity constraint as an unconditional constraint on what one may intend to do.

[92] It is clear from the context that Bardsley is primarily referring to "intention-in-action" here.

Other action theorists also seem to accept the exclusivity constraint as a constraint on what one can intend to do, among them Kutz (2000, pp. 21-22), Tuomela (2005), Brand (1984, pp. 99–100) and Searle (1983, p. 110). Kutz calls group-intentions "non-standard intentions" because they are not tightly linked to actions and "cannot be directly transformed from intentions that P to intentions to P." (p. 21) He even suggests that group-intentions might best be thought of as "either expressing a hope that we will do something, or standing in for an individual intention to promote our doing something." (p. 22) In effect, Kutz excludes the possibility that intentions to P may have content similar to group-intentions, presumably because he accepts the exclusivity constraint. Tuomela (2005) draws a distinction between "aim-intentions and "action-intentions". The intentions that can be directed at a joint activity as a whole are aim-intentions according to Tuomela, and these concern the bringing about of a state of affairs rather than the performance of an action (2005, p. 334). An action-intention concerns the direct performance of an action: it is an intention *to A*. About the conceptual constraints on what one can have an action-intention to do, Tuomela says the following: "If the intention concerns the direct performance of an action (e.g. when an agent intends to open the window) the agent must himself bring about the satisfaction solely by his own action." (2005, p. 329) While Tuomela sometimes talks about what he calls a "we-intention-in-action", the object of these intentions-in-action are not the whole joint performance but only the individual's contributing part of that performance (see Tuomela 2007, p.109).

Brand claims that it is "impossible in a logical or conceptual sense" that an agent intends *for another agent* to perform *an action* (1984, p. 99). While he concedes that an agent could directly move another agent's body, Brand claims that this "situation would be like one in which a mad physiologist stimulates [the other agent]'s brain so that his limbs move. [The other agent] would not be performing an action; rather something would be happening to him." (1984, p. 100) Brand thus fails to see the possibility that an agent might directly intend to move another agent's body even if this bodily movement is part of the other agent's action.

Put in terms of Searle's notion of intention-in-action, the own-action condition says that the content of one's intention-in-action can only concern the event that, together with the intention-in-action itself, constitutes one's action. This event of one's own action is typically assumed to be one's bodily movement. It is therefore helpful to introduce another possible constraint on the content of intentions-in-action, which I will call *the own-body constraint*. According to this constraint, the content of an agent's intentions-in-action only concerns the agent's own bodily movements. I will argue later that if the own-action condition is taken to imply the own-body constraint, then the condition will often be breached in cases where humans perform actions with tools.

Searle clearly embraces the exclusivity constraint in some form. He claims that if there is "intervening Intentionality" in the causal chain between the tokening of an intention-in-action and the event that is the intention's outcome, then that action cannot be successful. Searle illustrates this with the following example:

> Thus, suppose that unknown to me my arm is rigged up so that whenever I try to raise it, somebody else causes it to go up, then the action is his not mine, even though I had the intention-in-action of raising my arm and in some sense that intention caused my arm to go up. […] And that this is the right way to construe intentions-in-action is at least indicated by the fact that, when my intentions-in-action make explicit reference to the intentions of other agents, then in general the actions become the actions of those agents. Thus, suppose I know how my arm is rigged up and I want it to go up. My intention-in-action then is *getting the other agent to raise it*, not *raising it*. *My* action is getting him to raise it, *his* is raising it. (1983, p. 110)

The idea, then, is the following: In general, if I have an intention-in-action that includes in its conditions of satisfaction that it causes you to have an intention-in-action that in turn causes an action *A*, then *A* is performed by you. This excludes me from also performing *A*. The action *I* perform is instead getting you to perform *A*. Searle thinks that this line of reasoning remains valid even when I am unaware of the mediating role of your intention-in-action. However, it is clearly not valid in all such cases. We can see this by considering one of Searle's own examples (see 1983, p. 98): Gavrilo Princip's blow against Austria and vengeance of Serbia (performed by means of his killing of the Archduke Franz Ferdinand in Sarajevo on June 28, 1914).

Insofar as "striking a blow against Austria" is an action at all, it is Gavrilo's intentional action in spite of the fact that most of the links in the causal chain from intention-in-action to massive societal effects must have consisted of "intervening Intentionality". Perhaps Gavrilo was in some sense aware of this intervening Intentionality, at least upon reflection. In that case, Searle's line of reasoning is not even valid for some cases where an agent is aware of the mediating role of others' intentions in the causal chain between intention-in-action and outcome. Given Searle's choice of example in the quote above (raising one's arm), it is reasonable to take the exclusivity constraint to hold only, or only hold unconditionally at least, in cases where *A* is a *basic action*. A basic action is an action *A* that an agent can intend to perform "without intending to do another action by means of which he intends to do *A*." (Searle 1983, p. 100) In most circumstances, squeezing the trigger of a gun is a basic action, since one does not do this by means of intending to contract one's muscles in the arm and hand in such a way that the trigger gets pulled. This is unlike the case when Gavrilo intends to strike a blow against Austria. He intends to do this by means of firing a gun against the Archduke. In the case of basic actions, it is hard not to accept the exclusivity constraint as a fundamental constraint. But I will argue in the next section that the constraint does not always hold, even if its scope is restricted to intentions-in-action of basic actions.

To sum up, Tollefsen's modification of Bratman's account leads to a problem. By relying on the notion of intentions-in-action rather than plan-intentions, Tollefsen saws off the branch that allows the original account to avoid violating the exclusivity constraint. However, Tollefsen does not explain how individuals can have intentions-in-action of the form "I am intending-in-action to *J*'". I will now explain how this is possible.

4.3 Technologically extended intentions-in-action

My strategy will be to first argue that the conditions of satisfaction of the intention-in-action of a basic action are not always restricted to only concern an agent's own bodily movements. Human agents can have what I call basic but *technologically*

extended intentions-in-action. (An intention-in-action is "basic" if it is the intention-in-action of a basic action.) After establishing this, I argue in the next section that basic intentions-in-action can also, contrary to the exclusivity constraint, be *socially extended.* In effect, I am arguing that what may make up "one's own actions" is much less restricted than action theorists tend to assume. The own-action condition implies neither the own-body constraint, nor the exclusivity constraint. The reach of an agent's basic actions is not necessarily limited by the agent's body surface, nor by another agent's co-ownership of the action's physical movement or event.

As far as I can see, there is nothing in Searle's characterisation of intention-in-action as such that restricts the conditions of satisfaction to range only over the agent's bodily movements. I here seem to be in disagreement with Joëlle Proust (2003) who interprets Searle as embracing the own-body constraint. She claims that "an intention-in-action fails to represent any sort of further goal, such as 'switching on the light,' 'breaking a vase,' and so on. What it determines is, rather, a bodily movement." (p. 106) It is true that Searle frequently writes that an intention-in-action causes (and presents itself as causing) "a bodily movement", but I think this is a reflection of the choice of examples that he illustrates his theory with, such as raising one's arm, rather than the expression of a general constraint that is supposed to be part of his theory. Indeed, he says explicitly that we can make "intentional bodily movements where the conditions of satisfaction of our intentions go beyond the bodily movements." (1983, p. 99) However, Searle calls the intentions of such actions "complex intentions" (1983, p. 98). He characterises the self-reflexive content of the intention-in-action of a man who fires a gun in the following way: "This intention-in-action causes it to be the case that the trigger pulls, which causes it to be the case that the gun fires." (1990, p. 409) While the extended "reach" of the action is to be found in the content of the intention, this reach is represented *as* extended, as pertaining to events beyond the agent's immediate control. But there need be nothing complex (in terms of content) about intentions-in-action that reach beyond movements available to the agent considered as an unaided body.

Consider the well-known example of the blind man who explores his surroundings by moving and tapping a white cane in front of him (Merleau-Ponty, [1945] 2002, pp. 165–166). What, we can ask, is the blind man's 'experience of acting'? His experience is not that of manipulating his arm, wrist and fingers in order to cause the cane to move in certain ways, as if he was holding a cane for the first time in his life. Rather, the blind man simply taps the ground directly with the tip of the cane. The cane itself has faded out of his awareness and attention, just like our arms and hands are outside the focus of awareness and attention when we "use them" in action. We do not intend to move our joints, wrist and fingers in order to cause our hand to reach out and grasp an object, instead we simply reach out and grasp it directly. Similar things can be said about the perception involved in the skilled use of tools. The blind man's perceptual experience is that of touching the ground at the tip of the cane: he touches the ground *through* the cane. The experience is not that of touching the cane in his hand in order to *infer* what the ground is like. Such extension of our capacities for perception and action is typical of fluent use of tools, where the tools become "transparent" to the user. Following Andy Clark (2008, p. 31), I think that the best picture of the agency involved in such cases is that "of an extended or enhanced agent confronting the (wider) world" rather than that of a bare biological agent facing a tool.

Interestingly, neuroscientific findings suggest a subpersonal basis for the shift from phenomenological opaqueness to transparency. Maravita and Iriki (2004) argue that an agent's so-called body schema – a neural representation of the agent's body shape and posture – changes as the agent becomes fluent in using a tool to perform a task. This results in an extension of the agent's peri-personal space (its proximal action space). Bimodal neurons in the body schema of Japanese macaques normally only fire when a macaque touches something with its hand or sees something near its hand (hence *bi*modal). But after a macaque has learned to use a rake to reach for and collect food, these bimodal neurons also fire when the macaque touches something with the rake or sees something in the space around the rake (Maravita & Iriki 2004). Behavioural experiments suggest that such an extension of peri-personal space is also an effect of tool use in adult human subjects. In some neuropsychological

patients, so-called "visual-tactile extinction" (see di Pellegrino, Làdavas, & Farné, 1997) has been taken to indicate the boundary of peri-personal (or peri-hand) space. If a single tactile stimulus is presented to such a patient on their contralesional hand,[93] which is hidden from view on a table in front of them, then they have no problem detecting the stimulus. The same thing happens if they are simultaneously presented with a distracting visual stimulus on the ipsilesional side of their body, but not close to the ipsilesional hand. However, if the visual stimulus is presented immediately adjacent to the ipsilesional hand, then their performance in detecting the tactile stimulus drops significantly: the tactile stimulus is extinguished. Farné et al. (2007) studied one such patient's susceptibility to visual-tactile extinction before and after having used a rake to collect objects placed on a table. After tool use the space in which distracting visual stimuli resulted in extinction of the tactile stimuli was no longer restricted to the space immediately adjacent to the patient's ipsilesional hand, but extended into space where the rake's axis and tip had been.

While not conclusive, these parallel results suggest that extension of visual-tactile extinction is due to an extension of the patient's peri-personal space. This extension is induced by active tool use and encoded in the body schema (see Holmes et al. 2007 for some doubts). Thanks to the plasticity of the body schema, tools become incorporated into the suit of resources at hand that the cognitive system simply takes for granted (see Clark 2008, sect. 2.5). On this interpretation, these taken-for-granted resources define the agent's proximal action space. It should be noted though, that some behavioural experiments on adult human subjects are in tension with this extended body schema hypothesis (see Holmes, Calvert, & Spence, 2004; Holmes, Sanabria, Calvert, & Spence, 2007).[94]

---

[93] The contralesional hand is the left (right) hand if the lesion is in the right (left) hemisphere. The other hand is the ipsilesional hand.

[94] For example, Holmes et al. (2007) studied the effect of tool use on subjects' perception and action capabilities by comparing under what conditions and to what degree a visual distractor stimulus induced so-called "visual-tactile interaction". Visual-tactile interaction was measured by the effect that the visual distractor (one or two flashes) had on a subject's ability to discriminate between two tactile target stimuli (one or two vibrations felt through a stick). They found that visual-tactile interaction was more marked when (i) subjects held a single stick in their right hand

My argument is not hostage to any particular interpretation of these experimental results. But if the extended body schema hypothesis were confirmed, then it would demonstrate an interesting isomorphism between the personal-level phenomenon of "transparency in use" and its subpersonal underpinning. Personal-level phenomenology suggests that tool-using actions are often basic actions, just like bodily actions such as raising one's arm, and if the extended body schema hypothesis is correct, then this may be correlated with the tool's incorporation into the subpersonal representation of the agent's bodily action capacities. However, the phenomenological observations about tool use would not be undermined if this particular hypothesis were rejected. Within the framework of Searle's theory of intention and action at least, the phenomenology of action should be taken seriously. After all, the content of the intention-in-action is the content that is accessed when one has the experience of acting.

Searle makes some observations that fit this take on transparent tool use. First, he notes that what counts as a basic action is relative to an agent's skills (1983, p. 100). Hence, a skilled tool user can act on the world through the tool without having to do this by means of intending to manipulate the tool. Secondly, Searle does not want to restrict the kinds of causal chains that can be involved in the execution of an intention-in-action to those occurring within the body (see 1983, p. 110). If I have an intention-in-action to raise my arm that causes my arm to raise, then this counts as my intention-in-action being carried out successfully, even if someone, unbeknownst to me, has rewired my brain so that the motor commands from my brain go half way

---

to discriminate target stimuli presented on the same (right) side, compared to when (ii) the subjects held the stick in their right hand to discriminate target stimuli presented on the other (left) side (so that the axis of the stick crossed the body's midline), and compared to when (iii) subjects either held a single stick in their right hand or held one stick in each hand to alternate between discriminating target stimuli presented on their right side and on their left side. Holmes et al. (2007) suggest that these results can be explained by appeal to a general effect that maintained spatial attention toward one side of the body has on visual-tactile interaction on that side (it becomes more likely). If the body schema encodes an extended peri-personal or peri-had space after tool use, then it is not clear why the degree of visual-tactile interaction should vary between these experimental conditions.

around the world via the internet before reaching the effectors. In effect, there is nothing in Searle's account that precludes that the causal chain between an intention-in-action and the movement it represents extends beyond the agent's body. Besides Searle's habit of calling the intentions-in-action of tool-using actions "complex", there is nothing in his account that precludes the possibility of basic but technologically extended intentions-in-action.

I conclude that the contents of an agent's basic intentions-in-action are not constrained by the own-body constraint. I do not claim that the own-body constraint on the content of basic actions is explicitly accepted by many action theorists, but the idea that basic actions are bodily movements seems to be a tacit and unexamined assumption in much philosophy of action. For example, Davidson claims "that all primitive actions are bodily movements", where 'bodily movements' is "openhanded enough to encompass such 'movements' as standing fast, and mental acts like deciding and computing." (1980, p. 49) This assumption may be one source of resistance among action theorists to the possibility of intentions-in-action "to *J*". However, at least within the framework of Searle's theory of intention and action, the assumption should be rejected.[95]

Since the phenomenon of 'transparency in use' has been used to argue for the view that the vehicles of the contents of perception and action extend beyond the biological boundary of the human (or macaque) organism (Clark, 2008, Chapter 2; see Rupert, 2009, Chapter 8 for a critique of this kind of argument), it is worth pointing out that widening what counts as permissible contents of an agent's basic intentions-in-action does not depend on accepting such "vehicle externalism" (Hurley, 1998). The question of whether or not to accept the own-body constraint is a matter of what content it makes sense for an intention-in-action to have, given the kind of mental state that an intention-in-action is. It is not a matter of whether the vehicle of the intention-in-action may extend beyond the body surface of the agent. Neither is it a matter of whether the environment of the agent plays a role in fixing

[95] For a thorough discussion about basic actions and the use of tools, see Pols (2011, Chapter 3).

the content of its intentions-in-action. In other words, rejecting the own-body constraint does not depend on accepting either vehicle externalism or content externalism. The same can be said with regard to rejecting the exclusivity constraint.

4.4 Socially extended intentions-in-action

In order to reject the exclusivity constraint as an unconditional constraint on what one can intend to do, I will appeal to considerations similar to those that make rejection of the own-body constraint plausible. In other words, I will argue that intentions-in-action can be socially extended, not only technologically extended.

