Studies in the Philosophy of Sociality 11

Anika Fiebich *Editor*

Minimal Cooperation and Shared Agency

Springer

Studies in the Philosophy of Sociality

Volume 11

First book series in Philosophy of the Social Sciences that specifically focuses on Philosophy of Sociality and Social Ontology.

Covers a new and rapidly developing field that has become one of the key topics of the international philosophical world.

Aims at an interdisciplinary approach that will bring new perspectives to the study of such topics as communication, unintended consequences of action as well as social structures and institutions.

This book series publishes research devoted to the basic structures of the social world. The phenomena it focuses on range from small scale everyday interactions to encompassing social institutions, from unintended consequences to institutional design. The unifying element is its focus on the basic constitution of these phenomena, and its aim to provide philosophical understanding on the foundations of sociality. Studies in the Philosophy of Sociality covers the part of philosophy of the social sciences which deals with questions of social ontology, collective intentionality (e.g. collective knowledge, joint and collective action, joint mental states) and related philosophical topics. The series will include monographs and edited collections on philosophical and conceptual questions concerning social existence as well as conceptual and theoretical analyses of social notions and collective epistemology.

In principle, all phenomena dealing with sociality are covered as long as they are approached from a philosophical point of view, broadly understood. Accordingly, the works to be published in the series would in general be philosophical—without regard to philosophical schools and viewpoints—and they would meet the highest academic and intellectual standards are met. However, the series is interdisciplinary not only in an intra-philosophical sense but also in the sense of encouraging high-level work from other disciplines to be submitted to the series. Others who are active in the field are political scientists, economists, sociologists, psychologists, linguists, neuroscientists, evolutionary biologists, and researchers of artificial intelligence.

The resulting interdisciplinary approach will give new perspectives to the study of topics such as social interaction, communication, unintended consequences of action, social structures and institutions, the evolution of collective intentionality phenomena, as well as the general ontological architecture of the social world.

The series discourages the submission of manuscripts that contain reprints of previous published material and/or manuscripts that are below 150 pages / 75,000 words.

For inquiries and submission of proposals authors can contact christi.lue@springer.com, or contact one of the associate editors.

More information about this series at http://www.springer.com/series/10961

Anika Fiebich
Editor

Minimal Cooperation and Shared Agency

Springer

Editor
Anika Fiebich
Bremen, Germany

ISSN 2542-9094 ISSN 2542-9108 (electronic)
Studies in the Philosophy of Sociality
ISBN 978-3-030-29782-4 ISBN 978-3-030-29783-1 (eBook)
https://doi.org/10.1007/978-3-030-29783-1

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Chapter 1
Minimal Cooperation (Editorial Introduction)

Anika Fiebich

Abstract This chapter is an editorial introduction, which gives an overview of the single contributions to that volume. It will illustrate that 'minimality' in cooperation can be understood in various ways and that a distinction between internalist versus externalist approaches to cooperation explains why some of the accounts defended in this volume are incompatible with each other.

Keywords Cooperation · Minimality · Internalism · Externalism

This volume brings together top scholars in the philosophy of mind and action to discuss minimality in cooperation. In general, 'minimality' can be understood in a diverse number of ways. It can refer, for example, to the minimal or necessary criteria for cooperation or, alternatively, 'minimal' can be understood in the sense of 'basic' or 'simple'. Elucidating the variety of ways in which minimality can be interpreted in the analysis of cooperation provides a genuine contribution to the contemporary joint action debate. Notably, the main accounts in this debate (e.g., by Michael Bratman, Margaret Gilbert, Raimo Tuomela or John Searle) provide sufficient rather than necessary or minimal criteria for cooperation. A lot of discussion in the debate deals with robust rather than simpler and more attenuate cases of cooperation. Focusing on such minimal cases, however, may help to explain how cooperation comes into existence and how simple cases interrelate with more complex ones. Minimality in cooperation can be captured by focusing on particular aspects of cooperation, such as minimal social cognitive skills and competencies, or on how social roles and group memberships might deliver minimal cooperation constraints. Finally, discussing minimality in cooperation with respect to the peculiarities of different kinds of mental actions involved in cooperative activities (e.g., team reasoning or collective remembering) is also a fruitful enterprise.

A. Fiebich (✉)
Bremen, Germany
e-mail: ani.fiebich@web.de

A. Fiebich (ed.), *Minimal Cooperation and Shared Agency*, Studies in the Philosophy of Sociality 11, https://doi.org/10.1007/978-3-030-29783-1_1

The volume captures minimality in cooperation in the following ways:

(i) *From Minimal to Full Cooperation*

Starting from an account of full cooperation, Kirk Ludwig arrives at an account of minimal cooperation by asking what features can be discarded while at the same time retaining some form of social cooperation. Ludwig argues that cooperative behavior essentially involves cooperating agents who intend to contribute with one another to achieve a shared goal (i.e., agents who share an intention in a distributive sense). For this reason, cooperative behaviour involves, but goes beyond, mere coordinated behaviour. On this approach, minimally joint intentional action coincides with minimally cooperative behavior. Ludwig discusses four features that are jointly sufficient for fully cooperative behavior: (i) shared intention, (ii) absence of conflict, (iii) willingness to aid, and (iv) full effort. If (i) is the minimal condition, the failure of (ii)–(iii) should be compatible with cooperation. Competitive games and verbal arguments violate (ii) but nonetheless count as cooperative activities. Violations of (iii) are also compatible with cooperation and shared intention. One may intentionally sing a duet with another person in the absence of any error without being willed to help if she falters. One may even interfere with another's part (to promote one's own) without intending to undermine the joint effort. An example of a violation of (iv) is a work slowdown. This is a case in which full effort is withheld but workers have not ceased to cooperate with management entirely. All the cases in which cooperation occurs, Ludwig argues, include shared intention as their common feature, and there are no cases of cooperation in which it is missing. In addition, Ludwig argues that minimal cooperation does not require rationality and is compatible with coercion. Ludwig ends by comparing his account with 'game-theoretic cooperation', arguing that it is not the target notion of social cooperation, but at best provides rational underpinnings for shared intention (and hence genuine cooperation) in solving coordination problems.

(ii) *Context and Emergence of Minimal Cooperation*

Cédric Paternotte starts by discussing various ways in which minimality in cooperation can be understood by referring to (i) scale minimalism (i.e., cooperation involving a minimal number of interchangeable agents), (ii) conceptual minimalism (i.e., cooperation that builds upon a minimal conceptual core), (iii) ontological minimalism (i.e., cooperation involving a minimal number of entities needed for its characterization); and (iv) cognitive minimalism (i.e., cooperation requiring minimal cognitive skills and abilities). Reviewing a number of minimal approaches to cooperation in the literature, Paternotte illustrates that although these kinds of minimalism are not mutually exclusive, they are rarely all adopted simultaneously. Moreover, the different kinds of minimalism allow for pluralism. For example, an account may be minimal in a conceptual sense, because it essentially builds on the concept of joint commitment, while another account counts as conceptually minimal, because it does not allow for subtracting a specific mental attitude (e.g., a we-intention). Paternotte embraces pluralism even further by introducing what he calls a 'contextually minimalist approach' that focuses on cases of cooperation that occur in contexts in which

those factors that typically favor cooperative behavior are absent. Examples such as demonstrations, deliberation, and collective free improvisation are not minimal in either of the traditional senses. On the contextually minimal approach, however, such cases are minimal with respect to the contexts in which they occur.