What motivated my notion of basic but technologically extended intentions-in-action was the phenomenon of transparency in skilled tool use. Do people experience a similar shift from opaqueness to transparency with increased participation in a joint activity? I think this sometimes happens. Consider the example of two skilled ice skaters performing a figure dance. From an adopted first-person perspective, Axel Seemann (2009b) characterises their sense of joint control in the following way:

> [I]t isn't that *I* experience myself as being in a position to determine *your* doings by my actions. It is, rather that the experience really is one of *us* controlling our doings. And this experience is an embodied one: in the most obvious kind of case, such as the example of the figure dancers, our bodies really seem to form an experiential unit. The awareness I enjoy of your body in a dance is quite unlike the sensation of your flesh pressing against mine that I might be exposed to in a crowded subway carriage. Your body seems to form part of our joint interface with the world. (2009a, p. 508)

To my ears, Seemann's observation here rings true. Like in the case of skilled tool use, agents sometimes experience each other as transparent extensions of their own activity. The coordination of actions recedes to the background of attention and the agents experience themselves acting as one attending to their joint activity. The intentions-in-action here are basic, they are not executed via the execution of other intentions, whether one's own or those of one's co-agent.

My claim here is not that the whole activity – the entire dance on the ice for example – is part of the content of the participants' basic intentions-in-action, but rather a move or turn within that activity. The following joint move may be part of each intention's content: As one of the dancing ice skaters traces a curve on the ice, he lifts his partner up into a position where she has one leg held above her head and the other leg stretched out parallel to the ice (a so-called full Biellmann position). Here, the intention-in-action of each of the skaters will be dependent on the intention-in-action of the other, and the move as whole will be a basic action for each of them. The bodily movements of their partner will be part of the conditions of satisfaction of the intentions-in-action of each of them. Furthermore, the self-reflexive condition that an intention-in-action causes the movement or event that is specified in its content is satisfied for both of them. The intention-in-action of each skater causes not just their own bodily movements, but also the bodily movements of the other. After all, in the absence of their intention-in-action, the movements of the other would not be performed.

I have argued that the notion of 'socially extended intentions-in-action' is intelligible. It is also a useful notion, since it captures the kind of intentions-in-action that participants engaged in some skilful joint activities have, such as figure skaters or dancers. One mark of such joint activity would be that the participants have socially extended intentions-in-actions that overlap. This is one way in which we might understand the idea that intentions-in-action could have the content "to *J*" that is part of Tollefsen's account. It is also an alternative way of elaborating Searle's analysis of intention-in-action in *Intentionality* (1983) to deal with joint activity, which is different from Searle's (1990) own account of what he calls "collective intentions-in-action" (p. 410).

Searle thinks that when agents are acting jointly, each agent has a special kind of mental attitude, a "collective intention-in-action", that has the *form* "We are *A*-ing". It is thus different from a group-intention, which introduces the collectivity (the "we") into an account of joint action via its *content*. Searle's "collective intention-in-action" does not *represent* our *A*-ing as a joint activity; our *A*-ing is a joint activity

because each of us has a "we-mode" intention-in-action to *A*. Interestingly, Bratman (2009a, pp. 156–157) suggests that the reason Searle makes the move of positing this new kind of attitude is precisely that he endorses the own-action condition (and, I would add, that he accepts the exclusivity constraint). If the own-action condition is constitutive of what it is for a propositional attitude to be an ordinary "I-mode" intention, then the attitude that is characteristic of joint activity cannot be such an intention, but must be some other kind of intention-like attitude. However, like Bratman and others, I do not endorse the own-action condition for plan-intentions, nor do I endorse the exclusivity constraint for intentions-in-action. I thus see little reason to posit such a new intention-like attitude. I will hence not discuss Searle's own view of joint action further (for further criticism of Searle with which I agree, see Pacherie, 2007).

4.5 Objections and rejoinders

One initial objection to socially extended intentions-in-action might be the following. The sources of knowledge about our own and others' actions are different in important ways. We have no proprioceptive information about the bodily posture or movements of others. To the extent that such information is required for carrying out intentions-in-action and for monitoring the dynamic unfolding of action, there is reason to doubt that intentions-in-action can be socially extended. But no such proprioceptive information is available in tool use either, so such information does not seem to set a limit to what can part of the content of basic intentions-in-action.[96]

However, evolutionary considerations suggest that the boundary of the biological organism has an epistemic importance which at least makes it unlikely that extended intentions-in-action are prevalent. Even if vision is important for awareness and knowledge of our own actions, it has been argued that proprioceptive information

[96] Indeed, in circumstances where proprioceptive and visual information about our own actions are in conflict, our conscious reports about our own bodily movements seem to rely more on vision than proprioception, as for example the 'rubber hand illusion' demonstrates (Botvinick & Cohen, 1998).

has a kind of privileged status in our subpersonal cognitive machinery that is not enjoyed by perceptual information about the extra-bodily environment. In arguing against the hypothesis of extended cognition (or vehicle externalism), Kim Sterelny (2004) draws attention to the fact that an organism is a conglomeration of parts that have evolved together and thus become co-adapted to each other over time. Evolution has ensured that the internal "environment" of the organism is a friendly one: Information exchange is reliable, trustworthy and efficient. "[O]ver evolutionary time", as Sterelny (2004) puts it, "the internal informational environment of an agent will become more transparent." In contrast, the external environment is a "shared and sometimes contested space" inhabited by other agents who are potentially out to deceive and manipulate the organism. Hence, while there is selection pressure on perceptual systems to become more reliable and efficient, it is unlikely that external resources, such as tools, public information, or (I would add) other agents become as transparent to us as our internal resources. This is because perceptual systems pick up information from an epistemically hostile environment. But as Clark (2008, p. 103) points out in reply to Sterelny, these evolutionary considerations do not show that technological extensions of cognition are impossible, just that we should be sceptical about their prevalence.

Given that Sterelny is correct in characterising internal information flow as trustworthy and reliable, and external information flow as always potentially noisy and deceptive, technologically extended intentions-in-action ought to be quite rare, and arguably, socially extended intentions-in-action ought to be even more rare. The possibility of defection and deception is much greater in the case of other agents than in the case of tools. The objection, then, is that a socially extended intention-in-action will depend on the absence of a kind of vigilance and double-checking that we should expect to be the norm in interaction with other agents. However, like in the case of technological extension, these considerations do not show that socially extended intentions-in-action are impossible, just that that we should be sceptical about their prevalence. In addition, the epistemic difference between the internal and the external environment should not be exaggerated.[97] If, as Sterelny himself

[97] In a more recent paper, Sterelny (2010) comments on his earlier argument against

believes (see 2003), group selection has played an important role in the evolution of our pervasive tendency to cooperate, then human groups are in a sense made up of co-adapted parts that have evolved together.[98] Furthermore, my argument crucially depends on the phenomenology of participation in joint activity. In cases where an agent experiences his or her participation in a joint activity as transparent, then that should be reflected in the content of his intention-in-action. I am assuming here that Searle is right to identify an agent's experience of acting with the content of the agent's intention-in-action: The content of one's experience of acting just *is* the content of one's intention-in-action.

However, perhaps the experience associated with a socially extended intention-in-action is always mistaken or illusory. One might grant that a basic intention-in-action may have conditions of satisfaction that makes it socially extended, but argue that, in fact, the self-reflexive condition that the intention-in-action causes a jointly performed movement or event will never be satisfied. This would account for the phenomenological observations that I have been swayed by, but would arguably make socially extended intentions-in-action less interesting since they would not be appropriately causally connected to the (joint) activity. But it is not clear why one should think that the self-reflexive condition couldn't sometimes be satisfied. After all, an intention-in-action is just supposed to be one in a host of necessary factors that bring a movement or event about. The agent's body and the world must cooperate for the right things to happen even in the case of ordinary solo action.

---

the extended cognition hypothesis. He submits that his "initial contrast [between the internal and external environment] was overdrawn" and that his "earlier ideas on the importance of contested space were overinfluenced by Machiavellian models of social interaction." (p. 474)

[98] Less controversially, infants and toddlers are profoundly dependent on their caregivers and the group into which they are born. Even if there are cognitive mechanisms in older children and adults with the function of double-checking and vetting during joint activity, it does not seem plausible that such mechanisms are operating in young children, at least not during interaction with close kin or other familiar individuals.

One could object that socially extended intentions-in-action cannot have a *joint* or *shared* activity as its content. I have argued that the bounds of what counts as one's own basic actions are wider than action theorists have thought, but this does not make those actions joint or shared. Hence, the objector continues, I have not here gone to the heart of what is philosophically puzzling about shared intention, namely how the question of whether to do this or that can be a matter for us to settle together. As David Velleman (1997) points out, I can only intend to do something that is up to me to settle. But if it is up to me to settle whether or not to do something, then it cannot simultaneously be up to you. The puzzle is how an agent can exercise control over something while at the same time delegating this control to someone else.

What I am proposing here is not a solution to this puzzle. I am not making a proposal about how agents can share a future-directed intention or decision to engage in a joint activity, I am rather trying to suggest what characterises some joint activities, irrespectively of whether the participants jointly chose to enter the activity. I take it that some joint activities that are characterised by what Bratman calls "mutual responsiveness in action", but not necessarily by "mutual responsiveness of intentions" (see 1992, p. 339). What is joint is control and guidance of the unfolding of the joint activity, not a choice about whether or not to initiate it. As Bratman notes one's "plans will typically be at a level of abstraction appropriate to [one's] habits and skills." (p. 31) In many habitual and entrenched joint activities, this level of abstraction will arguably be found at a level that is above the level at which mutually responsive motor action occurs. This does not merely involve skilful joint manoeuvres like those performed by the dancing ice skaters, but also mundane joint actions that are performed habitually, such as when to people meet and shake hands or do a "high five".

That it is possible for participants' intentions-in-action to be socially extended does not, of course, establish that the intentions-in-action of infants and toddlers can be socially extended. Indeed, infants and toddlers do not appear to be in a good position to establish the kind of fluent expertise that makes coordination during joint activity

transparent. For this reason, I doubt that an account based on socially extended intentions-in-action, which I suggest that Tollefsen's account is, is applicable to children's joint activities *in general*, including cases such as two toddlers building a block tower together or engaging each other in social pretend play. But perhaps the conditions needed for two children, or a child and a caregiver, to have overlapping socially extended intentions-in-action do obtain in some simple forms of social play that involve repetition and predictable turn taking. This seems especially plausible in asymmetrical joint activities where a child is playing with an adult, since an adult co-participant is typically more disposed to be controlled and "manipulated" by the child than a peer will be (the adult can thus make up for the child's lack of fluent expertise).[99] Given the role of phenomenology in my argument for the possibility of socially extended intentions-in-action, and the difficulty of accessing what infants and toddlers experience, it is hard to draw any firm conclusion about what the role, if any, socially extended intentions-in-action might play in young children's participation in joint activities. The conclusion that such intentions-in-action have a role to play in some joint activities with adult participants, however, is much more secure.

Even with adult participants, the kind of sustained experience of fluency described by Seemann is the exception rather than the rule. As Kutz (2000) points out, "the complexity arising from problems of coordination renders collective activity salient, making it stand out against a background of unreflective, self-regarding activity." (p. 12) In line with my reflections on Sterelny's earlier argument against extended cognition, he also notes that the possibility of free-riding makes joint activity unlikely to become automatised and unreflective (p. 13). However, Kutz grants, in line with what I have been saying, that "well-rehearsed joint action may require no conscious deliberation or reflection." (p. 12) Indeed, in the case of some joint actions, such as when the ice dancers perform their well-rehearsed joint move,

[99] In a footnote, Tollefsen raises the intriguing possibility that, in light of young children's profound dependency on their caregivers, their intentions-in-action are actually socially extended as a rule rather than as an exception (2005, p. 93 n. 24).

conscious deliberation and reflection will not even be possible given the speed and timing requirements that must be met to avoid failure (and most likely, injury).

4.6 Conclusions

I have pointed out that couching Bratman's account of shared intention in terms of participants' intentions-in-action in order to avoid the mutual responsiveness problem, as Tollefsen (2005) does, is problematic. Such an account requires participants to have intentions-in-action with the content "to *J*", but unlike plan-like intentions "*that* we *J*", such intentions really do violate the exclusivity constraint. According to this constraint, one's intentions-in-action cannot range over the actions of another agent. I have argued that this constraint should not be accepted as an unconditional constraint on the possible contents of intentions-in-action. There is nothing in Searle's analysis of intention-in-action per se that precludes the contents of one's basic intentions-in-action to range beyond one's bodily movements, and the phenomenon of transparency during skilful tool use indicates that intentions-in-action often actually do have such wide-ranging contents. Thus, the own-action condition does not imply what I have called the own-body constraint. By analogy with such technologically extended intentions-in-action, phenomenological considerations also suggest that intentions-in-action can be socially extended. This means that the bounds of what counts as one's own actions are wider than action theorists typically assume. Neither technologically extended, nor socially extended intentions-in-action, fail to fulfil the own-action condition. As a result, the exclusivity constraint should be rejected. Two agents can have socially extended intentions-in-action with overlapping contents, so that they both guide and monitor the unfolding of their joint activity.

While I have expressed doubt regarding Tollefsen's account as a plausible account of *young children's* joint activities, I believe the account fills a lacuna in the philosophy of joint action. Bratman (1992) mentions that "mutual responsiveness in action" is a defining feature of joint activity, but he does not have much to say about this feature. Tollefsen's account – supported by my notion of socially extended intentions-in-

action – specifies an interpersonal pattern of mental states which I suggest is characteristic of some skilful or habitual joint activities or actions that involve tightly coupled mutual responsiveness in action. However, a potential objector might say the following about my application of Tollefsen's account to the skilful or habitual joint activities of planning agents such as adult human beings: "What you have provided is merely a notational variant of Bratman's account. The account captures a special limiting case that Bratman's account of shared intention is applicable to, namely the case where the plan-intentions mentioned in conditions (1) and (2) are present-directed rather than future-directed. The intentions in Bratman's first condition is intentions 'that we *J now*' rather than in 'that we *J* [at some future time *t*]'. You have been misled to think that Bratman's planning conception of intention requires that the intentions in his account must concern the future, but this is a mistake." Given Bratman's expansive notion of planning, this objection goes, even spontaneous, habitual or improvised joint actions are in some sense planned. In the next and final chapter, I show why such an expansive notion of planning should be resisted. Furthermore, even if it is accepted, there are good reasons for thinking that the intentionality of both individual and joint action cannot be reduced to the intentionality of intentions.

5 Action coordination without or beyond planning

Examples of joint activities in the philosophical literature include: lifting a heavy sofa together, travelling to Chicago together, playing chess, dancing a tango, painting a house together, singing a duet, playing basketball, walking together, washing the dishes together, executing a pass play in football, preparing a hollandaise sauce together, performing an outdoor ballet, and having a conversation (Bratman, 1999; Gilbert, 1990; Kutz, 2000; Searle, 1990; Tollefsen, 2005; Tuomela & Miller, 1988; Velleman, 1997). When they occur, many of these activities will typically be the result of advance planning. A basketball game will probably take place because the players planned to play at a certain time, one of them booked a court, and so on. Most travellers who go to Chicago will have planned the trip long in advance, and two people who are painting their house together will typically deliberate about which colour to pick days or even weeks before the day on which the painting commences. However, some of these activities may also be initiated spontaneously, in response to the situation of the here and now (it is difficult to imagine how the performance of an outdoor ballet and the joint painting of a house could be initiated spontaneously though!). For example, we might fall into a conversation about we did last weekend when we by chance meet each other on the street. Our joint lifting of a sofa might be performed in response the following situation: We are about to watch a film on television, and we both realise that wall-mounted flat screen is a bit too far away from the sofa to make for a good viewing experience. As you grab hold of one end of the sofa, you glance at me and mutter, “we better move this”. I immediately understand what should be done, and I grab hold of the other end, and we together move the sofa closer to the screen. This joint action would not be the outcome of planning in any ordinary sense.