Shaun Gallagher discusses the minimal conditions, resources and processes involved in the emergence of cooperation that is not preceded by the formation of a distal or prior shared intention. Communicative practices, like bodily gestures and speech acts, are first-order practices: they are not governed by reflective awareness or explicit planning, but are second-person practices, i.e., they are most often performed for others in social interaction. Gallagher discusses cases in which the meaning of such practices arises through the dynamics of the social interaction (e.g., when playing hopscotch together). Here he draws on Goldstein's notion of 'gestalt' according to which the relation between abstract and concrete cognitive processes should be understood on a foreground/background gestalt model rather than on a hierarchical one. On this account, the dynamical relations shape the relata in a way that cognitive processes like memory or imagination are not separated from the social interaction, but are continuous with, and intervene in, first-order communicative practices.

(iii) *Minimal Cooperation and Group Membership*

Glenda Satne and Alessandro Salice challenge Michael Tomasello's take on the 'Cooperative Evolutionary Hypothesis' (CEH) by providing an alternative minimal collective view. According to CEH, human-specific skills for cooperation draw essentially on the capacity to understand behavior in terms of mental states. Tomasello distinguishes between two forms of social intentionality as being part of a two-step evolutionary story: first, cooperation relying on joint intentionality in small-scale groups, which draws upon joint attention and coordination, and second, collective intentionality in large-scale groups that involves establishing a cultural common ground. The authors argue that social intentionality is collective from the onset with different stages of development, and point to a number of difficulties that Tomasello's account of joint intentionality bears in the light of developmental research. Following Michael Bratman, Tomasello argues that cooperation in small-scale groups relies on joint intentionality in which two (or more) agents share a goal to cooperate in accordance with, and because of, each other's sharing that goal, which is mutual common knowledge between them. This definition is cognitively demanding since the possession of mutual common knowledge presupposes a capacity for building meta-representations and for attributing mental states to oneself and others. Moreover, it excludes preverbal infants' helping behaviors as an instance of joint intentionality and cooperation. Reviewing a number of developmental findings, Satne and Salice argue that cooperation is present in ontogeny long before the capacity for mental-state attribution develops. On the minimal collective view, infants' helping behavior is a case of cooperation that involves social intentionality in virtue of having a sense of engagement and group membership.

Katherine Ritchie's defence of an externalist approach to minimal cooperation stands in opposition to the traditional internalist idea that cooperation essentially depends on mental factors like collective or shared intentions. On her approach,

minimal cooperation occurs when the members of a group play roles in an organizational structure and share a common goal. While Ritchie acknowledges that internalist approaches may be valid in small-scale cases of cooperation, she argues that they cannot be scaled up to large-scale cooperative activities. Her focus is on large-scale cases of cooperation, such as legislative bodies or corporate teams. In such cases group members may not be acquainted with other members or their mental states, but are nonetheless cooperatively pursuing a common goal. Notably, not even knowledge of the common goal may be required for there to be minimal cooperation in a large-scale case. The members of a spy network, for example, may not know about the goal of the network but could be minimally cooperating in virtue of their roles being functionally integrated to achieve an end. Ritchie highlights that minimal cooperation requires playing roles in an organized group structure, which rely on norms and obligations. This normative dimension of cooperation rules out that the parts of a car engine functioning in concord can be regarded as a case of minimal cooperation.

(iv) *Minimal (Social) Cognition in Cooperation*

Stephen Butterfill and Elisabeth Pacherie isolate some cognitively undemanding forms of cooperation that different hypothetical creatures, who possess divergent minimal social kills and cognitive abilities, are able to perform. They start by postulating a creature called 'Alphonso' and his kin, whose social cognitive skills are limited to tracking other people's action goals, which allows them to coordinate their own action towards a joint goal (e.g., moving a boulder together). Alphonso and his kin act on that goal, which is most salient; and where there is none, they simply pick up a goal, which is not too costly to end up acting alone, and pursue it under the assumption that their kins will join. On some occasions, however, it is too dangerous to act on that assumption (e.g., when hunting a large pig). In these cases salience is required as a proxy for common knowledge. Salience is induced, for example, by stimuli like loud noises ('salience heuristic'), and if an event is not salient enough, the characteristic response of others may work as a proxy ('triangulation heuristic'). The salience heuristic may fail, however, in situations, for example, where there are two pigs and neither is more salient than the other. In such cases, communicative skills such as gesturing and pointing are required, which are possessed by Alphonso's descendants called 'Beki'. The triangulation heuristic, in turn, may fail on occasions where it may be beneficial for the individual to hide a characteristic response for others. In cases such as this, deceptive skills are required, which are possessed by Alphonso's descendants called 'Bemi'. Communicative and deceptive skills are possessed by creatures called 'Kimi', who were born by parents stemming from both groups. Butterfill and Pacherie highlight that these examples illustrate that groups who do not possess sophisticated cognitive capacities, like belief reasoning, needed for joint action according to classical accounts, are still capable of engaging in cooperation based on minimal cognitive skills and different proxies.

Michael Wilby distinguishes between three families of collaborative activities: (i) massively shared agency, (ii) modest sociality, and (iii) minimal cooperation. Minimal cooperation can be basic in a sense that it is performed as an end in itself, or it can be a means to execute elements of plans within modest sociality such as proposed by Bratman. Wilby focusses on the latter and discusses the social cognitive

capacities that infants need to possess in order to engage in such cooperation. Coining the term Natural Intersubjectivity, Wilby argues for joint attention having an explanatory role and rationally coordinates function in dyadic cooperative activities in accordance with a prior shared intention. Focusing on cooperatively loaded actions, in which coordination is essential to their fulfillment, shows that when we share the intention to make a Hollandaise Sauce together, we need to identify the objects, locations, processes etc. that are required for our carrying out that intention, and we need to coordinate on the details (e.g., who is pouring). Paying joint attention to such identification and coordination allows for the realization of our shared planning.