Within the envelope of a joint activity that is the outcome of planning, participants will often have to perform joint actions spontaneously and in an improvised manner. Consider the pass I play to you during the second half of a football game that we both participate in. Due to the complexity and unpredictability of the game, it would

be useless for me to try to make a plan-like commitment to make the pass play, since it will be impossible to predict in when and under what circumstances the right opportunity will arise. To reliably achieve such a coordinated action, it will arguably not be sufficient that we have meshing plan-intentions to perform the pass play. Or consider the example of you and I washing dishes together. Within this activity, there will be joint actions that have to be performed with spontaneity and in an improvised fashion. If you hand me a glass that you have just rinsed so that I can dry it with a cloth, then my grasping of the glass and your release of it will have to be coordinated. Suppose that when you hand me a heavier object, such as a saucepan that you have just rinsed, you discover that I'm not able to hold the pan steady as you release your grip on the pan. If this would happen, you would probably be able to quickly return your hand to the pan and help me to steadily put it down. What is required for such feats of coordinated action is not only that the performance of each of our actions are mutually responsive to the other's action performance, but also that these performances are appropriately responsive to our common goal as well as to features of the environment (such as the shape, location and weight of the glass and the saucepan). Coordination that goes beyond deliberating and settling on plan-like commitments will thus often be required even if the joint activity as a whole is the outcome of advance planning. As Bratman himself points out, shared intentional activities are characterised not only by mutual responsiveness in intention, but also by "mutual responsiveness in action" (1992, p. 339).

The overall focus of this chapter is on providing some answers to the following question: How do participants successfully coordinate their actions in a joint activity in the absence of plans or beyond what their plans specify? The focus is on the second lacuna that I highlighted in chapter 1, namely how participants achieve mutual responsiveness in action. What I provide here is of course a very selective overview of research that is relevant to my question. It should not be taken as anything like an exhaustive review.[100]

[100] The empirical work I review here is highly selective. For a comprehensive overview of the empirical work on joint action, see Knoblich, Butterfill and Sebanz (2011).

The structure of the chapter is the following. I the next section, I consider what "planning" and "planned action" are in light of Bratman's planning theory of intention and intentional action. While I criticise Bratman for employing a too expansive notion of planning that obscures the difference between future-directed intentions (plans) and present-directed intentions (or intentions-in-action), I grant for the sake of this chapter's argument that there are present-directed plan-intentions and that that all intentional actions (including spontaneous ones) are the outcome of some kind of plan-intention. In the remaining two parts of the chapter (sections 5.2 and 5.3) I then aim to demonstrate—by reviewing empirical work on motor control and coordination in both individual and joint action—that even if we would grant that all intentional actions involve plan-intentions, the purposiveness and intentionality of those actions cannot be reduced to the contents of plan-intentions. When it comes to individual action, I argue for this claim in section 5.2.1, where I review empirical work on motor planning and motor control. A similar argument for joint action is made in section 5.3.1, where I present some empirical work on planning and control mechanisms for joint action. In sections 5.2.2 and 5.3.4, I consider the role of self-organisation in bodily coordination and control, within the individual (5.2.2) and across individuals (5.3.2). Considered together, the empirical work reviewed here shows that what Risjord calls "the standard picture" of joint activity is wrong. According to this picture, joint activity is imbued with joint intentionality only in virtue of a shared intention that distributes tasks and roles between the participants (see section 1.3.2).

5.1 Planning and action control

When we say that we planned to perform an action *F*, we normally mean that we decided to perform *F* before the time of action arrives. We commit and prepare in advance for carrying out the action. However, when I here write of an action *F* that it was planned, what I mean is simply that the action is the outcome of a plan-intention, that is, a plan-like commitment to act.[101] This notion of planned action is clearly not

[101] This will typically be a plan-like commitment to *F*, but it need not be. Bratman

an everyday notion. While Bratman thinks that the features that are constitutive of the mental state of intention are most perspicuous in light of the role that future-directed intentions plays in our agency, plan-like commitments to act, or plan-intentions for short, may also be present-directed. Such a present-directed intention to *F*, furthermore, can be formed or acquired without having been preceded by a future-directed intention to *F*. Suppose that I am queuing at the checkout line in a supermarket and I suddenly spot a basket beside me that is full of my favourite kind of chocolate bars. If I just decide to buy one without first deliberating about it and immediately reach out and grab it, then this will be an action of mine that is guided by a present-directed intention, a plan-like commitment to *F now*. What makes the action intentional is, according to Bratman, the plan-intention that caused the action (in the appropriate way).[102] Clearly, the notion of planning at work here is much broader than the folk notion.

According to an everyday notion of planning, what I plan to buy in the grocery store are the items that appear on my shopping list, rather than the chocolate bar that I spontaneously decide to while waiting in the checkout line. Planned actions are actions that one has settled on performing in advance. Of course, as I reach out to grasp and put the chocolate bar in my shopping basket, I represent the goal that my action is directed toward, and this representing occurs in advance of the represented outcome actually being brought about, just not very far in advance. If representing a goal in this way counts as having plan, then we should say that all actions that are the outcome of an intention-in-action are planned, even if the intention-in-action and the

---

argues that for an agent to intentionally *F*, the agent must either have intended to *F* or have intended to *G*, where *F* is in what Bratman calls *G's* "motivational potential" (1984, 1987, Chapter 8). Bratman does not give a specification of the conditions under which *F* is in *G's* motivational potential, but his view is opposed to what he calls "The Simple View", according to which an agent intentionally *Fs* only if the *F*-ing is an outcome of an intention to *F*.

[102] This explication of intentional action in terms of action caused by a plan is by no means an idiosyncratic feature of Bratman's theory, but is also characteristic of theories of action proposed by, for example, Alvin Goldman, Myles Brand and Alfred Mele (see Preston, 2013, Chapter 2).

bodily movement was not preceded by a prior intention. On Searle's view of action, this would mean that all actions are planned actions.

However, I think there are good reasons for distinguishing between present-directed intentions (intentions-in-action) on the one hand, and future-directed intentions (plan-intentions, prior-intentions) on the other hand. As David Velleman (1991, 2007) points out, the functional role and the norms that Bratman identify the mental state of intention with make the idea of intentions that are present-directed problematic. With the exception of the demand that intentions should control conduct, the norms of rational planning that Bratman takes to be constitutive of intentions seems to be either inapplicable to present-directed intentions or to be superfluous for the purpose of acting in the present. The demand for means-end coherence is not applicable since there is no room for planning or deliberation (I do not consider various means for retrieving the chocolate bar before I settle on reaching for it and pick it up). On the other hand, the demand for consistency is superfluous since one will not—if one has been a rational planner—have any intentions that conflict with the present-directed intention. If one had, then the present-directed intention could not have formed in a rational fashion without deliberation and reconsideration. If I was on a diet and had settled on not buying any chocolate prior to entering the supermarket, then I would presumably not have formed a present-directed intention to get the chocolate bar (off course, I could have given in to temptation and bought the chocolate bar anyway, but in that case, I would have bought the chocolate bar despite my resolution not to do so, rather than because I reconsidered and rescinded my resolution). But in the absence of any such conflicting intentions, the mental state that guides my action of getting the chocolate bar can play no role in rationally coordinating my action with my intentions. The reason that the role and norms of intentions are most perspicuous in the case of future-directed intentions may thus be that these are the only intentions that exist (given Bratman's conception of intention). Arguably, Searle's distinction between prior intentions and intentions-in-action has the advantage of not lumping together what appear to be two kinds of states that play different roles in the exercise of our

agency. The distinction also has the advantage of respecting a folk psychological distinction between planned and spontaneous action.

Bratman grants that it is possible that not all our voluntary and purposive actions are caused by present-directed intentions, but he suggests that such actions may actually be the outcomes of "long-standing personal polic[ies]", plan-like conditional commitments to respond in a certain way if one is in a certain type of situation (1987, p. 126). He considers his action of reaching up and catching a ball that someone unexpectedly throws towards him for example. While he concedes that this action seems to be too automatic and unreflective to be the outcome of a plan-like commitment to act, he suggests that it may be the outcome of his previously adopted "policy of protecting myself from flying objects" (1987, p. 126). Bratman does not commit to the claim that such a treatment is possible for *all* apparently spontaneous actions, but the move allows him to expand the explanatory scope of his planning theory of intention and intentional action. As far as I can see though, maximising explanatory scope is the only motivation for generally treating spontaneous actions as the outcome of personal policies in this way. The downside of such a treatment is that it more or less eliminates the idea that some actions are performed in a genuinely spontaneous manner.

While I think there are good reasons to doubt that Bratman's planning theory of intention and intentional action is suitable for making sense of spontaneous action, I will for the sake of argument accept that when we perform intentional actions, these actions will always be the result of a present-directed plan-intention (which may or may not have been preceded by a future-directed intention or personal policy). Furthermore, I will for the sake of the argument accept that this applies both to individual action and to joint action. So even spontaneous joint actions, such as a sudden pass play that two footballers make, are the outcome of a shared intention. The point that I want to make in this chapter is that even if we concede this, individual action control or joint action control still cannot be fully understood in terms of processes of forming, acquiring, revising or rescinding intentions. Bratman's claim that "intention and action are not separately controlled by the agent,

but rather the agent's control of here action goes by way of her intention" (1987, p. 55) will have to be qualified. At the very least, the research I review in this chapter shows that Bratman's theory is incomplete in significant ways.

5.2 Online control and coordination of individual action

In this section, I argue that successful performance of individual action cannot merely be the result of a straightforward execution of a plan-intention. Consider what is required for successful performance of a simple bodily action, such as reaching out and grasping a cup. On a naïve picture of action control, what enables this performance is an intention to grasp the cup, which in detail represents the movement trajectory of the hand, the angle adjustments to be made to the arm, hand, and finger joints, as well as what force to exert once the fingers touch the handle of mug. This detailed plan then results in a sequence of motor signals that triggers neuromuscular changes that result in the intended action being performed. Since this type of control is carried out without guidance from feedback, it is called 'open-loop control' (feedback would close the loop). However, given the required degree of accuracy, the degrees of freedom involved, and the context-sensitive nature of the appropriate sequence of motor signals, this is not a computationally viable way of controlling actions.

It may seem that the solution is to rely on sensory feedback to fine-tune the movement on the fly, and thus obviate the need to plan all the details of the movement in advance. But this gives rise to a problem. While we clearly do rely on sensory feedback in monitoring and guiding most of our everyday activities, our motor system does not have time to wait for sensory feedback before adjusting the movements involved in most of our bodily actions. There is not enough time for such "closed-loop control". Arguably, there is also not enough time for action control to go by way of control of present-directed intentions. For a neural signal to propagate from motor cortex to the muscles and back through proprioceptive sensory pathways takes between 250 and 400 ms (Denier van der Gon, 1988; Itō, 1984; quoted in Grush, 2003, p. 76). Such a delay is too long for effective online control of, for

example, the fast movements that Bratman performs when he reaches out and catches the ball that was unexpectedly thrown at him, or the movements that I execute as I quickly swoop down my hand to catch a glass that has just slipped out of your hand.

5.2.1 Emulator-based action control

According to various current theories of motor control, the control dilemma just described is solved by what is called 'pseudo-closed-loop control'. The basic idea is that we use "internal models" to generate simulated feedback during action execution. This simulated feedback is used to make adjustments on the fly during execution (Desmurget & Grafton, 2000; Grush, 1997, 2004; Jeannerod, 1994; Wolpert, Miall, & Kawato, 1998). In pseudo-closed-loop control, the motor command from the brain is thus not only propagated to one's muscles, an "efference copy" of this command is also sent to a subsystem that I will call an "emulator" (following Grush, 1997). The emulator simulates the input-output mapping of the controlled system. If I am about to throw a ball at Bratman, then my arm is the controlled system. The emulator implements a "forward model" that computes the effects of the propagation of the motor signal to the controlled system. If the model is accurate, then it implements the same input-output loop as a closed-loop control system would have implemented. Since the machinery underpinning the emulator is in the brain rather than in the peripheral nervous system, the emulator can deliver the simulated feedback to the controller in time for anticipatory adjustments to be made before and during action execution (because the feedback signal does not need to travel such a long distance). In this way, we are able to perform smooth and precise movements despite the sluggish "wetware" that implements our action control system. The simulated feedback can also be used "offline" when we imagine ourselves performing an action from the first-person perspective (as a climber might imagine a move she'll have to make later when she starts to climb the route that she is facing). That the simulated feedback used during online motor control is also used during such motor imagery is suggested by the fact that the activities of actually doing something and of imagining doing are constrained in very similar ways, so

that, for example, the time it takes for someone to walk a certain distance is the same as the time it takes for the same person to imagine walking that distance (Decety, Jeannerod, & Prablanc, 1989; for discussion, see Currie & Ravenscroft, 2002, pt. 2; Butterfill & Sinigaglia, 2012). It is also suggested by the fact that the same cortical areas are involved both in motor control and in motor imagery (Parsons et al., 1995; Stephan et al., 1995).

While forward models enable smooth and precise movement execution, they cannot ensure that the executed movement is appropriate given the agent's beliefs, goals and situation. This depends on whether the motor signal received by the forward model is the appropriate signal in the first place. Delivering appropriate input to the forward model requires "motor planning"; that is, it requires computing which motor signals to send to the effectors in order to bring about a certain outcome (the action goal). It has been suggested that an emulator can perform this motor planning too by implementing what is called an "inverse model". An inverse model takes a representation of a desired outcome as input and computes what the appropriate motor command is for performing an action that is likely to bring that outcome about.

Now, the workings of emulation-based motor planning and motor control processes are not something we normally have reflective access to. This indicates that these processes do not go by way of control over present-directed intentions (that is, by way of the formation, acquisition, revision or rescindment of such intentions). While the motor imagery can be under conscious control, the online use of the simulated feedback of forward models in action control is not something we have reflective access to or conscious control over. (Of course, we do have conscious control of the offline use of motor imagery for action planning, as the preceding example of the climber's imaginative rehearsal of her future move shows.) But even if we did have reflective access and conscious control over these processes, there is good reason to think that what we in that case have access to and control over is not present-directed intentions. This is because the goal representations that are used by the emulators are arguably coded in a different format from the goal representations of plan-intentions

(Butterfill & Sinigaglia, 2012; see also Pacherie, 2008 on "M-intentions"). The plan-intentions that are inputs and outputs to personal-level planning and deliberation are propositional in format. This is what allows them to interact with other propositional attitudes like beliefs and desires during practical reasoning. This is also what allows us, as Butterfill and Sinigaglia point out, to intend outcomes that can only be picked out with quantification and identity, such as "that one cross seven distinct bridges in 48 hours without yet specifying which bridges or hours." (p. 12) But the representations involved in motor planning and control are in a format that is unable to encode such contents. Recall that emulators are involved in motor imagery as well as online motor control. Based on this commonality of neural resources, Butterfill and Sinigaglia argue that we can draw conclusions about the format of the representations involved in motor planning and control from premises about what it is possible for us to experientially imagine ourselves doing (from a first-person perspective). Since it is arguably impossible to *experientially* imagine oneself performing the activity of crossing seven distinct bridges in 48 hours without yet specifying which bridges or hours (as opposed to propositionally imagine *that* one does it), the format of the representations used in motor planning and control cannot be propositional in format. If the representations had that format, then we should arguably be able to engage in such imaginings. Furthermore, it seems that one cannot experientially imagine doing something biomechanically impossible either. For example, the climber cannot in this way imagine herself reaching up from the ground to grab the hold at the top of the ten-meter high rock face that she is standing in front.[103] However, one can surely propositionally imagine that one performs such a biomechanically impossible action, and as long as one didn't believe that it was biomechanically impossible, then one could also intend to perform that action. This suggests that the format of the representations involved in emulator-based motor planning and control cannot be propositional in format. The different constraints on

[103] As Currie and Ravenscroft (2002, pp. 77–78) note, we may *visually* imagine that our body changes shape and configuration in a biomechanically impossible way. This may be difficult to distinguish phenomenologically from imagining doing something that is biomechanically impossible. However, this does not mean that motor imagery and visual imagery are not two distinct forms of experiential imagining.

what one can intend and what can be represented by the motor system show that emulator-based motor planning and control is not mediated by control of present-directed intentions.

It may be objected that the whole control problem that 'pseudo-closed-loop control' is supposed to solve is based on the assumption that the main sensory feedback that one gets about the execution of one's action is proprioceptive. However, in many cases, one also receives exteroceptive sensory feedback that carries information about one's movement. This may arrive after a much shorter delay than proprioceptive feedback. After all, visual feedback will arrive to the retina with the speed of light. This, one might think, will provide time for reflective access and conscious control of even the fine-grained details of action control. Indeed, it is true that visual input typically plays a key role in enabling the motor control system to reliably estimate the current state of the body during action execution (see footnote 96). However, even such closed-loop aspects of motor control actually occur outside the agent's conscious awareness. This is vividly illustrated by an experiment that Pierre Fourneret and Marc Jeannerod (1998) carried out. Fourneret and Jeannerod instructed subjects to trace a straight line toward a target on a graphic tablet using a stylus held in their right hand. The movement of their hand was obstructed from the subject's view by a mirror placed in horizontally above the tablet, but they could see the reflection of a real-time visualisation of the stylus' trajectory in the mirror (from a computer screen that was placed straight above the mirror).