(v) *Minimalism in Team Reasoning and Solution Thinking*

Francesco Guala explores minimality in cooperation with respect to the social cognitive demands that are required to engage successfully in different kinds of economic games. Defending what he calls the 'minimal mind-reading' (MMR) program, he argues that successful cooperation involves belief-less reasoning and that reasoning about beliefs often hampers rather than facilitates successful engagement in a coordinative activity. Using the example of the Hi-lo game, Guala points to empirical findings that challenge the assumption that people use the 'primary salience' principle to predict another agent's behavior, i.e., they choose the solution, which comes first to mind. Rather, they appear to use the 'Schelling salience' principle by converging on the solution, which is most likely to solve the particular problem they face. Guala illustrates the application of that principle in team reasoning and solution thinking, highlighting that rational actions can be deduced from the goal of the agents in team reasoning, and that identifying the solution helps to infer the individual strategies in solution thinking. In both cases reasoning about the other agent's beliefs is not required. Guala ends by emphasising that belief-less reasoning is not only successful in economic games, but also cognitively less demanding and ontogenetically prior to cooperation that requires reasoning about other agents' beliefs. In line with pluralist theories of social cognition, he argues that people are typically engaged in belief-less cooperation in everyday life.

Abraham Roth asks whether or not the team reasoning conception of joint action allows for cases of minimal cooperation, where participants act jointly but are largely motivated by proprietary (i.e., private) reasons not shared by others. On the team reasoning conception, the members of a team are reasoning about what they should do together. That is, when acting on a collective intention stemming from team reasoning, they are acting for the corresponding collective reasons. This does not seem to rule out that the agents act for proprietary reasons as well. But is this a real possibility on the team reasoning view? Roth argues that the team reasoning view will often rule out proprietary reasons in joint action. Sometimes they can be allowed, but only when they figure as reasons for how the individuals act within a framework that is established by team reasoning. According to Roth, on the team reasoning view, proprietary reasons are subordinate to shared or collective reasons. Surprisingly, the team reasoning view cannot countenance the notion of minimal joint action understood in terms of participants largely motivated by proprietary reasons. As soon as proprietary reasons figure as reasons in a non-subordinate manner, individuals are no longer genuinely acting with each other.

(vi) *Experiences and Mental Agency in Minimal Cooperation*

Corrado Sinigaglia and Stephen Butterfill discuss the role of motor representations in the experiences of agents who perform a joint action, i.e. an event involving two or more agents acting together with a purpose. They focus on experiences of action that reveal what is being done rather than those that reveal who is acting. Arguing for a parallel between individual and joint action, Sinigaglia and Butterfill start by elucidating the role of motor representations in individual actions. Motor representations specify action outcomes but, unlike intentions, they only concern present actions and they are tied to the body and its capacities to act. Reviewing a number of neuro-scientific and behavioral studies, the authors show that motor representations both cause and shape the experiences of individual actions by influencing, for example, judgments about actions and their consequences. Drawing on a recent study, which shows that collective goals can be represented motorically, they hypothesize that motor representations of a collective goal of joint action may also cause and shape the experiences of agents acting together.

Santiago Arango-Muñoz and Kourken Michaelian argue for collective remembering as a form of shared mental agency that can be minimal in various ways. For example, the collective reconstructed memories may have some, but not necessary all, episodic features. Episodic features include, but are not limited to, such things as color, shape, emotion, first-person perspective, intensity and phenomenology. The authors point to findings from memory research demonstrating that genuinely collective episodic memory is present in some long-married couples and other small-scale groups, but not in large-scale groups such as societies. Moreover, the authors argue that collective memory implies collective metamemory but that the form of metacognition at work in emergent collective memory is minimal insofar as it needs to be understood as an individual-level capacity rather than an emergent group-level capacity itself.

The distinction between externalism and internalism is helpful to elucidate the different target questions that are addressed when accounting for minimality in cooperation (Fig. 1.1). The question ‘wherefore do we cooperate?’ lies at the heard of any approach to cooperation. Externalists answer that question by referring to the notion of a common goal that the agents pursue together. Notably, ‘common goal’ is understood here teleologically rather than mentalistically. Internalists, in contrast, argue that the cooperating agents pursue a shared goal or shared intention, which presupposes mutual common awareness or mutual common knowledge of having such a goal or intention as well as social cognitive skills such as tracking goals or attributing mental states to oneself and others. Whereas externalists focus in their analysis of cooperation primarily on external characteristics (e.g., the context in which an instance of cooperation takes place), internalists pay less attention to external factors but are occupied with what is going on in the mind of the cooperating agents (e.g., which reasons motivate them to cooperate).

The contributions to this volume illustrate that minimalism in cooperation can be defended in various ways, potentially leading to divergent outcomes. An externalist approach to cooperation, for example, may argue for minimal cooperation in virtue of the members of a group pursuing a common goal by playing roles in an

EXTERNALISM	INTERNALISM
Wherefore? common goal	*Wherefore?* shared goal or shared intention
Who? groups (e.g., with respect to scale, single members of a group having particular roles, hierachies etc.)	*Who?* groups (e.g., as plural subjects sharing an intention, or the personal intentions of agents are interlocked in a way that a shared intention emerges)
What? behavioral coordination towards a common goal	*How?* cognitive skills and competencies
Where? contextual features	*Why?* reasons, beliefs, desires, affections, etc.

Fig. 1.1 Leading questions in externalist vs. internalist approaches to cooperation

organizational structure, without appealing to mental factors such as shared intentions at all (see Chap. 6). An internalist approach to cooperation, in contrast, may argue that minimal cooperation coincides with minimal joint intentional action (see Chap. 2). Moreover, we can distinguish between different kinds of minimalism, some of which are neutral with respect to the internalist-externalist distinction such as conceptual or ontological minimalism. Others kinds of minimalism, in contrast, are not neutral. For example, contextual minimalism and scale minimalism account for minimality in cooperation in an externalist sense, whereas cognitive minimalism draws upon an internalist view. An approach to cognitive minimalism (e.g., Chap. 7) may postulate a variety of sets of minimal (social) cognitive skills and capacities that enable different kinds of minimal cooperation. An approach to contextual minimalism, in turn, may account for minimal cooperation by referring to the absence of contextual features that typically favour its emergence (e.g., Chap. 3).

Most of the contributions in this volume focus on collective actions in minimal cooperation in one way or the other. However, cooperation can also be minimal in virtue of mental agency being genuinely collective, i.e., non-reducible to the mental agency of the single individuals involved in the cooperating activity (e.g., in collective remembering; Chap. 12). Some philosophical questions, such as whether and to what extent minimal cooperation can be understood as a basic collective action in an instrumental or causal sense, have not been addressed in this volume at all. Other questions remain open for empirical future research; for example, whether motor representations of a collective goal cause and shape the experiences of the cooperating agents in the same way as they shape and cause the experiences of a single agent who pursues a personal goal (see Chap. 11).

Acknowledgments This research has been supported by a grant on Minimal Cooperation, led by Corrado Sinigaglia and Hykel Hosni at the Philosophy Department of the University of Milan, Italy, which has been finished for Anika Fiebich in December 2019. Thanks to the members of the Centre for the Study of Social Action at the University of Milan and the Centre for Agency, Values and Ethics at Macquarie University for fruitful scientific discussions.