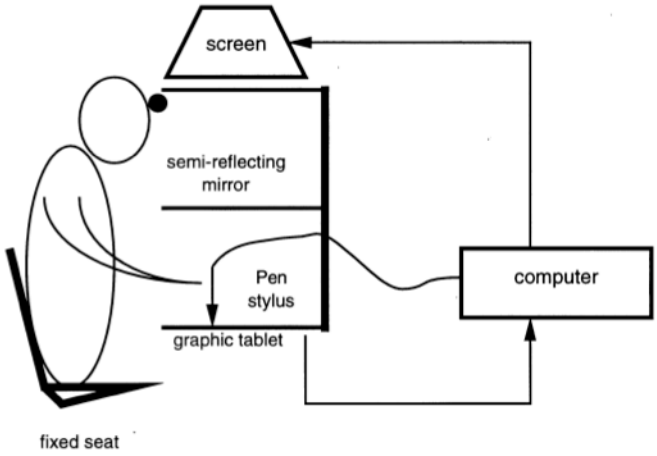

Figure 5 The set up of Fourneret and Jeannerod's experiment (figure from Fourneret and Jeannerod 1998).

In one session, the subject's task was to pick which one of thirteen different vectors on a card that best represented the actual hand movement that they had performed toward the target. In another session, the subject was instead afterwards asked to, with their eyes closed, recreate the line they had drawn. What the subjects did not know was that on some trials, the trajectory they saw reflected in the mirror was displaced to the left or the right by an angle of 2, 5, 7 or 10 degrees by the computer. The subjects had no problem correcting for the displacement introduced by the computer, so that they always traced a line *on the screen/mirror* that ended straight on target. Of course, in these trials with displacement, the line that was actually traced by the subject's hand underneath the mirror would not have been straight since it included movements that compensated for the computer's displacement. But when asked to estimate or recreate the movement trajectory that their arm had actually traced beneath the mirror, they estimated or recreated that trajectory as straight or almost straight (when the displacement was 10 degrees, the average estimated displacement was 2.5 degrees). Figure 6 shows an example of how the trajectory actually drawn and the trajectory that was seen by the subject could differ.

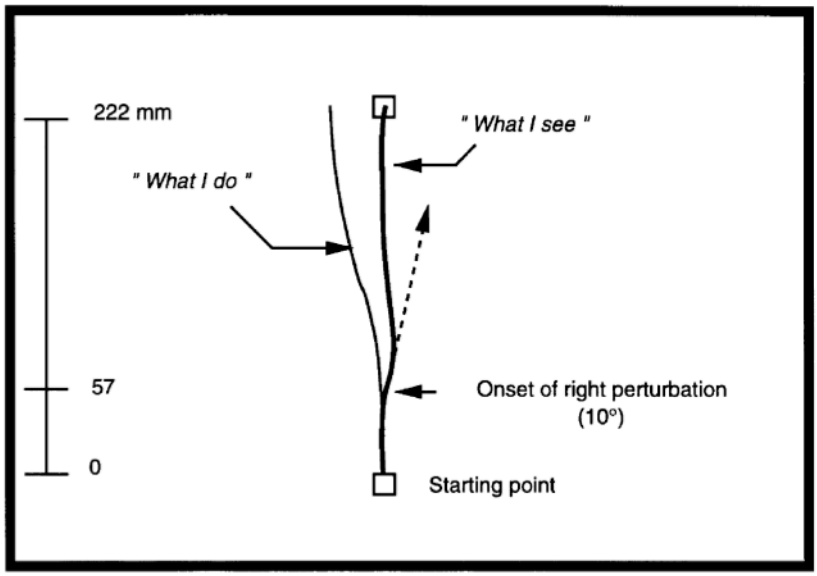

Figure 6 The performance of one subject on a trial where the computer introduced a displacement of the displayed trajectory (from Fourneret and Jeannerod 1998).

Given these incorrect estimates, the subjects seem to have been unaware of the corrections that must have been made by the motor control system in order to cancel the displacement introduced by the computer. This suggest that even when action execution is monitored and guided by closed-loop control processes that involve visual feedback, the loops normally doesn't go by way of the agent's conscious awareness and control of present-directed intentions. Note that this is just one illustration of how visually guided motor control appear to be partly encapsulated from processes of planning, reflection and conscious control. The empirical data that supports Milner and Goodale's (2006) so-called "two visual systems" hypothesis also indicates that visually guided motor control often proceeds without any interference or guidance from conscious awareness.[104] Such data also put pressure on Bratman's treatment of spontaneous actions as the outcome of personal policies, as it suggests that control of action sometimes simply bypasses systems for planning and reasoning.

[104] According to the "two visual systems" hypothesis, there are two more or less distinct streams in the brain that process input from the retinas. A ventral processing stream supports visual categorisation, reasoning and planning, whereas the largely independent dorsal processing stream supports visually guided motor control.

5.2.2 Intrapersonal coordination by self-organisation

The emulator-based pseudo-open-loop control model is a solution to a problem created by the demands on speed and timing that the real world puts on action control systems. Another problem for coordination and control arises from the sheer number of different component parts of the body that must be put into appropriate relations with each other for an action or activity to be performed successfully. If one reduces the problem to that of merely relating the body's hinge joints (elbows, knees, etc.) to each other in appropriate ways, then we have about a hundred degrees of freedom to deal with when we perform an action that involves the whole body (Turvey, 1990). (Note that this includes most of the bodily actions we perform, as various compensatory changes must almost always be made to retain balance and stability when we move a part of our body.) Given limited computational resources, a cognitive system must therefore be designed to keep the number of independent variables under executive control to a minimum.

According to one hypothesis, bodily coordination and control is achieved by exploiting the formation of so-called "coordinative structures" or "synergies"—structures formed by neural and muscular components that temporarily function as *self-organising* units of coordination in the execution of some specific task. Motor "plans" can then refer to simple control variables that cause whole coordination units to adaptively self-organise rather than to adjustments of body segments or joint angles (Turvey, 1990, p. 942). As an example, work by Bonnard and Pailhous (1991) suggests that when people are trying to intentionally control their walking speed, they do this by simply modifying the force exerted in their steps until the rhythmic beat (the cadence) of the walking fits the beat that they associate with the desired speed. This works because the components of the system that is achieving the walking self-organise so that cadence reliably tracks speed (in the face of various circumstances, so that the system automatically adapts if the walker puts on heavy boots or a backpack for example). Compensations in terms of the duration of different phases of the strides are here achieved without any need for centralised control. The intentional control of the walking speed can thus be achieved by present-directed intentions to walk with a certain speed, and this has the effect that

the agent increases or decreases the force exerted in their steps. The feedback that the agent then gets from the resulting cadence can then be compared to the goal speed in order to determine whether more or less force should be exerted in the agent's steps (other feedback such as optic flow is probably used as well, although in Bonnard and Pailhous' experiment, the subjects were walking on the treadmill so they could not rely on such visual feedback). We are not normally aware of the other changes in organisation of the walking cycle that occur as a consequence of this control loop. Self-organisation of coordinative structures is thus an adaptive aspect of action coordination that typically occurs without conscious awareness and outside the control of present-directed intentions.

Self-organisation has primarily been studied within the framework of Dynamical Systems Theory (DST), a very general theory for describing and explaining how various kinds of complex systems change over time. To clarify the difference between self-organisation and centralised control, I will here introduce some basic aspects of a DST-approach to bodily coordination. From the point of view of DST, the body in action is a complex and partly self-organising system that changes over time. To say that the body in action is self-organising is to say that the order of the system is not prescribed or specified by something outside the body itself, nor by some component within it. When applying DST to the behaviour of a system, one looks for different global behavioural modes of the system, and then tries to relate transitions between these modes to changes in one or several 'control parameters'. The different global behavioural modes are also referred to as different values of an 'order parameter' or 'collective variable'.

Dynamical systems research on bodily coordination has almost exclusively focused on rhythmic movements. In all rhythmic movement, several oscillators (say, swinging limbs or fingers, postural sway movements, rocking chairs, or swinging handheld pendulums) become entrained to each other, synchronised in-phase (>>, <<) or anti-phase (><, <>). The rhythmic limb movements that are produced by humans and by other animals exhibit the same properties as the rhythmic movements that are produced by various inanimate physical systems. These movement patterns

can all be modelled as coupled oscillators, which can be succinctly characterised by a simple equation that express the relation between oscillation frequency (control parameter) and the relative phase difference (order parameter) between the oscillators (known as the Haken-Kelso-Bunz (HKB) equation). The equation predicts, for example, that anti-phase coordination of two rhythmically moving limbs or fingers will make a transition to in-phase coordination as the rhythm frequency increases. In-phase and anti-phase coordination are both attractors in the state space of the dynamical system made up of the moving limbs, but the in-phase attractor is stronger. [105] One can observe this phenomenon by following these simple instructions: Start to move both of your index fingers rhythmically in anti-phase coordination at a low frequency. Now, slowly increase the frequency with which you move your fingers. At some point, you will switch into in-phase coordination. The HKB-equation predicts that these are the only two stable relative phase modes. The model also makes two more specific predictions, which have been confirmed in various studies of bimanual coordination, such as in the finger-wriggling paradigm just described (Kelso, 1995, pp. 58–59). First, it correctly predicts that there will be a nonlinear increase of fluctuations in the order parameter near the phase transition (so-called "critical fluctuations"). Secondly, it correctly predicts that when the system is near a phase transition, it will take much longer for it to relax back into its current phase if the system is perturbed from the outside (so-called "critical slowing down").

An example of a real world form of bodily coordination that has been elegantly explained by concepts and principles from DST is the gait and the switches between

[105] Dynamical systems are conceptualised as occupying a state or phase space that represents which states or phases (of the order parameter) that the system can occupy given a certain control parameter value. Stable states from which the system is unlikely to be perturbed from are called "attractors" and are represented as being at the bottom of valleys in the state space landscape. The area in the state space from which the system will move into an attractor is the attractor's "basin of attraction". States that the system will move away from are "repellors" and are the hills of the state space landscape. In the case of the horse's locomotion system, when the horse is moving slowly forward, there is only one attractor, the attractor for walking. As the speed increases, however, the state space will gradually start to change so that eventually, another attractor, for trotting, will pull the system into a new gait mode.

gait modes that quadrupeds exhibit when they change speed (see Schöner, Jiang, & Kelso, 1990). For example, take the system of a horse in locomotion. The horse has at least three types of gait (walk, trot, or gallop) with each type *typically* being used within a certain speed range. When the horse moves slowly, it tends to walk; when it moves at moderate speed it trots; when it moves very fast, then it gallops. As it happens, horses tend to use the gait type that consumes the least energy given current speed. We can imagine the horse's "choice" of gait (walk, trot, or gallop) at different speeds as the outcome of a control switch that puts the horse in a different gait when it passes certain speed thresholds. For example, when the horse reaches a certain speed while walking, the switch signals the relevant parts of the neural systems that underpin the horse's locomotion to go into "trot-mode". Such a centralised control solution to the problem of choosing the gait that optimises energy consumption is clearly *not* self-organised. The explanation that appeals to self-organised coordination instead falls back on very general principles and concepts from DST to succinctly describe the system's behaviour and predicts its evolution. The system's behaviour consists of the gaits and gait transitions as a function of speed. The order parameter of the system is the gait mode, and the control parameter is the horse's speed. Note that the order parameter only has three stable states (also called 'phases'), while the control parameter is continuous. When the speed gradually changes, the order parameter is instead stable until some critical point when the horse suddenly switches to a new gait. This switch between different order parameter values is called a phase transition or a bifurcation. It is plausible that under normal circumstances, a horse does not purposefully control its gait but merely its speed (and we might speculate that just as a human being uses cadence to determine current speed, a horse might use both its gait mode and cadence to determine its current speed).[106]

Proponents of a DST approach in cognitive science often pit it in opposition to representational approaches (such as the common coding theory, which is introduced

[106] Note that there are no critical fluctuations between a horse's gait modes when the speed is near a phase transition. Kelso suggests that this is because there are special neural mechanisms that stabilise the gait to avoid these fluctuations (1995, pp.72-74).

in section 5.3.1). But note that explanations based on dynamical models like the Haken-Kelso-Bunz equation are not *incompatible* with explanations that appeal to internal representations and information processing. Dynamical models are supposed to be completely neutral about what the mechanisms underpinning a system are. In other words, these mechanisms may well be in the business of manipulating internal representations and processing information. Nevertheless, the research done under the auspice of DST should at least make us cautious not to prematurely appeal to representation and prediction in explaining various forms of coordination, such as synchronisation of rhythmic movements for example. The principles of DST have been used to make sense of self-organisation in a wide range of systems, including weather systems, the formation of snowflakes, galaxies and all kinds of organic forms (Thelen & Smith 1996, p.50). None of these, arguably, are information processing systems. If even simple pendulums entrain—that is, their cycles become coupled through phase interlocking—then there is no reason to think that entrainment in bodily coordination must be mediated by mechanisms that trade in representations.

To sum up, in the case of individual action, there are clearly levels of motor control and coordination that do not occur by way of the agent's control of present-directed intentions. Both the formation of self-organised coordinative structures and fine-grained centralised motor control (based on the integration of real and simulated feedback) allow an agent's bodily movements to quickly adapt to the changing and uncertain circumstances of action execution. These means of coordination and control are "intelligent" in the sense that they enable an agent's actions to be performed in an appropriate way given the context and the agent's goals. In addition, the representations of outcomes that are involved in emulator-based motor planning and control also lend purposiveness to the actions themselves. As Butterfill and Sinigaglia (2012, p. 6) point out, if a representation represents the outcome of an action and reliably increases the likelihood that this outcome is brought about, then the action is at least partly purposive in virtue of this representation (this is why intentions are typically taken to make actions goal-directed). However, neither self-organisation nor motor planning and control are normally under the control of one's

plan-intentions. Part of what makes our activity and our actions appropriately related to our goals and intentions are thus not represented in our intentions: The intentionality of individual action outstrips the intentionality of our intentions.

5.3 Online control and coordination of joint action

For the same reasons that control and planning of individual action cannot be exhaustively explained by an agent's control of her intentions, control and planning of joint action cannot be exhaustively explained by appeal to such control of intentions either. When you and I successfully do a "high five", what the correct force, joint angle adjustments and so on are for each of us depend on the details of the other's movement. This suggests that each of us must represent and predict not only our own action, but also the other's action (dyadic coordination). Furthermore, we have to be able to (triadically) coordinate our actions based on these predictions in relation to our common goal of doing a high-five (in such a way that our palms are slapped together to make a satisfying loud sound without either of us injuring our hand) (Knoblich & Jordan, 2002; Sebanz et al., 2006).

Subpersonal control structures and representations play an important role in enabling the successful performance of such mundane joint actions. First of all, as we have seen, such structures and representations enable the participants to successfully execute the individual actions that are component parts of the joint action. Secondly, the emulator-based systems for motor planning and control also enable the agent to recognise and predict of the actions of others. Finally, there are subpersonal mechanisms that are dedicated for enabling triadic coordination in the context of joint activity (between several participants in relation to a common goal). The existence of such mechanisms suggests that actions are jointly purposive not only in virtue of the content of plan-like intentions but that they may also be jointly purposive in virtue of the contents of representations processed in subpersonal systems for motor planning and motor control (in the same way that an individual action is individually purposive in virtue of the goal representations that are processed during motor planning and control). Results in cognitive psychology and

cognitive neuroscience thus not only supports the view that the intentionality of individual action outstrips the intentionality of intention, the "joint intentionality" of joint action also outstrips the intentionality of "shared intentions".

In the second subsection, I go on to review some research that suggests that the formation of self-organised coordinative structures may have a role to play not only in intrapersonal bodily coordination and control, but also in interpersonal bodily coordination and control during joint activity.

5.3.1 Emulator-based action control and common coding

When one is performing a joint action together with someone else, one's motor control and motor planning must be adaptive to what one's co-participant is doing or is about to do. In order to do this, one would be well served by capacities for representing and predicting the actions of others. According to the so-called "common coding theory", such capacities are underpinned by the same neural system and representational format that enables planning and monitoring of one's own action (Hommel, 2009; Prinz, 1997). Various behavioural studies and phenomena can be explained by this theory. For example, common coding may be responsible for the so-called "chameleon effect", the tendency of people to unconsciously adopt the posture, gestures and manners of people they interact with. This tendency seems to be independent of why people engage in interaction and whether they have any special positive attitude toward their interlocutor before the interaction (Chartrand & Bargh, 1999). Behavioural experiments also show that response reaction times shorten when participants concurrently see someone else perform the response action, whereas they grow longer when they see someone perform a different action (e.g. Brass, Bekkering, & Prinz, 2001; Kilner, Paulignan, & Blakemore, 2003; Stürmer, Aschersleben, & Prinz, 2000). Such phenomena indicate some kind of direct perception-action influence that could be the result of common coding. As Sebanz et al put it, "observing an action leads to corepresentation", which may lead to facilitation or impairment of action execution, depending on whether the observed and performed action are congruent or not (2005, p. 1234). According to the

common coding view, this is a consequence of the fact that perception and performance of action are underpinned by a shared representational system, where actions are coded in terms of their perceptual effects (Hommel, 2009; Prinz, 1997). Formulated in terms of the emulation framework of motor control and motor planning that I presented in section 5.2.1, forward and inverse models might thus not only subserve the prediction and planning of one's own actions, but also the prediction of others' actions and the retrodiction of their prior goals (Gärdenfors, 2007; Wolpert, Doya, & Kawato, 2003).

Along with the behavioural evidence for common coding, there are converging neurophysiological findings that support the view. So-called "mirror neurons" are a possible neural correlate of the common sensorimotor action code. Pellegrino et al. (1992) discovered that there are bi-modal neurons in the ventral premotor cortex of macaques that fire both when the macaque observes an experimenter performing an action of type *F* (e.g. a reaching-and-grasping of a small object, say) and when the macaque itself performs an action of that same type *F*. By plausible extrapolation, mirror neurons exist in the human premotor cortex as well. There are many different accounts of the role of mirror neurons, but according to at least one proposal, their primary function in action observation is to predict the continuation of an observed action (Fogassi et al., 2005; see Jacob, 2009 for discussion). Clearly, such predications can play an important role in facilitating coordination in a joint activity.

But note that part of the behavioural evidence for the common coding view was that the execution of one's own action is impaired (slowed down) if one is simultaneously observing another agent perform a different action. A system that leads to such interference effects may clearly hinder rather than help in the context of a joint activity. When the actions that the participants need to contribute to the joint performance are all of the same type, corepresentation may of course improve performance (this may be the case for example, when two people do a "high five" or when they lift and carry a sofa together). But in many joint activities, the contributions of the participants will instead be of different but complementary types.

When I tilt the moka pot to pour coffee in the mug you are holding up, we normally don't want you to start tilting the mug to one side and pour the coffee on the floor!

In recent years, cognitive neuroscientists have found "counter-mirror neuron" activity in human subjects, and these may play a role in mediating coordination in joint activities that require participants to perform complementary actions (Catmur, Gillmeister, Bird, Liepelt, & Brass, 2008; Heyes, 2010; Newman-Norlund, van Schie, van Zuijlen, & Bekkering, 2007). A counter-mirror neuron fires when a subject observes a certain type of action (e.g. someone else grasping an object), and when one performs another type of action that is associated with that observed action type (e.g. the subject herself releasing an object). Many have taken the findings of counter-mirror neurons to show that people have two kinds of bimodal neurons, one type that are used to represent and simulate the actions of others (mirror neurons) and another type that is used to represent and simulate appropriate responses to the actions of others (counter-mirror neurons) (e.g. Newman-Norlund et al., 2007; Knoblich & Sebanz, 2008, p. 2026). But it may be that mirror and counter-mirror neurons are just bimodal neurons that gets trained through associative learning during development to respond in different ways to perceived actions (see Heyes, 2010 for this "associative hypothesis"). Mirror neurons could get their mirroring properties as a result of a developing child having perception-action links forged while interacting with adults who imitate them for example, or while observing themselves moving in front of mirrors, or as a result of participating in games, sports and dances where they perform similar actions simultaneously with others. But other types of social interaction would produce other perception-action links between complementary actions. If this associative hypothesis is correct, then the common coding theory may be true not because it correctly characterises a basic architectural feature of the human cognitive system, but because a common code is a reflection of the kind of social interactions and the kind of environment in which humans typically develop. Whether or not this hypothesis about the origin of the mirroring and counter-mirroring properties is true, various kinds of bimodal neurons may clearly play a role in mediating habitual and entrenched forms of coordination in the context of joint activity. But if the common coding theory is generally correct about

the architecture of our perception and action systems, then we should expect there to be mechanisms that allow participants involved in joint activities to cope with potential interference effects.

5.3.2 Shared task representations

Sebanz et al. (2005) hypothesize that representations and predictions or one's own and other's actions are kept apart but coordinated through what they call "higher-level task representations" in the context of joint activities (see also Atmaca, Sebanz, Prinz, & Knoblich, 2008). The idea is that, in some circumstances, participants not only represent the actions of co-participants, but also the tasks that they themselves and their co-participants are expected to perform. Each participant would then not only represent and predict their own tasks and actions for the purpose of controlling their part of the joint action, but they would also represent and predict the tasks and actions of their co-participants.

The hypothesis receives support from behavioural experiments that demonstrate interference effects not only when a participant sees a co-participant perform an action that is incongruent with their own action, but also when they see a stimulus in the environment that a co-participant is expected to respond to with such an incongruent action. In a simple but ingenious experiment, Sebanz et al. (2003) first presented individual participants with a two-choice reaction time task. The subjects were seated facing a computer screen that displayed the stimuli that they had to respond to using two buttons that were placed beside each other in front of them. In each trial of the experiment, a hand with an index finger that was pointing to the right, to the left or straight ahead was presented to the subject. There was also a ring on the hand's index finger that was either red or green. The task of the subject was to push one of the buttons (say, the right button) when the ring had one colour (say, red), and the other button (left) when the ring had the other colour (green). The direction of the pointing finger was a feature of the stimulus that was task-irrelevant. Experiments with similar two-choice tasks with task-irrelevant spatially associated stimulus features have consistently shown that responses that are spatially

compatible with the task-irrelevant feature (say, stimulus feature: right-pointing finger; response: press right button) are performed quicker than when that feature is spatially neutral. Responses that are spatially incompatible with the task-irrelevant feature (stimulus feature: left-pointing finger, response: press right button) are on the other hand performed slower compared to when that stimulus is spatially neutral. However, this spatial (in)compatibility effect disappears in "go-nogo" conditions where the subject only has to respond with one type of action to one type of stimulus (e.g., "press the left button if the ring colour is green, otherwise do nothing"). In other words, a two-choice task with spatially associated choices (press the right or the left button) is required for the spatial (in)compatibility effect to occur.

The task just described is a version of the so-called Simon task (Simon & Rudell, 1967), which was replicated by Sebanz et al. (2003, Experiment 1) with the expected results. Crucially, Sebanz et al. added a third condition in which the two-choice set-up was presented to pairs of subjects, with each member of the pair being responsible for one button. Note that with this setup, each subject performed a task that is exactly like the task performed by participants in the individual "go-nogo" set-up.

Figure 7 The social Simon task. Considered individually, each member of the participant dyad performs a task that is equivalent to the "go-nogo" condition in the non-social Simon task. The pointing direction of the index finger on the screen is the task-irrelevant stimulus, while the colour of the ring on the same finger is the task-relevant stimulus. (Figure from Sebanz et al., 2003)

If each member of the pair is considered separately, then we should expect that the performance of each member should roughly be the same as the performance of the individual participant performing the "go-nogo" version of the original Simon task, that is, there ought to be no interference effect. But the spatial (in)compatibility effect actually reappeared in this social version of the Simon task. This suggests that each participant was considering herself faced with the two-choice task with spatially associated choices again, but with control of the one button delegated to her co-participant (but still under her supervision and monitoring). Sebanz et al.'s (2005) explanation of this "social Simon effect" is that a participant's predictions of their own and the other's actions interact with a representation of the tasks that they and their co-participant are facing (a "shared"[107] task representation). According to this explanation, when a subject perceives the co-participant's relevant stimulus feature, the representation of the co-participant's task rule is activated and this in turn triggers a simulation of the co-participant's expected response.[108] Because of common coding, the triggered simulation interferes with the participant's planning and execution of their own response.

Note that the stimuli were presented one at a time, so there was no simultaneous execution and observation of action. The subjects sat beside each other in front of the screen and took turns performing their respective task, and there is nothing in the task that requires the subjects to have a common goal. There is no actual interdependence between the actions of the two participants. So if Sebanz et al.'s interpretation of the social Simon effect is correct, then a "shared task representation" seems to be automatically created in each participant, even if such a

[107] "Shared" in the sense of replicated in each participant.

[108] It is actually not required that the participant perceives the co-participant's action for the spatial (in)compatibility effects to occur. Sebanz et al. (2003, Experiment 2) carried out an experiment where the participants performed the social Simon task without any auditory or visual feedback of their co-participant's action but the spatial (in)compatibility effect occurred nevertheless. In at least one variant of Sebanz et al.'s experiment (Tsai, Kuo, Hung, & Tzeng, 2008), the (in)compatibility effects on reaction times were exhibited even when participants merely believed that they were in a joint action context, without being able to observe a co-participant at all (however, a study by Welsh et al. (2007) seems to contradict this).

representation hinders more than it helps in the context of the experiment (and serves no actual coordinative function).

Sebanz et al.'s explanation of the social Simon effect that I have presented here—based on shared task representations plus corepresentation—is of course not the only possible explanation. The empirical findings are, as empirical findings are, open to alternative interpretations. For example, Wenke et al. (2011) argue that that the interference created by the task-irrelevant stimulus feature (the direction of the pointing finger) is automatically treated as a cue regarding whose turn it is to act. On this view, the (in)compatibility effect is an effect of the (in)compatibility of agents rather than of responses. When a participant sees the stimulus that she is supposed to respond to, but also sees a finger pointing toward the other participant, this interferes with the process of identifying which of the two agents—herself or her co-participant—that is supposed to respond. Wenke et al.'s competing explanation of the social Simon effect is of course compatible with the existence of shared task representations, but if correct, it undermines the motivation to posit such representations to explain the social Simon effect.

It is hard to know to what extent something like shared task representations would be useful in joint activities outside the psychology laboratory. In the social Simon task, participants are literally being given tasks rules that associate actions with particular stimuli by the experimenters. In many joint activities "in the wild" such task rules will often be harder to identify though. But it is not unreasonable to suppose that representations of task rules may sometimes play a role in enabling coordination in the wild too. Take again the example of two people washing dishes together. Here, the task distribution between two individuals is usually known to both, one is washing and rinsing the dishes, handing over items to the other participant who wipes them dry with a cloth and then puts them into a cupboard, and this can, stretching language a bit, be described in terms of task rules which map stimuli to responses (e.g. if I see you offer a new item for me to dry, then grab it, wipe it and put it in the cupboard).

The social Simon effect suggests that people tend to treat situations involving others as a kind of joint action situation, even if there is no actual interdependence between their own and others' actions. Arguably, what matters is whether or not the participants frame what they are doing as part of something done together with the other, or if they frame it as something they are doing individually.[109] This is the upshot of both the shared task representation account, and the agent-identification account of Wenke et al. Both these accounts posit mechanisms that are dedicated to the execution, monitoring and control of action in a joint action context. Importantly, these mechanisms appear to come online automatically and unreflectively, even in situations where there is no actual need for the agents to represent and predict other's tasks and actions. This suggests that an agent's emulation-based motor planning and control processes automatically modulated by their social context. In contexts where people are performing tasks that are actually interdependent, it is plausible that they will be more likely to frame what they are doing as a joint endeavour than in a context such as that given by the social Simon task.[110] In such contexts, shared task representations are likely to help rather than hinder, since such representations will enable participants to predict their co-participants' actions.

5.3.3 Common coding and joint perceptual effects

Recall that, according to the common coding view, the common "code" of action perception and action performance consists of perceptual effects of actions (Hommel, 2009). Now, doing something together with a partner has different perceptual effects compared to oneself doing one's part of that joint action on one's own. If contributions to joint actions are coded in term these perceptual effects, then we should expect that a member of a group will perform her part of a joint action faster if she perceives another group performing a similar joint action (more similar

[109] Exactly how to explicate what such framing consists in is a moot point—perhaps it is a matter of whether or not the participants think they have a Bratmanian shared intentions "that we participate in the experiment".

[110] See Bacharach (2006, pp. 81–87) on the interdependence hypothesis concerning group identification.

perceptual effects) compared to if she merely perceives an individual performing an action that is similar to her part in a non-group context (less similar perceptual effects).

Tsai, Sebanz and Knoblich (2011) had participants perform a simple reaction-time task with a confederate in order to test this prediction. Each participant sat with a confederate facing a screen where two hands were displayed, each hand with its index finger raised above a key. In one condition, the hands that the participant and the confederate saw were two left hands (see the left pane in Figure 8). The "group" consisting of the participant and the confederate were thus observing the left hands of the two members of another group (the two left hands must belong to two different individuals who, it is natural to assume, are standing next to each other, suggesting that they are part of the same group). This condition was thus characterised by "inter-group congruency". In the other incongruent condition, the participant and the confederate were instead facing one left hand and one right hand, palms facing each other, as if they were a pair of hands belonging to one and the same agent (see the right pane in Figure 8).

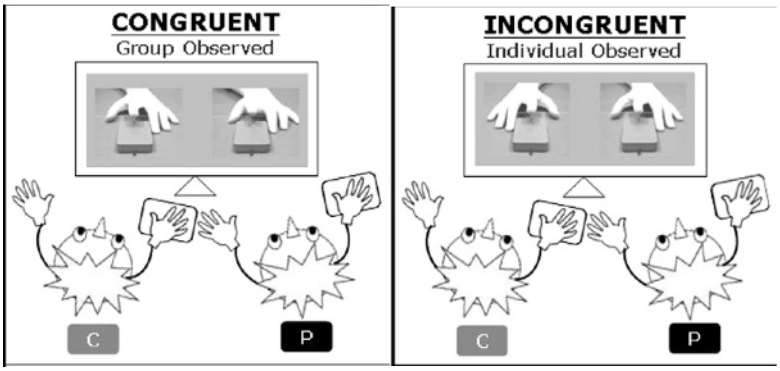

Figure 8 The two conditions in Tsai et al.'s (2011) group mimicry study.

In all conditions and trials, the only instruction given to the participants was that they press a key when the ipsilateral hand immediately in front of them moved. In other words, what the confederate did and what the contralateral hand on the screen did was never relevant to the participant's task.

In half of the trials on which the participant was supposed to respond, both hands on the screen moved simultaneously, and on half of them, only the hand in front of the participant moved. What the confederate did varied between two conditions. In the numerically compatible condition, the confederate's task paralleled that of the participant. In other words, the confederate responded with a key press when the hand in front of him or her moved. In the numerically incompatible condition, the confederate's task was instead to respond when the hand in front of the participant moved (the contralateral hand, from the confederate's perspective) but not when the hand in front of the confederate moved. In effect, in the numerically compatible condition, both the participant and the confederate responded ("WE response" in Figure 9) when both hands on the screen moved, and only one of them responded when only one hand moved on the screen. In the numerically incompatible condition on the other hand, only the participant responded when both hands on the screen moved ("ME response" in Figure 9), and both the participant and the confederate responded when only the hand in front of the participant moved.

As expected, the participant responded significantly faster when there was numerical compatibility in the "inter-group congruency" condition: The participant responded faster by himself or herself in response to one hand moving but she responded faster if the confederate also responded when both hands on the screen moved.