Chapter 2
What Is Minimally Cooperative Behavior?

Kirk Ludwig

Abstract Cooperation defines the social world. What are its limits? What draws the line between cooperation (even if not full cooperation) and mere coordination? What is its relation to joint intentional action? This chapter argues that minimal cooperation—in the sense of cooperation at play in our understanding of the human social world—coincides with minimal joint intentional action, where this is characterized in terms of members of a group having intentions directed at their bringing something about together in accordance with a common plan at the time of action—a shared intention. The strategy is to start with an account of full cooperation and then to consider what features can be discarded while some form of cooperation is retained and then to argue that the account of minimal conditions for shared intentional action captures the sense in which agents are still cooperating. The chapter considers cooperative activities in which there is a constitutive conflict of goals (e.g., competitive games), cooperation on regulating what are constitutively non-cooperative activities (the laws of war), forms of cooperation that make possible forms of conflict that are not embedded in them (quarreling), halfhearted or grudging effort (a work slowdown), failure to provide aid where needed (Bratman's Ayn Rand singers), or interference with others (partial sabotage of others' contributions). Finally, the chapter takes up the question whether a decision-theoretic notion of cooperation captures the ordinary social concept without appeal to shared intention (no), whether coercion is compatible with it (yes), and whether it requires rationality (no).

Keywords Cooperation · Shared intention · Joint action · We-intention · Competition · Conflict · Game-theoretic cooperation · Coercion

K. Ludwig (✉)
Department of Philosophy, Indiana University Bloomington, Bloomington, IN, USA
e-mail: ludwig@indiana.edu

A. Fiebich (ed.), *Minimal Cooperation and Shared Agency*, Studies in the Philosophy of Sociality 11, https://doi.org/10.1007/978-3-030-29783-1_2

2.1 Introduction

Cooperation admits of degrees. When factory workers stage a slowdown, they do not cease to cooperate with management in the production of goods altogether, but they are not fully cooperative either. Full cooperation implies that participants in a joint action are committed to rendering appropriate contributions as needed toward their joint end so as to bring it about as efficiently as they can, consistently with the type of action and the generally agreed upon constraints within which they work. Their contributions are sensitive to information (where available) about how others are contributing, in the sense that they adjust as needed their contributions in light of information about how other contributors are faring, to ensure effective pursuit of their joint end. This includes rendering aid to other participants if needed, insofar as they are able.

Full cooperation entails that those cooperating are engaged in a joint intentional action. Some prominent studies of joint intentional action focus exclusively on cases of full cooperation; others seem to presuppose it as a condition on joint intentional action.[1] But not all joint intentional action is fully cooperative. One example

[1] Michael Bratman has focused attention a subset of collective or joint intentional actions that involve full cooperation (Bratman 2014, 1993, 1992). Margaret Gilbert's account of shared intention as a joint commitment to intend to *J,* which is understood in turn as a matter of the parties being committed to acting *as far possible* as if they were a single body intending to *J*, seems to presuppose they are committed to full cooperation (Gilbert 2006, p. 137). Tuomela's account likewise seems to focus on full cooperation (Tuomela 2013). In an early paper on cooperation, Tuomela clearly is thinking of full cooperation when he aligns cooperative behavior with acting on a group intention (Tuomela 1993), and he would classify many of the cases I discuss (e.g., "minimizing contributions" as in the work slow down, and conflict of goals in competitive games, and in quarreling) as non-cooperative actions. However, he distinguishes between cooperative and non-cooperative joint action types and engaging in them cooperatively or not. By a *cooperative joint action type* he means one in which the participants have the possibility of helping others in their roles (p. 98). So competitive games which involve a conflict of interest amongst the players (or teams of players) are ruled out, as well as action types in which the participants simply cannot respond to the others, e. g., a team saboteurs who work independently for operational security (but why is that *per se* non-cooperative?). By *engaging non-cooperatively* he means failing to give full effort, or hindering others, or free-riding, or failing to give aid to others (less helpfully: not having a cooperative attitude). So one can engage cooperatively in a non-cooperative joint action type like playing chess. It will be clear in subsequent discussion that I divide the landscape differently, treating these as different ways of moving away from full cooperation. In more recent work (Tuomela 2010, 2013), building on (Tuomela 2000, 2007, Chap. 7), Tuomela distinguishes between pro-group I-mode cooperation and we-mode cooperation. The latter corresponds roughly to his target notion in the earlier paper, where the members of the group construct themselves as a social group and think of themselves in terms of a non-aggregative 'we' (2010, p. 65). There is not space to go into a comparison here, but a few remarks are in order. First, what Tuomela defines as pro-group I-mode cooperation is not cooperation on my account since it does not involve a shared intention. The case of mixing a hollandaise sauce I discuss at the end of Sect. 2.10 is an example. Taking advantage of what one expects another to do, and even being mutually responsive, doesn't constitute cooperation. Of course, there could be no objection to using "pro-group I-mode cooperation" as a technical term in describing the relevant sort of action, except that it is misleading. It is not what we have in mind by cooperation in ordinary contexts. See Sect. 2.9 in this connection.

is the work slowdown. Another example is provided by competitive games like chess and football, or sports like boxing and wrestling, where participants are clearly not intending to contribute to the pursuit of all of the goals of the others engaged in the activity, even when those goals are internal to the type of activity in question, but instead intend actively to frustrate some of them. Most accounts of joint intentional action have difficulty with these cases. At the same time, they are cases in which there is some cooperation. In this paper, I provide a taxonomy of forms of non-cooperative behavior within the context of behavior that is still to some degree cooperative. I argue that the minimal conditions of joint intentional action define minimal cooperative behavior, that is, that minimally joint intentional action is per se minimally cooperative behavior.[2] This is to agree with Searle that "The notion of a we-intention, of collective intentionality, implies the notion of *cooperation*" (Searle 1990), but also that the notion of cooperation implies the notion of a we-intention (see also Tuomela 1993, 2010, 2013). This corresponds to the ordinary notion of working together toward the same end. I define in precise terms what that comes to, by appeal to the shared plan account of joint intention, and how it is possible in cases in which it seems that one or more participants are in one or more ways acting so as to frustrate the contributions of other participants to their joint action.