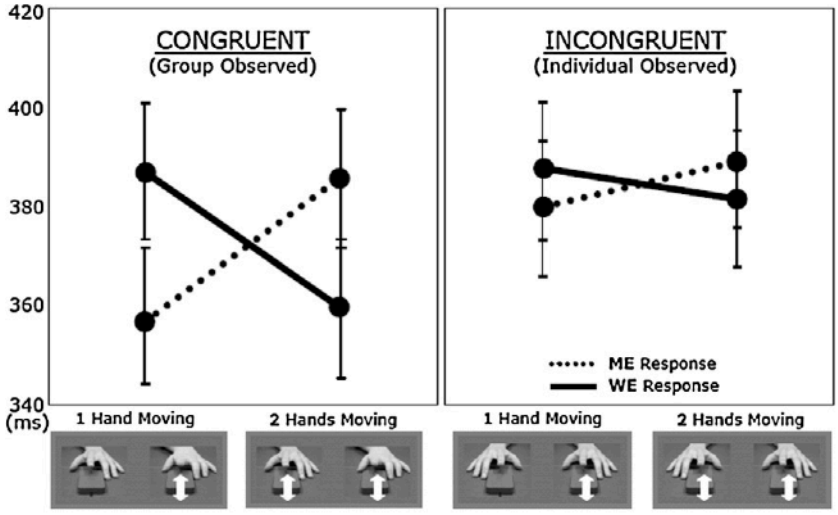

Figure 9 The results of Experiment 1 reported in Tsai et al. (2011).

Note that there was no actual interdependence between the actions of the confederate and the participant, and the participant was not aware of this effect on their reaction times. Plausibly, it was only when the participant took the two hands on the screen to belong to two agents that Tsai at al. found a significant difference. This interpretation is strengthened by the fact that Tsai et al. in a second experiment got the same results when inter-group congruency/incongruency was induced in a different way. Instead of using two left hands in the congruent condition, they used a right and a left hand that had different skin colours, but with palms facing each other.

I suggested in the previous section that a plausible interpretation of the social Simon effect is that participants tend to treat each other as partners in a joint endeavour (possibly with a common goal) even when there is no actual interdependence between their actions. Conceiving of what they are doing as parts of a joint action has the top-down effect that they represent the other's task (stimulus-response relation) or possibly whose turn it is to act. Similarly, this "group mimicry" effect suggests that if people conceive of themselves as taking part in a joint action, then they will take into account not just the perceptual effects of their own contribution but also the other's contribution (here I am going beyond what Tsai et al. claim about their results). This means that observing similar perceptual effects created by what appears to be a joint action of another group will activate representations involved in the performance of an individual's own part of such a joint action. Like the effects of shared task representations, these effects are automatic and not under the conscious control of the participants. Below the level of practical reasoning and personal-level planning, our cognitive systems for action planning and action perception thus differentiate between behaviourally identical actions based on whether they are mere individual actions or actions performed as parts of joint actions.

5.3.4 Interpersonal coordination by self-organisation

In this section, I review some work that shows that coordinative structures that may simplify bodily coordination and control can be formed through not just the self-organisation of components belonging to one and the same body, but also through

the self-organisation of components that belong to several bodies. A speculative upshot of this is that motor plans may trigger adaptive changes in coordinative structures that span across an interpersonal system (by changing the value of a control parameter).

As I mentioned in section 5.2.2, the so-called HKB-equation succinctly models the relations between the oscillation frequency (control parameter) and the relative phase difference (order parameter) in a system of coupled oscillators. What is striking is that the HKB-equation that I presented is that can be used not only predict and characterise a range of phenomena that occurs in bimanual coordination. The equation appears to capture the properties of rhythmic *interpersonal* bodily coordination too. For example, two human beings, one walking in front of and in view of the other, tend to spontaneously end up in two stable patterns which are equivalent to pace (in-phase) and trot (anti-phase) gait modes among horses that I discussed earlier (Harrison & Richardson, 2009). The stable states, critical fluctuations and critical slowing down that can be observed in the bimanual coordination in human subjects also characterises the dynamics of two people who try to coordinate the rhythmic movements of, for example, their wriggling fingers or their swinging lower legs. In the interpersonal case, the coupling that enables coordination is forged through visual rather than neuromuscular connections, but this is irrelevant to the whether the general coordination principles of DST apply to the system or not. It does not matter whether the connection between the oscillators are mechanical or optical, the same model can predict the behaviour of the system in question.

The interpersonal entrainment that occurs when two people intentionally try to, for example, synchronise their finger-wriggling to the pace of a metronome, also occurs unintentionally, when they have visual information about each other's movements but are intentionally trying to perform some other task. Richardson et al. (2007) demonstrated that two people in visual contact with each other's movements unintentionally entrained their rocking movements into an in-phase rhythm. The participants were told that the study intended to investigate the ergonomics of

rocking chair movements and how different postural positions influence the rocking movement. This allowed the experimenters to manipulate the gaze of the participants by instructing them to gaze at certain locations. Other experiments indicate that it is people's gaze behaviour, induced by visual tracking of the other's movements, which mediate the entrainment process (Schmidt & Richardson, 2008). But it is not the only mediator. Further experiments have also shown that entrainment of postural sway occurs when people are talking to each other, even when they are not visible to each other. Instead, it is the entrainment of the participants' speech rhythm propagates through the body to induce entrainment of postural sway (Fowler, Richardson, Marsh, & Shockley, 2008; Shockley, Richardson, & Dale, 2009). The same model thus predicts that entrainment will occur in all kinds of systems that are made up of coupled oscillators, regardless of whether the oscillator movements are produced by pendulums or limbs, or whether the entrainment is mediated by vibrations in a wall, neuromuscular signals or the attentional tracking of movement or sounds between people. Note that it is hard to see how this could be explained with a "control switch" explanation, whereas it follows naturally from DST.

In light of this work on coordination dynamics, there is no reason to think that compensatory changes that occur in response to changes in control parameters might be the result of self-organisation in systems—coordinative structures—that range across individuals. Recall the hypothesis that one way in which bodily coordination and control is made possible—in spite of the huge number of components that have to be brought into appropriate relations with each other in order to facilitate successful action performance—is that many components self-organise into coordinative structures that can then be controlled adaptively by merely adjusting control parameters (such as increasing the force exerted in each step as one walks, say). The DST research on bodily coordination of rhythmic movements shows that coordinative structures may form interpersonally as well as intrapersonally. This suggests that control of joint activity can be simplified in a way similar to the way that control of individual activity can be simplified. Perhaps individuals can control interpersonal coordinative structures by intentionally controlling certain control parameters, and then rely on the components of their own and others' bodies self-

organise in appropriate ways. Shockley et al. seem to suggest that something like this is possible in the quote that follows. They propose that synergies formed through entrainment in social interaction may account for much of the coordination that occurs in face-to-face communication, and describes the notion of an interpersonal coordinative structure in the following way:

> If joint cognitive tasks are similarly achieved via a functionally defined cross-person organization, then perturbing/constraining the actions of one component of the cross-person coordinative structure of one member (e.g. a relevant body segment, optical information, cognitive constraints) should result in rapid compensatory changes in other components of the cross-person structure (e.g. changes in movement patterns of a body segment, looking patterns, or cognitive kinematics in the other member of the pair). In other words, if the cross-person coordinative structure consists of a certain relation among body segments and cognitive states, then constraints on the (action) effectors of one person should affect the movements and/or cognition of the other member of the pair as readily as cognition can affect one's own efforts. (Shockley et al., 2009, p. 315)

When it comes to the kind of skilful and habitually entrenched interpersonal bodily coordination that may be observed when two skilled dance partners do a well-rehearsed performance for example, interpersonal coordinative structures may be present and perhaps play an important role. Coordination may be mediated both by visual and tactile information, as well as through both individuals becoming entrained to patterns external to them, such as the rhythm of the music. The moves of one dancer constrains the moves the other, just like movement of a bodily segment in one dancer constrain the movement of another bodily segment in the same dancer. Just like an individual may change walking speed by simply intentionally increasing exerted force, and let the changes within the cycle of walking (consisting of the stance phase, double support phase, and the swing phase) self-organise, an individual may intentionally change a control parameter for an interpersonal coordinative structure and let its components across the agents involved self-organise. The presence of these coordinative structures can in such cases simply be taken for granted by each participant's action system, just like the workings of units of coordination comprised of components of its own body are taken for granted by executive control systems in the individual. While this is no more than speculation at

this point, I can see no reason why it cannot play some role in enabling the coordination of joint activities.

Of course, I am not suggesting that this type of emergent coordination could be the whole story of the coordination involved when two people dance together for example. Emulator-based motor planning and motor control will be no doubt be involved in breaking up coordinative structures and patterns that are the result of entrainment. It is hard to see how the DST approach could be scaled up to explain more all the coordination involved in, for example, two people washing dishes together (and if it could, I imagine that this explanation would provide us with any useful understanding of how this coordination was made possible). This is not to say that ideas from DST might not play a role in explanations of how various forms of complex coordination and motor control are possible. No doubt, motor control and motor planning, emergent coordination, as well as personal-level planning and practical reasoning are often all at work simultaneously both in individual and joint action. To understand how these different systems and forces interact in joint action will be challenging indeed (Knoblich et al., 2011).

5.4 Conclusion

I have presented different ways in which participants who are engaged in a joint activity coordinate their actions in ways that do not rely on plan-like commitments that are the end-points of practical reasoning. On the one hand, systems underlying action performance enable online motor control of action beyond what is specified by such plan-like commitments. The representations that are involved in such action control not only enable fine-grained and smooth action execution but also, as we have seen, action observation and the prediction of the actions of others. While perception and prediction are of course important for successful coordination of many joint activities, the fact that perception and action shares a representational system (has a "code" in common) has been shown to create interference effects under experimental conditions where subjects are asked to perform actions while they are observing the performance of different action. A proposal I picked up from

Sebanz et al. was that when agents engage in joint activities (which often require agents to perform different complementary actions), they have "shared task representations" that represent the division of labour among the participants, mapping certain stimuli to certain agents and responses. Such representations will enable to predict what others' are likely to do within the context of a joint activity based not on the actions but based on the presence of stimuli in the environment that they are supposed to respond to.

The workings of these mechanisms and representations that are involved in the motor planning and control of both individual and joint actions typically are not accessible to an agent's deliberation, planning and reasoning. This means that an agent's control of his action, and several agents' control of their joint action, does not merely occur by way of the control of intentions. Furthermore, there are general principles of emergent coordination that facilitates both intrapersonal and interpersonal bodily coordination. An important question for the cognitive science of joint action to pursue is how representational centralised mechanisms interact with, exploit, and counteract coordination that emerge from self-organising coordination structures. Another important question is how all this harmonically interact with systems for planning agency and practical reasoning (see Butterfill & Sinigaglia, 2012). These questions all arise both for individual and for joint action.

6 Conclusions

Philosophers have typically assumed that the only kind of small-scale tightly coupled multi-agent activity that is philosophically interesting is an activity that is coordinated and caused by a "shared intention" (or a "joint" or "collective" intention). According to most accounts, such a shared intention consists of a mutually known pattern of intentions, commitments and beliefs among the participant. The pattern structures the distribution of roles and tasks for the sake of the participants' pursuit of a common goal. In this thesis, I have argued that some kinds of small-scale multi-agent activities that are not the outcome of this kind of shared intention are still interesting kinds of genuine joint activity. This is not just a terminological issue (if it is, then it is a terminological issue with substantial consequences). Thinking that joint activities caused by shared intentions are the paradigmatic form of joint activity has arguably created a false dualism between allegedly bona fide joint activity, where agents together make up a collective 'we', and other multi-agent activity, where they merely count as a distributive 'we'. The chapters of this thesis are held together by two lacunas that have been left relatively unexplored in the wake of this false dualism. The first concerns how to make sense of the apparently joint cooperative activities of agents that lack the capacities for planning and "mindreading" that one must have in order to be a party to a shared intention. The second concerns how participants who have a shared intention are able to coordinate their actions "online"—that is, during action execution as a joint activity unfolds—without recourse to plans that specify in advance what they should do.

Chapters 2 and 3 were devoted to the first lacuna. I there developed conceptual and theoretical resources for thinking about the joint activities of relatively cognitively unsophisticated agents such as young children and non-human primates. As I have repeatedly pointed out, accounts of joint activity such as Bratman's are too cognitively and conceptually demanding in light of the socio-cognitive profiles of such agents. A common theme in chapters 2 and 3 was to accommodate an account

of joint activity to agents who have the concept of 'goal' but not that of 'belief'. In chapter 2, I showed the widely accepted requirement that agents must have common knowledge of each other's goals or intentions in order act jointly requires that they have the concept of belief. Such a requirement is thus one of the reasons why most accounts of joint activity are too conceptually demanding for agents with such a socio-cognitive profile.

However, I showed that if participants lack the concept of belief, then one of the two main motivations for the common knowledge requirement—to filter out certain "concealment" cases that intuitively aren't cases of genuine joint activity—actually dissipates. Concealment cases can only be constructed if the participants have the concept of belief, so there is no need to rule out such cases when it comes to participants who lack the concept. Furthermore, I have suggested that there is a kind "openness" that only requires of participants that they have the concept of goal but not that of belief could satisfy the other main motivation for the common knowledge requirement, to make sense of the idea that joint activities are non-accidentally coordinated. This openness requires of the agents that they are each aware of a mutual dependency that must exist between their goals, but this awareness need not itself be mutual. But I also submitted at the end of chapter 3 that there are even weaker notions of joint activity that may be accidentally coordinated. Agents can thus be participants in such joint activities having neither common knowledge nor any awareness of mutual goal-dependency.

An upshot of my analysis of the role of common knowledge in accounts of joint activity was an initial argument for a kind of pluralism about kinds of joint activity is further established in chapter 3. Even if we restrict our analysandum to small-scale, more or less egalitarian and highly interdependent joint activities, these come in several different genuine kinds (that is, we need not consider cases of large-scale industrial action, flash mobs, or the activities of structured groups such as corporations, committees and political parties to be pluralists about genuine joint activity). The most robust kind of joint activity, in which only agents who have the concept of belief can participate, demands that the agents have common knowledge

of each other's goals or intentions with regard to the joint activity. A less robust kind of joint action does not require common knowledge but must still be non-accidentally coordinated, and thus demand of the participants that they each have some awareness of mutual-goal dependency. Note that the notion of "openness" as awareness of mutual goal-dependency was merely a specification of what a full account of such openness needs to achieve. Providing such a full account will require further research.

In chapter 3, I constructed an account of a kind of joint activity driven by what I called a "joint goal". The account was supposed to be an explication of a definition of cooperative activity as that of "two or more individuals acting together to achieve a common goal" that has been influential in the study of the cooperative capacities and proclivities of young children and non-human animals (Boesch & Boesch, 1989; Brownell et al., 2006; Naderi et al., 2001). While this kind of joint activity falls short of being a shared intentional activity in Bratman's sense, it does not merely consist of a collection of coordinated pursuits of individualistic goals; it involves agents having a common goal and being in a position to mutually benefit from this fact. When two or more agents are in such a position, they have a "joint goal". A large part of chapter 3 was taken up by a discussion of what the criteria for having a "common goal" should be (alternatively put, for having the "same goal" or to "share a goal"), in order for this notion to earn its place in a definition of cooperation. Among other things, I argued that agents do not need to represent a goal in the same way in order for it to be their common goal. Furthermore, a common goal need not have collective content ("we"-content): it need not have contents such as "that we catch the prey" or "that I do my part in our hunting of the prey", but may be agent-neutral ("that the prey is caught") or only implicitly identify the agent of the action ("to catch the prey"). These features of my account set it apart from e.g. Bratman's account, and arguably, most accounts of joint activity that appeal to the notion of shared, joint or collective intention. Both features also help make my account less cognitively and conceptually demanding than Bratman's account. Participants need not engage in Level-2 perspective taking in order to take advantage of the fact that they have a common goal, and they do not need to have the concept of "that we *J*".