It should be acknowledged that 'cooperation' is used in many different contexts. For example, in biology it is extended to interactions among insects and bacteria, where neither individual nor shared intention is present, though we can find analogues of agents, their actions, and utilities, and analyze combinations of "actions" in game-theoretic terms.[3] Investigations into the evolution of cooperation take as their target systems of "actors" which can be described as "selecting" Pareto

Second, we-mode cooperation is, on Tuomela's view, what it is to engage in joint intentional action. His account is nonreductive, using a preanalytic notion of joint intentional action in specifying what it is for a group to share an intention in the we-mode. Thus, his account of cooperation is ultimately circular (though he argues it is not viciously circular). In contrast, I offer a non-circular reductive account in Sect. 2.3. If correct, it provides deeper illumination of the concept of cooperation. Third, the we-mode requires conditions, like rationality and mutual belief in conditions for success being met, which are not present in all cases of joint intentional action or cooperation—see note 6 and Sect. 2.11. For further discussion of Tuomela's account of we-intentions and shared intentions, see (Ludwig 2015a, 2016, sec. 16.1).

[2] Paternotte (Paternotte 2014) offers an analysis of "minimal cooperation," but it is not clear that his is the same project as mine. I seek a core notion that underlies all forms of social cooperation. Paternotte focuses on certain cases of cooperation that don't fit analyses of joint intentional action that call for fairly strong conditions on common knowledge or belief, e.g., a demonstration in which those who come out are uncertain whether others will. He calls this minimal cooperation. He crafts a set of conditions designed for these cases but argues that minimal cooperation must be rational (pp. 60–61). Yet he allows that cooperation in general (in classical cases) does not have to be rational (60). Thus, the conditions he defines are not common to all cases of cooperation. They do not characterize minimal cooperation in the sense I have in mind. I will return to the question whether minimal cooperation requires rationality in Sect. 2.11.

[3] See (Sterelny et al. 2013) for a sense of the range of systems which have been treated under the label of 'cooperation'.

optimal[4] sets of "action strategies." The uses in biology (where we are not dealing specifically with agents) are examples, I think, of linguistic exaptation. The target notion for the purposes of this paper is what I will call social cooperation, the sort that is characteristic of the human social world, where the cooperators are agents acting intentionally. This does not limit social cooperation to the human world, but it does limit social cooperation to genuine agents, and, if I am right, relatively sophisticated agents.[5] I will return in Sect. 2.9 to the question whether a game-theoretic account in terms of strategic reasoners is adequate to characterize this notion.

In Sect. 2.2, I give a fuller characterization of the conditions for fully cooperative behavior. The question then is how far one can deviate from fully cooperative behavior and still have something that deserves to be called cooperative. When we have reached the limits, we have minimally cooperative behavior. In Sect. 2.3, I outline the account that I argue captures both the minimal conditions for shared intention, and so joint intentional action, and the minimal conditions for cooperation. In Sect. 2.4, I discuss cooperative behavior that embeds constitutively a conflict of goals, as in the case of competitive games, and show that the account of shared intention accommodates it. In Sect. 2.5, I discuss cooperation in moderating conflict that is not in its nature embedded in cooperative behavior, such as that involved in the laws of war. In Sect. 2.6, I discuss forms of conflict that rely on certain forms of cooperation for them to be possible, where the forms of cooperation are not built into the conflict. An example is a couple fighting over how to respond to finding out that their son is taking drugs. In Sect. 2.7, I consider cases in which not all participants in joint intentional action cooperate fully on reaching their goal

[4] A set of strategies is Pareto optimal if there is no other set of strategies that makes every player at least as well off and some player better off.

[5] Fiebich (2017) argues that the conditions required by the shared intention account of cooperation may be too cognitively demanding to allow for children younger than five to be counted as cooperating in that sense, even though studies find that children as young as two engage "in cooperative problem solving activities and social games involving parallel or complementary roles" (sec. 3.1). It is not clear to me that my own account of shared intention requires more than can be comfortably attributed to children who appear to engage in joint activities with others. It does not include, e.g., any common knowledge or mutual belief requirement. It does not appeal to sophisticated sui generis concepts of joint intentional action. It does require the concept of intentional action and so of intention, but ordinary attributions of intention to small children seem to presuppose this since they presuppose children have the concept of agency by embedding action verbs in the complement of attitude attributions. Plans can differ in their complexity and sophistication, as Fiebich notes, but the account I offer encompasses very simple as well as very complex plans. However, suppose that very young children do not have even a rudimentary concept of intention (as a commitment to action as distinct from a desire which carries with it no such commitment), but engage in activities which we are inclined to describe from the external perspective as cooperative. Without settling exactly how to describe its cognitive aspects, we can allow that it involves a kind of proto-social cooperative behavior. At some developmental stage, it is plausible that children engage in something that is on its way to but not yet comparable to the forms of cooperative behavior that adults routinely engage in. I do not want to deny this. My interest lies, however, in what it is that this is a stage on the way to becoming. For to understand the developmental story about cooperation, we need to understand what its target is.

in the sense that they either refrain from offering aid when possible, or actively interfere with the efforts of others. In Sect. 2.8, I consider grudging or halfhearted effort, as in the case of the work slowdown in which a group intentionally delivers jointly a subpar performance in some joint task. In Sect. 2.9, I consider whether coordinated behavior in which individuals realize a mutual benefit, but which does not involve a shared intention, might still count as cooperative behavior (the game-theoretic conception). Section 2.10 responds to the objection that cooperation excludes coercion, while the account allows joint intentional action under threat. Section 2.11 responds to the objection that cooperation is constitutively rational. Section 2.12 is a brief summary.

2.2 What Is Fully Cooperative Behavior?

We are looking for minimally cooperative behavior, which is one end of a spectrum at the other end of which is fully cooperative behavior. We begin therefore with fully cooperative behavior and then consider how far we can retreat from it and still have something that looks like cooperation. In explaining fully cooperative behavior, I list central features of it. The claim is that every feature on the list is necessary for behavior to be fully cooperative and that they are jointly sufficient.

(1) **Shared Intention**. Fully cooperative behavior is jointly intentional behavior and so entails joint intentional pursuit of a common goal that is minimally expressed in what the shared intention of those who cooperate aims to bring about.
(2) **Absence of Conflict**. Fully cooperative behavior excludes conflict of purpose that is embedded in, or is presupposed or supported by, the joint action that the cooperating group is engaged in.
(3) **Willingness to Aid**. Fully cooperative behavior requires

    (a) [*Non-interference*] not only that participants not actively interfere with the contributions of others, except to the extent it is necessary in part to fulfill a participant's own role and is not at the expense of success in the joint enterprise, but also
    (b) [*Mutual Aid*] that participants in it be ready and willing to provide aid to others[6] in performance of their parts in their joint action, insofar as it is compatible with their performing their own parts and the aid offered does not otherwise detract from their joint pursuit of their goal (balanced appropriately against other considerations such as personal cost and the requirements of morality); and

[6] This can include as Tuomela notes (1993, pp. 93–4) helping secure preconditions for another's performance of his task, helping the performance itself, and shielding him from various things that might interfere with his performance of his task.

(c) [*Attention to Others*] that participants pay attention, with an eye toward meeting desiderata #3a-b, to what other members of the group with whom they are acting intentionally are doing and their progress in successfully carrying out their parts in it, insofar as this is possible compatibly with carrying out their own parts.