In chapters 4 and 5, I moved on to the second lacuna in the philosophical literature on joint action. In chapter 4, I reframed Tollefsen's (2005) account of shared intention-in-action, and argued that it is apt primarily as an account of skilful joint activities of adult human beings, rather than as an account of children's joint activity. I highlighted what appeared to be a problem with her account, namely that it is in conflict with a widely accepted constraint on the content of intentions-in-action. According to this "exclusivity constraint", one cannot intend to perform another agent's action, even if one might intend *that* she perform it. I showed that the exclusivity constraint should not be accepted as an unconditional constraint on the contents of intentions-in-action: one may intend to perform a basic action that belongs both to oneself and to another agent. Based on the phenomenology of tool use, I first argued that intentions-in-action of one's basic actions may be "technologically extended", meaning that their contents are not restricted to concern the agent's bodily movements. In analogy with this, I then argued that the phenomenology of some skilful joint activities supported the idea that one's basic intentions-in-action may be "socially extended", in violation of the widely accepted exclusivity constraint. This argument rests on the details of Searle's account of intention-in-action, and at the end of the chapter, I imagined a potential objector saying that the account was actually just a notational variant of Bratman's account. The account, the objection went, is actually subsumed under Bratman's more general account of shared intention, since plan-intentions need not be about the future, but may also concern what one should do *now*. At the beginning of chapter 5, I presented some reasons why Searle's distinction between prior intention and intention-in-action in some ways is clearer than Bratman's distinction between future-directed and present-directed intentions. These are also reasons for resisting the view that Tollefsen's account is merely a notational variant of a special limiting application of Bratman's account.

In chapter 5, I showed—by reviewing work in cognitive psychology, cognitive neuroscience and human movement science—that even if we grant that Bratman's planning framework is applicable to spontaneous joint actions, the joint intentionality

of joint actions in general cannot be reduced the contents of the participants plan-intentions. There are mechanisms and representations involved in the motor planning and control of both individual and joint actions that typically are not accessible to the agent's deliberation, planning and reasoning. Furthermore, it is partly in virtue of these mechanisms and representations that the actions are imbued with purposiveness. This means that an agent's control of his action, as well as several agents' control of their joint action, does not merely occur by way of control of intentions. Furthermore, I presented work that shows there are general principles of coordination that facilitates both intrapersonal and interpersonal bodily coordination. The workings of such principles enable motor planning and motor control systems to exploit the formation of coordinative structures that reduces the degrees of freedom that must be centrally controlled during the unfolding of an action or activity. Following others, I pointed out that difficult questions are unanswered regarding how the various states, mechanism and forces that contribute to the coordination of joint activities interact in harmonious ways.

Bibliography

Alonso, F. (2009). Shared Intention, reliance, and interpersonal obligations. *Ethics*, *119*(3), 444–475.

Anscombe, G. E. M. (1969). *Intention*. Ithaca: Cornell University Press.

Apperly, I. A. (2011). *Mindreaders: The Cognitive Basis of "Theory of Mind"*. Psychology Press.

Apperly, I. A., Riggs, K., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, *17*(10), 841.

Atmaca, S., Sebanz, N., Prinz, W., & Knoblich, G. (2008). Action co-representation: The joint SNARC - effect. *Social Neuroscience*, *3*(3), 410–420.

Bacharach, M. (2006). *Beyond Individual Choice: Teams and Frames in Game Theory*. Princeton University Press.

Back, E., & Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition*, *115*(1), 54–70.

Baillargeon, R., Scott, R., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118.

Becchio, C., Sartori, L., & Castiello, U. (2010). Toward You: The Social Side of Actions. *Current Directions in Psychological Science*, *19*(3), 183–188.

Behne, T., Carpenter, M., & Call, J. (2005). Unwilling versus unable: infants' understanding of intentional action. *Developmental Psychology*, *41*(2), 328–337.

Bermúdez, J. L. (1995). Ecological Perception and the Notion of a Nonconceptual Point of View. In J. L. Bermúdez, A. Marcel, & N. Eilan (Eds.), *The Body and the Self* (pp. 153–173). MIT Press.

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

Boesch, C. (2005). Joint cooperative hunting among wild chimpanzees: Taking natural observations seriously. *Behavioral and Brain Sciences*, *28*(5), 692–693.

Boesch, C., & Boesch, H. (1989). Hunting behavior of wild chimpanzees in the Taï National Park. *American journal of physical anthropology*, *78*(4), 547–573.

Bonnard, M., & Pailhous, J. (1991). Intentional Compensation for Selective Loading Affecting Human Gait Phases. *Journal of Motor Behavior*, *23*(1), 4–12.

Botvinick, M., & Cohen, J. (1998). Rubber hands "feel" touch that eyes see. *Nature*, *391*(6669), 756.

Brand, M. (1984). *Intending and Acting: Toward a Naturalized Action Theory*. MIT Press.

Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica*, *106*(1-2), 3–22.

Bratman, M. (1984). Two Faces of Intention. *The Philosophical Review*, *93*(3), 375–405.

Bratman, M. (1987). *Intention, plans, and practical reason*. Harvard University Press.

Bratman, M. (1992). Shared Cooperative Activity. *The Philosophical Review*, *101*(2), 327–341.

Bratman, M. (1993). Shared Intention. *Ethics*, *104*(1), 97–113.

Bratman, M. (1997). I Intend That We J. In G. Holmström-Hintikka & R. Tuomela (Eds.), *Contemporary action theory: social action* (pp. 49–63). Dordrecht: Kluwer.
Bratman, M. (1999). *Faces of intention: selected essays on intention and agency*. Cambridge University Press.
Bratman, M. (2000). Reflection, Planning, and Temporally Extended Agency. *The Philosophical Review*, *109*(1), 35–61.
Bratman, M. (2006). Dynamics of sociality. *Midwest Studies in Philosophy*, *30*(1), 1–15.
Bratman, M. (2007). *Structures of Agency: Essays*. Oxford University Press.
Bratman, M. (2009a). Modest sociality and the distinctiveness of intention. *Philosophical studies*, *144*(1), 149–165.
Bratman, M. (2009b). Shared Agency. In C. Mantzavinos (Ed.), *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice* (pp. 41–59). Cambridge University Press.
Breheny, R. (2006). Communication and Folk Psychology. *Mind & Language*, *21*, 74–107.
Brinck, I., & Gärdenfors, P. (2003). Co–operation and Communication in Apes and Humans. *Mind & Language*, *18*(5), 484–501.
Brosnan, S. F., Salwiczek, L., Bshary, R., Brosnan, S. F., Salwiczek, L., & Bshary, R. (2010). The Interplay of Cognition and Cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1553), 2699–2710.
Brownell, C. (2011). Early Developments in Joint Action. *Review of Philosophy and Psychology*, *2*(2), 193–211.
Brownell, CA, & Carriger, M. (1990). Changes in Cooperation and Self-Other Differentiation during the Second Year. *Child Development*, *61*(4), 1164–1174.
Brownell, CA, Ramani, G., & Zerwas, S. (2006). Becoming a social partner with peers: cooperation and social understanding in one- and two-year-olds. *Child development*, *77*(4), 803–821.
Butterfill, S. A. (2011). *What is Joint Action? A modestly deflationary approach*. Unpublished manuscript.
Butterfill, S. A. (2012a). Joint action and development. *The Philosophical Quarterly*, *62*(246), 23–47.
Butterfill, S. A. (2012b). Interacting Mindreaders. *Philosophical Studies*, 1–27.
Butterfill, S. A., & Sinigaglia, C. (2012). Intention and Motor Representation in Purposive Action. *Philosophy and Phenomenological Research*. Advance online publication. doi:10.1111/j.1933-1592.2012.00604.x
Call, J., Hare, B., Carpenter, M., & Tomasello, M. (2004). "Unwilling"versus "unable": chimpanzees' understanding of human intentional action. *Developmental Science*, *7*(4), 488–498.
Campbell, J. (2002). *Reference and Consciousness*. Oxford University Press.
Campbell, J. (2005). Joint Attention and Common Knowledge. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint Attention: Communication and Other Minds* (pp. 287–297). Oxford University Press.
Carpenter, M. (2009). Just How Joint Is Joint Action in Infancy? *Topics in Cognitive Science*, *1*(2), 380–392.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant behavior & development*, *21*(2), 315–330.
Catmur, C., Gillmeister, H., Bird, G., Liepelt, R., & Brass, M. (2008). Through the looking glass: counter-mirror activation following incompatible sensorimotor learning. *European Journal of Neuroscience*, *28*(6), 1208–1215.
Chalmeau, R. (1994). Do chimpanzees cooperate in a learning task? *Primates*, *35*, 385–392.
Chalmeau, R., & Gallo, A. (1995). Cooperation in primates: Critical analysis of behavioural criteria. *Behavioural processes*, *35*(1-3), 101–111.
Chant, S., & Ernst, Z. (2008). Epistemic Conditions for Collective Action. *Mind*, *117*(467), 549–573.
Chartrand, T., & Bargh, J. (1999). The chameleon effect: the perception-behavior link and social interaction. *Journal of personality and social psychology*, *76*(6), 893–910.
Clark, A. (2008). *Supersizing The Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press.
Clark, H., & Marshall, C. (1981). Definite reference and mutual knowledge. In A. Joshi, B. Webber, & I. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge University Press.
Cohen, P.R., Levesque, H. J., & Smith, I. (1997). On Team Formation. In G. Holmström-Hintikka & R. Tuomela (Eds.), *Contemporary Action Theory Volume 2: Social Action*. Synthese.
Cohen, Philip R., & Levesque, H. J. (1991). Teamwork. *Noûs*, *25*(4), 487–512.
Csibra, G., & Gergely, G. (2007). "Obsessed with goals": Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, *124*(1), 60–78.
Cubitt, R. P., & Sugden, R. (2003). Common knowledge, Salience and Convention: A Reconstruction of David Lewis' Game Theory. *Economics and Philosophy*, *19*(2), 175–210.
Currie, G., & Ravenscroft, I. (2002). *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford: Oxford University Press.
Davidson, D. (1963). Actions, Reasons, and Causes. *The Journal of Philosophy*, *60*(23), 685–700.
Decety, J., Jeannerod, M., & Prablanc, C. (1989). The timing of mentally represented actions. *Behavioural Brain Research*, *34*(1–2), 35–42.
Denier van der Gon, J. J. (1988). Motor control: Aspects of its organization, control signals and properties. In *Proceedings of the Seventh Congress of the International Electrophysiological Society*. Amsterdam: Elsevier.
Dennett, D. C. (1998). *Brainchildren: Essays on Designing Minds*. MIT Press.
Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences*, *4*(11), 423–431.
Di Pellegrino, G., Làdavas, E., & Farné, A. (1997). Seeing where your hands are. *Nature*, *388*(6644), 730.
Drea, C., & Carter, A. (2009). Cooperative problem solving in a social carnivore. *Animal Behaviour*, *78*(4), 967–977.

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

Farne, A., Serino, A., & Ladavas, E. (2007). Dynamic size-change of peri-hand space following tool-use: determinants and spatial characteristics revealed through cross-modal extinction. *Cortex; a journal devoted to the study of the nervous system and behavior*, *43*(3), 436–443.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal Lobe: From Action Organization to Intention Understanding. *Science*, *308*(5722), 662–667.

Fourneret, P., & Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia*, *36*(11), 1133–1140.

Fowler, C., Richardson, M., Marsh, K., & Shockley, K. (2008). Language use, coordination, and the emergence of cooperative action. In A. Fuchs & V. K. Jirsa (Eds.), *Coordination: Neural, behavioral and social dynamics* (pp. 261–279). Springer.

Gallese, V. (2010). Of Goals and Intentions. In Grammont, Franck, Legrand, Dorothée, & Livet, Pierre (Eds.), *Naturalizing Intention in Action* (pp. 201–225). MIT Press.

Gärdenfors, P. (2007). Mind-reading as control theory. *European Review*, *15*(02), 223–240.

Georgeff, M., Pell, B., Pollack, M., Tambe, M., & Wooldridge, M. (1999). The belief-desire-intention model of agency. In J. P. Müller, A. S. Rao, & P. S. Munindar (Eds.), *Intelligent Agents V: Agents Theories, Architectures, and Languages* (pp. 1–10). Paris, France: Springer Verlag.

Gergely, G, & Csibra, G. (2003). Teleological reasoning in infancy: the naïve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292.

Gergely, György, Bekkering, H., & Király, I. (2002). Developmental psychology: Rational imitation in preverbal infants. *Nature*, *415*(6873), 755–755.

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. (2006). Rationality in Collective Action. *Philosophy of the Social Sciences*, *36*(1), 3–17.

Gilbert, M. (2008). Two Approaches to Shared Intention: An Essay in the Philosophy of Social Phenomena. *Analyse & Kritik*, *30*, 483–514.

Gilbert, M. (2009). Shared intention and personal intentions. *Philosophical studies*, *144*(1), 167–187.

Glüer, K., & Pagin, P. (2003). Meaning theory and autistic speakers. *Mind and Language*, *18*(1), 23–51.

Gold, N., & Sugden, R. (2007a). Collective Intentions and Team Agency. *Journal of Philosophy*, *CIV*(3), 109–137.

Gold, N., & Sugden, R. (2007b). Theories of team agency. In F. Peter & H. B. Schmid (Eds.), *Rationality and Commitment* (pp. 280–312). Oxford University Press.

Gräfenhain, M., Behne, T., Carpenter, M., & Tomasello, M. (2009). Young children's understanding of joint commitments. *Developmental Psychology*, *45*(5), 1430–1443.

Gray, J. (1978). Notes on data base operating systems. *Operating Systems*, 393–481.
Grush, R. (1997). The architecture of representation. *Philosophical Psychology*, *10*(1), 5–23.
Grush, R. (2003). In defense of some "Cartesian" assumptions concerning the brain and its operation. *Biology and Philosophy*, *18*, 33–93.
Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *The Behavioral and Brain Sciences*, *27*(3), 377–96; discussion 396–442.
Hamann, K., Warneken, F., Greenberg, J. R., & Tomasello, M. (2011). Collaboration encourages equal sharing in children but not in chimpanzees. *Nature*, *476*(7360), 328–331.
Hamann, K., Warneken, F., & Tomasello, M. (2011). Children's Developing Commitments to Joint Goals. *Child Development*, *83*(1), 137–145.
Harris, P., Kavanaugh, R. D., Wellman, H. M., & Hickling, A. K. (1993). Young children's understanding of pretense. *Monographs of the Society for Research in Child Development*, *58*(1), 1–107.
Harrison, S. J., & Richardson, M. J. (2009). Horsing Around: Spontaneous Four-Legged Coordination. *Journal of Motor Behavior*, *41*(6), 519–524.
Hay, D. (1978). Cooperative interactions and sharing between very young children and their parents. *Developmental Psychology*, *15*(6), 647–653.
Heal, J. (1978). Common Knowledge. *The Philosophical Quarterly*, *28*(111), 116–131.
Henderson, A. M. E., & Woodward, A. L. (2011). "Let's work together": What do infants understand about collaborative goals? *Cognition*, *121*(1), 12–21.
Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews*, *34*(4), 575–583.
Hill, K. (2002). Altruistic cooperation during foraging by the Ache, and the evolved human predisposition to cooperate. *Human Nature*, *13*(1), 105–128.
Holmes, N., Calvert, G., & Spence, C. (2004). Extending or projecting peripersonal space with tools? Multisensory interactions highlight only the distal and proximal ends of tools. *Neuroscience Letters*, *372*(1-2), 62–67.
Holmes, N., Sanabria, D., Calvert, G., & Spence, C. (2007). Tool-use: capturing multisensory spatial attention or extending multisensory peripersonal space? *Cortex; a journal devoted to the study of the nervous system and behavior*, *43*(3), 469–489.
Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychological Research Psychologische Forschung*, *73*(4), 512–526.
Hughes, C., Fujisawa, K. K., Ensor, R., Lecce, S., & Marfleet, R. (2006). Cooperation and conversations about the mind: A study of individual differences in 2-year-olds and their siblings. *British Journal of Developmental Psychology*, *24*(1), 53–72.
Hurley, S. L. (1998). *Consciousness in Action*. Harvard University Press.
Hutto, D. (2012). Elementary Mind Minding, Enactivist-Style. In A. Seemann (Ed.), *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience* (pp. 307–341). MIT Press
Itō, M. (1984). *The cerebellum and neural control*. New York: Raven Press.
Jacob, P. (2009). A Philosopher's Reflections on the Discovery of Mirror Neurons. *Topics in Cognitive Science*, *1*(3), 570–595.

Jacob, P. (2012). Sharing and Ascribing Goals. *Mind & Language*, *27*(2), 200–227.
Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery, *17*(2), 187–202.
Keijzer, F. (2012). The Sphex story: How the cognitive sciences kept repeating an old and questionable anecdote. *Philosophical Psychology*, 1–18. Advance online publication. doi:10.1080/09515089.2012.690177
Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. MIT Press.
Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological movement on action. *Current Biology*, *13*(6), 522–525.
Knoblich, G, Butterfill, S., & Sebanz, N. (2011). Psychological Research on Joint Action: Theory and Data. In B. Ross (Ed.), *The Psychology of Learning and Motivation*, Vol. 54 (pp. 59–101). Burlington: Academic Press.
Knoblich, G., & Jordan, J. S. (2002). The mirror system and joint action. *Advances in Consciousness Research*, *42*, 115–124.
Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1499), 2021–2031.
Kutz, C. (2000). Acting Together. *Philosophy and Phenomenological Research*, *61*(1), 1–31.
Lewis, D. K. (1969). *Convention*. Cambridge, MA, USA: Harvard University Press.
Ludwig, K. (2007). Collective Intentional Behavior from the Standpoint of Semantics. *Nous*, *41*(3), 355–393.
Mackie, J. L. (1965). Causes and Conditions. *American Philosophical Quarterly*, *2*(4), 245–264.
Maravita, A., & Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences*, *8*(2), 79–86.
Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., & Flavell, J. H. (1974). The Early Development of Inferences about the Visual Percepts of Others. *Child Development*, *45*(2), 357–366.
Melis, A. P., Hare, B., & Tomasello, M. (2006). Chimpanzees recruit the best collaborators. *Science*, *311*(5765), 1297–1300.
Mendres, K. A., & De Waal, F. B. M. (2000). Capuchins do cooperate: the advantage of an intuitive task. *Animal Behaviour*, *60*(4), 523–529.
Merleau-Ponty, M. (2002). *Phenomenology of Perception*. (C. Smith, Trans.). London: Routledge. (Original work published 1945).
Miller, K., & Tuomela, R. (2001). What are Collective Goals? In M. Kiikeri & P. Ylikoski (Eds.), *Explanatory Connections: electronic essays dedicated to Matti Sintonen, http://www.helsinki.fi/tint/matti/*. Retrieved from http://www.helsinki.fi/tint/matti/miller.pdf
Miller, S. (1986). Conventions, Interdependence of Action, and Collective Ends. *Nous*, *20*(2), 117–140.
Miller, S. (1995). Intentions, Ends and Joint Action. *Philosophical Papers*, *24*(1), 51–66.
Miller, S. (2001). *Social Action: A Teleological Account*. Cambridge University Press.
Millikan, R. G. (2004). *Varieties of Meaning: The 2002 Jean Nicod Lectures*. MIT Press.

Milner, D., & Goodale, M. (2006). *The Visual Brain in Action* (2nd ed.). Oxford University Press.
Moll, H., & Tomasello, M. (2007). Cooperation and human cognition: the Vygotskian intelligence hypothesis. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*, *362*(1480), 639–648.
Moore, C., & Dunham, P. J. (Eds.). (1995). *Joint Attention*. Psychology Press.
Morton, A. (1996). Folk Psychology is not a Predictive Device. *Mind*, *105*(417), 119–137.
Naderi, S., Miklósi, Á., Dóka, A., & Csányi, V. (2001). Co-operative interactions between blind persons and their dogs. *Applied Animal Behaviour Science*, *74*(1), 59–80.
Newman-Norlund, R. D., Van Schie, H. T., Van Zuijlen, A. M. J., & Bekkering, H. (2007). The mirror neuron system is more active during complementary compared with imitative action. *Nature Neuroscience*, *10*(7), 817–818.
Noë, R. (2006). Cooperation experiments: coordination through communication versus acting apart together. *Animal Behaviour*, *71*(1), 1–18.
Nøttestad, L., Fernö, A., & Axelsen, B. (2002). Digging in the deep: killer whales' advanced hunting tactic. *Polar Biology*, *25*(12), 939–941.
Onishi, K., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, 308(5719), 255–258.
Pacherie, E. (2000). The Content of Intentions. *Mind & Language*, *15*, 400–432.
Pacherie, E. (2007). Is Collective Intentionality Really Primitive? In M. Beaney, C. Penco, & M. Vignolo (Eds.), *Mental processes: representing and inferring* (pp. 153–175). Cambridge: Cambridge Scholars Press.
Pacherie, E. (2008). The phenomenology of action: a conceptual framework. *Cognition*, *107*(1), 179–217.
Pacherie, E. (2011). Framing Joint Action. *Review of Philosophy and Psychology*, 1–20.
Pacherie, E. (2012). The Phenomenology of Joint Action: Self-Agency vs. Joint-Agency. In A. Seemann (Ed.), *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience* (pp. 343–389). MIT Press.
Pacherie, E., & Dokic, J. (2006). From mirror neurons to joint actions. *Cognitive Systems Research*, *7*(2-3), 101–112.
Parsons, L. M., Fox, P. T., Downs, J. H., Glass, T., Hirsch, T. B., Martin, C. C., Jerabek, P. A., Lancaster, J. L. (1995). Use of implicit motor imagery for visual shape discrimination as revealed by PET. *Nature*, *375*(6526), 54–58.
Paternotte, C. (2010). Being realistic about common knowledge: a Lewisian approach. *Synthese*, 1–28.
Peacocke, C. (2005). Joint Attention: Its Nature, Reflexivity, and Relation to Common Knowledge. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint Attention: Communication and Other Minds* (pp. 298–323). Oxford University Press.
Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, *91*(1), 176–180.
Perner, J., & Roessler, J. (2010). Teleology and causal understanding in childrens' theory of mind. In J. H. Aguilar & A. A. Buckareff (Eds.), *Causing Human*

Action: New Perspectives on the Causal Theory of Action (pp. 342–395). MIT Press.
Perry, J. (1979). The Problem of the Essential Indexical. *Noûs*, *13*(1), 3–21.
Petersson, B. (2007). Collectivity and circularity. *Journal of Philosophy*, 138–156.
Petersson, B. (2011). Project application: Intentional agency and agent perspectives. Retrieved from http://internt.ht.lu.se/media/documents/project-192/RJ_12_14.pdf
Petit, O., Desportes, C., & Thierry, B. (1992). Differential probability of "coproduction" in two species of macaque (Macaca tonkeana, M. mulatta). *Ethology*, *90*(2), 107–120.
Pettit, P., & Schweikard, D. (2006). Joint Actions and Group Agents. *Philosophy of the Social Sciences*, *36*(1), 18–39.
Phillips, W., Barnes, J. L., Mahajan, N., Yamaguchi, M., & Santos, L. R. (2009). "Unwilling" versus "unable": capuchin monkeys' (Cebus apella) understanding of human intentional action. *Developmental Science*, *12*(6), 938–945.
Pika, S., & Zuberbühler, K. (2008). Social games between bonobos and humans: evidence for shared intentionality? *American Journal of Primatology*, *70*(3), 207–210.
Pols, A. (2011). *Acting with Artefacts* (Doctoral dissertation). Technische Universiteit Eindhoven.
Preston, B. (2013). *A Philosophy of Material Culture: Action, Function, and Mind*. Routledge.
Prinz, W. (1997). Perception and Action Planning. *European Journal of Cognitive Psychology*, *9*(2), 129–154.
Proust, J. (2003). Action. In B. Smith (Ed.), *John Searle* (pp. 102–127). Cambridge University Press.
Rakoczy, H. (2006). Pretend play and the development of collective intentionality. *Cognitive Systems Research*, *7*, 113–127.
Rakoczy, H. (2008). Pretence as Individual and Collective Intentionality. *Mind and Language*, *23*(5), 499–517.
Recanati, F. (2007). Content, mode, and self-reference. In S. L. Tsohatzidis (Ed.), *John Searle's Philosophy of Language: Force, Meaning, and Mind*. Cambridge University Press.
Richardson, M., Marsh, K., Isenhower, R., Goodman, J., & Schmidt, R. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, *26*(6), 867–891.
Risjord, M. (2012). *Improvised Joint Performance*. Unpublished manuscript.
Roth, A. (2003). Practical Intersubjectivity. In F. F. Schmitt (Ed.), *Socializing Metaphysics: The Nature of Social Reality* (pp. 65–91). Rowman & Littlefield.
Roth, A. (2004). Shared Agency and Contralateral Commitments. *The Philosophical Review*, *113*(3), 359–410.
Roth, A. (2011). Shared Agency. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*. Retrieved from http://plato.stanford.edu/archives/spr2011/entries/shared-agency/

Rubinstein, A. (1989). The Electronic Mail Game: Strategic Behavior Under "Almost Common Knowledge". *The American Economic Review*, *79*(3), 385–391.
Rupert, R. (2009). *Cognitive Systems and the Extended Mind*. Oxford University Press.
Schiffer, S. R. (1972). *Meaning*. Oxford University Press.
Schmidt, R., & Richardson, M. (2008). Dynamics of interpersonal coordination. In A. Fuchs & V. K. Jirsa (Eds.), *Coordination: Neural, behavioral and social dynamics* (pp. 281–308). Springer.
Schöner, G., Jiang, W. Y., & Kelso, J. A. S. (1990). A synergetic theory of quadrupedal gaits and gait transitions. *Journal of Theoretical Biology*, *142*(3), 359–391.
Searle, J. R. (1980). The intentionality of intention and action. *Cognitive Science*, *4*(1), 47–70.
Searle, J. R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press.
Searle, J. R. (1990). Collective Intentions and Actions. In P. R. Cohen, J. Morgan, and M. E. Pollack (Eds.), *Intentions in Communication* (pp. 401–415). MIT Press.
Searle, J. R. (1992). *The Rediscovery of the Mind*. MIT Press.
Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76.
Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition*, *88*(3), 11–21.
Sebanz, N., Knoblich, G., & Prinz, W. (2005). How Two Share a Task: Corepresenting Stimulus-Response Mappings. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1234–1246.
Seed, A. M., Clayton, N. S., & Emery, N. J. (2008). Cooperative problem solving in rooks (Corvus frugilegus). *Proceedings of the Royal Society B: Biological Sciences*, *275*(1641), 1421–1429.
Seemann, A. (2009a). Why We Did It: An Anscombian Account of Collective Action. *International Journal of Philosophical Studies*, *17*(5), 637–655.
Seemann, A. (2009b). Joint Agency: Intersubjectivity, Sense of Control, and the Feeling of Trust. *Inquiry*, *52*(5), 500–515.
Seemann, A. (Ed.). (2012). *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*. MIT Press.
Shockley, K., Richardson, D., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science*, *1*(2), 305–319.
Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*(3), 300–304.
Skyrms, B. (2009). Commentary on Tomasello's "Why We Cooperate". In *Why We Cooperate* (pp. 137–146). MIT Press.
Sober, E., & Wilson, D. S. (1998). *Unto Others: Evolution and Psychology of Unselfish Behavior*. Harvard University Press.
Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Harvard University Press.

Stander, P. E. (1992). Cooperative hunting in lions: the role of the individual. *Behavioral Ecology and Sociobiology*, *29*(6), 445–454.
Stephan, K. M., Fink, G. R., Passingham, R. E., Silbersweig, D., Ceballos-Baumann, A. O., Frith, C. D., & Frackowiak, R. S. (1995). Functional anatomy of the mental representation of upper extremity movements in healthy subjects. *Journal of Neurophysiology*, *73*(1), 373–386.
Sterelny, K. (2003). *Thought in a Hostile World: The Evolution of Human Cognition*. Wiley-Blackwell.
Sterelny, K. (2004). Externalism, Epistemic Artefacts and the Extended Mind. In R. Schantz (Ed.), *The Externalist Challenge* (pp. 239–254). Walter de Gruyter.
Sterelny, K. (2010). Minds: extended or scaffolded? *Phenomenology and the Cognitive Sciences*, *9*(4), 465–481.
Stürmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence effects with manual gestures and postures: A study of imitation. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1746–1759.
Tanner, J., & Byrne, R. (2010). Triadic and collaborative play by gorillas in social games with objects. *Animal Cognition*, *13*(4), 591–607.
Tebbich, S., Taborsky, M., & Winkler, H. (1996). Social manipulation causes cooperation in keas. *Animal Behaviour*, *52*(1), 1–10.
Tollefsen, D. (2005). Let's Pretend! Children and Joint Action. *Philosophy of the Social Sciences*, *35*(1), 75–97.
Tollefsen, D., & Dale, R. (2012). Naturalizing joint action: A process-based approach. *Philosophical Psychology*, *25*(3), 385–407.
Tomasello, M. (2009). *Why We Cooperate*. MIT Press.
Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees understand psychological states-the question is which ones and to what extent. *Trends in Cognitive Sciences*, *7*(4), 153–156.
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*(5), 675–691.
Tomasello, M., & Hamann, K. (2012). Collaboration in young children. *The Quarterly Journal of Experimental Psychology*, *65*(1), 1–12.
Tsai, C.-C, Kuo, W., Hung, D., & Tzeng, O. (2008). Action co-representation is tuned to other humans. *Journal of Cognitive Neuroscience*, *20*(11), 2015–24.
Tsai, C.-C., Sebanz, N., & Knoblich, G. (2011). The GROOP effect: groups mimic group actions. *Cognition*, *118*(1), 135–40.
Tuomela, R. (2005). We-intentions revisited. *Philosophical Studies*, *125*(3).
Tuomela, R. (2007). *The Philosophy of Sociality: The Shared Point of View*. Oxford University Press.
Tuomela, R., & Miller, K. (1988). We-intentions. *Philosophical Studies*, *53*(3), 367–389.
Turvey, M. (1990). Coordination. *The American psychologist*, *45*(8), 938–53.
Vanderschraaf, P., & Sillari, G. (2009). Common knowledge. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Spring 2009 Edition)*. Retrieved from http://plato.stanford.edu/archives/spr2009/entries/common-knowledge/
Velleman, J. D. (1991). Book Review: Intention, Plans and Practical Reason. *The Philosophical Review*, *100*(2), 277–284.

Velleman, J. D. (1997). How To Share An Intention. *Philosophy and Phenomenological Research*, *57*(1), 29–50.
Velleman, J. D. (2007). What good is a will? In A. Leist (Ed.), *Action in context* (pp. 193–215). Berlin: Walter de Gruyter.
Verba, M. (1993). The beginnings of collaboration in peer interaction. *Human Development*, *37*(3), 125–139.
Visalberghi, E., Quarantotti, B., & Tranchida, F. (2000). Solving a cooperation task without taking into account the partner's behavior: the case of capuchin monkeys (Cebus apella). *Journal of comparative psychology*, *114*(3), 297–301.
Ward, P. I., & Enders, M. M. (1985). Conflict and cooperation in the group feeding of the social spider Stegodyphus mimosarum. *Behaviour, 94*, *1*(2), 167–182.
Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child development*, *77*(3), 640–63. doi:10.1111/j.1467-8624.2006.00895.x
Warneken, F., & Gräfenhain, M. & Tomasello, M. (2012). Collaborative partner or social tool? New evidence for young children's understanding of joint intentions in collaborative activities. *Developmental Science*, *15*(1), 54–61.
Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child development*, *77*(3), 640–663.
Watson, J. S., & Ramey, C. T. (1987). Reactions to response-contingent stimulation in early infancy. In J. Oates & S. Sheldon (Eds.), *Cognitive Development in Infancy* (pp. 77–85). The Open University.
Wellman, H., Cross, D., & Watson, J. (2000). Meta-analysis of theory-of-mind development: The truth about false belief. *Child development*, *72*(3), 655–684.
Welsh, T. N., Higgins, L., Ray, M., & Weeks, D. J. (2007). Seeing vs. believing: Is believing sufficient to activate the processes of response co-representation? *Human Movement Science*, *26*(6), 853–866. doi:10.1016/j.humov.2007.06.003
Wenke, D., Atmaca, S., Holländer, A., Liepelt, R., Baess, P., & Prinz, W. (2011). What is Shared in Joint Action? Issues of Co-representation, Response Conflict, and Agent Identification. *Review of Philosophy and Psychology*, *2*(2), 147–172.
Wilby, M. (2010). The simplicity of mutual knowledge. *Philosophical Explorations*, *13*(2), 83–100.
Wolpert, D., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*, *358*(1431), 593–602.
Wolpert, D., Miall, R., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, *2*(9), 338–347.
Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34.
Wooldridge, M. (2000). *Reasoning about rational agents*. MIT Press.
Zawidzki, T. (2008). The function of folk psychology: mind reading or mind shaping? *Philosophical Explorations*, *11*(3), 193–210.