(4) **Full Effort**. Fully cooperative behavior requires participants to intentionally give full effort in doing their parts, relative to the type of task, that is, that they aim do their parts as best as they can, rather than, e.g., half-heartedly, or in a way that they recognize to be inefficient or suboptimal relative to what they understand themselves to be able to do, given the nature of the task (balanced appropriately against other considerations such as personal cost and the requirements of morality).

(1) is a minimal condition and it will occupy us below. (2) rules out competitive games and forms of cooperation that provide merely the preconditions for forms of conflict or that moderate or regulate aspects of what would otherwise be unconstrained conflict. However, (2) should not be construed so as to rule out people cooperating for different reasons. For example, two people may cooperate to elect someone for different reasons. One for his principles and another for money. They may cooperate fully regardless of whether their purposes in working together are different. (2) should also not be construed as ruling out full cooperation in cases in which participants have longer term goals which they know will bring them into conflict, for they may for all that fully cooperate on an interim task, each thinking "sufficient unto the day is evil thereof," that is, postponing until it is necessary the inevitable conflict they see between them. (3) requires the participants both (a) to refrain from interfering with others and (b) to be ready to aid others in performance of their parts so far as that is compatible with their doing their own part and does not otherwise make it less likely that they will achieve their goal, and in consequence (c) to pay attention so far as possible to what the others are doing. (3a–c) provide a generalization of Bratman's requirement of mutual responsiveness in what he calls modest sociality in (Bratman 2014) and earlier called shared cooperative activity (Bratman 1992). It does not require actual responsiveness in behavior or planning, but only where there is scope for it given the nature of the action undertaken. So, for example, it does not rule out joint action in which the members of the group cannot interact, as in prepackaged behavior, e.g., our synchronizing our watches and performing appropriate actions in sequence at widely separated locations. (4) requires wholehearted effort. (4) rules out a work slow-down from being a fully cooperative activity. Dragging one's feet, making a halfhearted effort, being lazy, or having to be urged on at every step of the way is not fully cooperative behavior.

2.3 What Is a Shared Intention?

Shared intentions are attributed using plural noun phrases, as in [1],

[1] We intend to build a boat

where the object of the intention is understood to involve the group picked out by the subject term doing something together, that is, jointly, rather than individually, so that when the intention is successful the group does that thing together intentionally. The two basic options are that

(i) a shared intention is an intention of a group per se, that is, a group-level psychological state, or
(ii) a shared intention is a matter of each member of the group having an appropriate sort of intention directed at their doing something together.

On the first view, the subject of the intention attributed in [1] is the group that 'We' refers to. On the second view, the subject term in [1] is understood distributively, so that [1] attributes to each member of the group a certain kind of intention directed at their building a boat together. I will call views in the latter category distributive accounts of shared intentions.

I have argued in favor of the multiple agents account of plural agency (Ludwig 2007, 2016). On this view, to say that we built a boat is to say that there is a boat building event (stripped of the implication that there are agents of it) of which each of us, and only those among us, were direct agents.[7] On this view, the distributive/collective ambiguity in plural action sentences traces to a scope ambiguity, namely, whether the event quantifier introduced by the action verb takes wide or narrow scope with respect to a distributive quantifier over members of the group. If the multiple agents account is correct, then there is no need for a group-level agent in the case of plural attributions of intention on their collective reading, since there is no group agent that is the object of the intention. Thus, I adopt a distributive account of shared intention. Following Tuomela and Miller (1988), I call intentions of the sort agents have when sharing an intention *we-intentions*. They are intentions of individuals, not of groups, and the term (as I use it) is merely meant to help pick out whatever sort of intention is distinctive of participating in collective intentional action. The key question for distributive accounts of shared intention is what makes we-intentions special.

What is special about we-intentions lies either in their mode or their content. Mode accounts include (Searle 1990; Gilbert 2013). The mode account is commit-

[7] Some authors use 'joint action' to mean what I mean by joint intentional action (Tuomela 1993). Terminologically I think is better to make the distinction with the modifier 'intentional' to avoid confusion.

ted to the view that verbs of intention are ambiguous in their use in attributing intentions directed at individual and at collective action (of the relevant sort).

However, it tells against this that verbs of intention are perforce interpreted in the same way across clauses in which the verb is elided in the second clause and there is a shift from collective to distributive readings, as in [2].

[2] They intend to build a boat together and sail it individually.

Since the elision is an abbreviation, it is proper only if the verb is used in the same sense in both clauses. If the mode account were correct, the result would be a zeugma, like 'He left the room and his hat' (Zwicky and Sadock 1975), but it is not. Thus, we should reject mode accounts of we-intentions. If we focus on the content, we can take it to involve sui generis concepts pertaining to joint intentional action, as in Tuomela's account (Tuomela 2005b, 2013), or to involve only concepts at play already in our understanding of individual agency. If the latter is possible, then it is to be preferred. I have argued for a reductive account of we-intentions, which I call the shared plan account, elsewhere (Ludwig 2007, 2016). I sketch the main idea of the account here and then argue in the rest of this chapter that it draws the line between minimally cooperative and non-cooperative activity in the right way. The taxonomy of accounts of we-intentions is given in Fig. 2.1.

The shared plan account is motivated by projecting an account of the content of individual intentions to the plural case. In the individual case, if I intend to build a boat, I intend to bring it about in accordance with a plan I have at the time of acting that I build a boat. If *we* intend to build a boat, then, in parallel with the individual intention, we intend to bring it about in accordance with a plan *we* have at the time of acting that we build a boat (together). Thus, we intend that we implement a collective action plan for building a boat. If I instantiate this to me, then we get [1].

[3] I intend that we bring it about in accordance with a (collective action) plan we have at the time of action that we build a boat together.

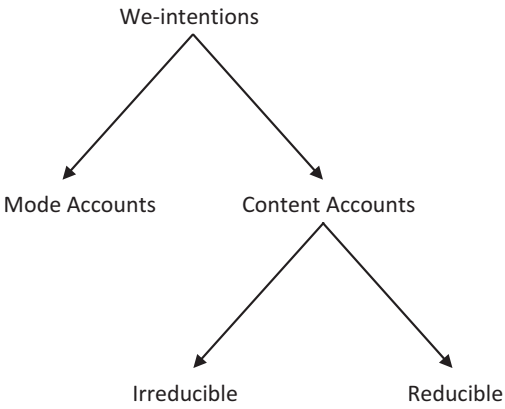

Fig. 2.1 Taxonomy of accounts of we-intentions

An individual intention of the form '*x* intends that ….' is short for '*x* intends to bring it about in accordance with a plan *x* has at the time of acting that …'. Thus, we get [4], abbreviated in [4a].

[4] I intend that I bring it about in accordance with a (individual action) plan I have at the time of acting that we bring it about in accordance with a (collective action) plan we have at the time of acting that we build a boat together.

[4a] I intend to bring it about that we build a boat together in accordance with a shared plan.

This secures that each individual's intention is directed toward an action of his own, but one which itself aims at joint action on the part of members of the group in accordance with a common plan.

Several quick remarks are in order. First, the account does not require that I think I have the power to ensure the joint action occurs by myself. I only have to think that I can make a contribution, given other conditions in place, to bringing it about that we act in accordance with a shared plan (Ludwig 2016, sec. 14.11). Second, it does not require that everyone know every aspect of the plan in accordance with which they act. It is enough that there is a plan that they can all think of which is the same and that each knows at least his own part in it (Ludwig 2016, sec. 14.13).[8] Third, in specifying the plan we allow, as in the case of individual action, a margin of error around the canonical plan conception, that is, certain deviations from the canonical plan are close enough to count as satisfying the intention (Ludwig 2016, sec. 7.4, 14.13). Fourth, this analysis does not include any common knowledge or belief requirement on shared intention or shared intentional action (Ludwig 2016, sec. 14.15).[9] Fifth, the account is reductive and so non-circular. If it is adequate as an

[8] Paternotte (2012) has objected that shared plans are not necessary for cooperation or joint intentional action. The examples he has in mind are demonstrations and flash mobs. But the participants do share a plan in these cases, in the sense I require. The plan for the demonstration is defined by where it is to take place and what it is a protest about, and it leaves room for improvisation in the light of onsite developments. The plan for the flash mob is likewise defined by the place and the action specified. To intend to act in accordance to the same plan with others does not require everyone to know who all the other members of the group are. They have to have a way of locating the group they are acting with, but they can do this by description rather than enumeration. For some action types like a demonstration, everyone has roughly the same role, so one thinks of the group as consisting of those filling a demonstrator role in the location of the demonstration. In the case of a flash mob, which is arranged by a mastermind, different people or groups of people might be given different roles and not know the roles of others. But that is fine too. They don't have to know the roles of others, or who they are, to carry out an intention to act in accordance with the same plan. They have to know their own part, and to know that there is a reasonable chance that in doing their own part they contribute to the group acting according to the same plan. Knowing someone is organizing the flash mob and assigning roles is enough both for this and for being able to think about the group with which one acts.

[9] Here is an example from Paternotte (2012, p. 49): a hundred people want to demonstrate against, say, racial inequality, but they don't know beforehand whether others will participate; nonetheless, they all go out to demonstrate and do so together intentionally, and this is a form of cooperation toward a shared goal. Something similar can be said about flash mobs (ibid.), or prisoners who

account of minimal cooperation, it avoids the charge of defining cooperation in part by appealing to the idea in the definiens. This contrasts with Tuomela's account of cooperation in terms of shared intention, which appeals to an unanalyzed notion of joint intentional action in the definiens (Tuomela 2005a, 2013).

Thus, my thesis, more precisely, is that minimal cooperation consists in a group executing successfully a shared intention in the sense just given.

2.4 Competitive Games

By design competitive games, like chess, boxing, bridge, or basketball, are not fully cooperative joint activities. At the same time they are essentially collective intentional action types that require cooperation. Thus, these activities embed conflict constitutively in a form of cooperation. How is this compatible with the participants intending to act in accordance with a shared plan?[10]

Competitive games are governed by constitutive rules. The participants must jointly follow them, at least by and large, to count as engaged in the activity type (Ludwig 2014, 2017; Searle 1969). That is why they are essentially intentional collective action types. The rules specify a type of pattern of collective action. For example, the rules of chess specify constraints on sequences of actions of the players involved. Any allowable sequence fits the pattern. To play, the participants must jointly intentionally aim at instantiating the pattern, by jointly intentionally following the rules. In this respect they cooperate. At the same time the action type includes as a constitutive feature (or a regulative ideal) that each participant aim at a final state (winning) that only one of them can achieve. Competitive games therefore fail to meet the No Conflict condition (2) on fully cooperative activities. However, it is constitutive of the activity type itself that participants aim to engage in that they each strive to win. Given this, they are in fact cooperating on competing with each other in a certain form of activity. The competition is embedded in what it is that they are cooperating on. Therefore, if either of them were to fail to strive to win, he would not be fully cooperating in the activity, but instead undermining it. Full cooperation in competitive games requires knowingly aiming at a goal that will frustrate what one knows to be a goal of the other in the same activity. Competitive games may then seem an odd case of failing to fully cooperate with others, since if one

intentionally cooperate (in the full sense) in the Prisoner's Dilemma (see Sect. 2.9) without knowledge that the other is doing so (this rules out even the weakest forms of common knowledge). See (Blomberg 2015) for additional arguments.

[10] See (Nguyen 2017) for a sense of the difficulties that beset theorists in coming to grips with competition in cooperation in the philosophy of sport. Nguyen looks for some further thing that the competition promotes for the players so that winning is not really the goal each strives to achieve. I think this is unnecessary. The cooperation consists in the shared intention to bring into existence an activity type governed by constitutive rules that include that each strives to win the game. The reasons we decide to realize such action types are a separate matter. Sometimes it is recreation, sometimes it is prize money, sometimes it is to humiliate one's opponent, and so on.

were to aid competitors in all their goals in pursuing the activity, one would thereby be failing to engage fully with them.

How can full cooperation in an activity type frustrate full cooperation in it? There is no real paradox here. The sense in which competitive games fall short of full cooperation is not that each participant fails to be fully committed to what they intend to do together, but that they have goals constitutive of doing it that cannot in principle be simultaneously realized.

Competitions may appear to raise a problem for the Willingness to Aid condition (3) on full cooperation. (3b), [Mutual Aid], requires that participants be ready help the others so far as compatible with doing their own part. But in competitive games it may appear that this requires them to help the other win, since that is part of what it is for the other to do his part. Yet, this is incompatible with doing one's own part. However, (3b) includes the proviso that one aid another only insofar as it does not prevent one from fulfilling one's own part or otherwise undermine the joint activity. Helping one's competitor to win undermines one's own contribution to the activity type, and undermines the goal of a competition directly. Consequently, although at first sight competitions appear to be incompatible with (3b), they are consistent with it. (3a), [Non-interference], may seem to present a separate problem. For it is undeniable that in competitive games like chess, one is aiming to frustrate the opponent's goal of winning. However, since one's own role requires that one try to prevent one's opponent(s) from winning, to fail in this would be to fail to play one's own role in the plan. Moreover, since the other aims to win despite one's own efforts to prevent him (for that is the point of the game), to fail to try to prevent him would after all be interfering with a condition on his playing his appropriate role. So superficially what looks like a violation of (3a) is not. One is aiming to frustrate the other player's goal of winning but not his contribution to the joint activity, playing chess, as such.

Bratman's account of shared cooperative activity and of modest sociality has a prima facie difficulty with competitive games. In his early account he required that, for two people to be engaging in shared cooperative activity, they each intend (i) that they *J* (for cooperative neutral action type *J*), and (ii) that they *J* by way of each of their intentions that they do so, and (iii) that the subplans formed by each in pursuit of that end mesh with one another (Bratman 1992). Their subplans mesh provided that they are consistent with one another. This leaves open whether all or only some of their subplans are to mesh. In *Shared Agency* (2014) Bratman requires for modest sociality that the participants intend that *all* of their subplans mesh. But this appears to rule out any form of joint intentional action in which there is competition, for competition entails that some of the subplans of each participant associated with their intentions to engage in the activity are not consistent with some of the subplans of at least some other participants (in team competitions one's subplans are not necessarily in conflict with those of other team members).[11]

[11] I believe Bratman can modify his account to respond to this objection (Ludwig 2015b). For competitive activities defined by constitutive rules, one can require not that all subplans associated with each participant's intention to participate mesh, but only that all associated subplans that are

What about Gilbert's account? Since Gilbert holds that for us to intend to *J* is for us to be "committed together to constitute, as far as possible, a single body" that intends to *J* (Gilbert 2006, p. 137), for us to intend to play chess together would be for us to be committed to constitute, as far as possible, a single body that *intends* to play chess, and so, to constitute, as far as possible, in turn, a single body that *plays* chess. But what does this come to when playing chess involves competing against one another? Playing chess against myself isn't very much like playing chess with someone else. I can't really aim to frustrate my own goals. I can't hide my strategy. I can't plan to take my opponent by surprise with a discover check or a knight fork. Perhaps it might be said that we are only to do so *as far as possible*. But how far does that take us? Does it take us as far as actually playing chess? The trouble is that the idea of emulating a single body that has a goal (carrying a piano upstairs) doesn't fit very well a goal that embeds competition between distinct agents.

In contrast, the shared plan account handles competition in a particularly simple way. There is obviously a plan type which the competitors each intend to contribute to their jointly implementing (and so to making it the case that there is a plan they share in acting). The plan type is specified by the rules of the game together with a specification of who plays which roles (e.g., white or black in chess). That participants are each to compete against the other participants (sometimes as teams) is a constitutive element of the plan that they intend to jointly implement. There is no conflict between two people intending act in accordance with a common plan to play chess and each intending to win because, while it is part of the plan that they compete against one another, it is not part of the plan that they both win.[12]

2.5 Conflict Constrained by Cooperation

A more puzzling case is conflict constrained by cooperation, where the kind of conflict itself is not constitutively governed by rules, as are competitive games. A paradigm is modern warfare between nation states governed by the laws of war. In this case the rules are regulative rather than constitutive. The activity type, which constitutively involves a conflict of goals,[13] can exist independently of being governed by

focused on bringing into existence *the action type which is the goal of their shared intention* mesh (e.g., those directed at instantiating *playing chess*, as opposed to winning).

[12] Competitive games stand in for a range joint intentional actions in which there is some built in conflict of interest though they are in their nature joint intentional actions. For example, in bargaining or negotiating there is a conflict of interest as well a common interest that subsumes it. Each side wishes to get the best deal possible (perhaps within certain limits, if neither wants to take too great advantage of the other), but they engage in a cooperative interaction designed to arrive at a mutually acceptable transaction or agreement.

[13] It is possible to subsume combat under a cooperative scheme, as when champions are chosen to fight as part of a decision procedure. But while war subsumes combat, the two concepts are not the same. War may decide an issue between two sides but it is not a matter of their agreeing to abide by the outcome but rather of the winning side being in a position to dictate to the other without their agreement.

the rules. Consequently, there is no sense in which there is cooperation on being in conflict, as there is in competitive games. In the case of competitive games this is possible because there is an action type that the participants aim to instantiate that builds in competition. The problem this poses is threefold. First, how is cooperation in war possible at all? What does it come to? Second, if it can involve cooperation, is it an example of cooperation in the absence of joint intentional action? Third, if it involves cooperation, in what way, if it does, does it involve the intention of the parties to act in accordance with a shared plan? If there is cooperation in war that is not joint intentional action or that does not involve intending to act in accordance with a shared plan, then my account of minimal cooperative action is incorrect. I will argue that what cooperation is implied by the laws of war, when relativized to its proper domain, does not pertain to making war itself, and that once we have the proper domain of cooperation in view, it is intuitively joint intentional action and conforms to the shared plan account of joint intentions.

Constraints on the conduct of war have an ancient pedigree, but codes for the conduct of modern warfare trace their origins back to the Lieber Code governing the conduct of the Union armies in the US Civil War, signed by President Lincoln on 24 April 1863. These formed the basis for the Hague Conventions on land warfare of 1899 and 1907, which were superseded by the First through Fourth Geneva Conventions (1949) in the aftermath of the Second World War, and subsequent Protocols, I-II (1977) and III (2005). One hundred and sixty-nine states have ratified the four Geneva Conventions, including all the members of the United Nations. One hundred and seventy-four have ratified Protocol I. One hundred and sixty-eight have ratified Protocol II, while 73 have ratified Protocol III. The Geneva Conventions govern both the conditions under which a war is acceptable (jus ad bellum) and constraints on acceptable conduct in war. They are designed to mitigate the worst aspects and consequences of war.

The First and Second Geneva Conventions concern primarily the humane treatment of sick and wounded members of armed forces on land and at sea. The Third concerns the treatment of prisoners of war. The Fourth concerns humanitarian treatment of civilians in a war zone. Protocol I (1977) is an amendment to the Geneva Conventions that extends provisions of the original four in light of developments in modern warfare. For example, it outlaws attacks on aircrew parachuting from an aircraft in distress without giving them an opportunity to surrender. It prohibits indiscriminate attacks on civilians or civilian targets, and the use of biological weapons, nuclear weapons, and land mines, whose scope of destruction cannot be limited. Protocol II (1977) provides protections for victims of internal armed conflicts, and Protocol III (2005) provides for the use of the Red Crystal emblem for medical and religious personnel in time of war in addition to the Red Cross and Red Crescent.

Wars are not joint intentional actions. The parties to war are not executing a shared plan in making war on each other. War in itself is constitutively non-cooperative. Those waging a war of conquest are not trying to do something with those they aim to conquer. They may aim to force the subject population to cooperate once conquest has

been completed, but the act of war itself aims to destroy organized resistance and to frustrate the opponent's aims.[14] It is all the same to an invading army if their opponents are defeated in battle, simply lay down their arms, or are destroyed by a plague or act of nature. Their goals are not affected by this. They do not require resistance to satisfy their war aims. So it is not in the pursuit of war that the parties to war cooperate.[15]

The laws of war have the character of side agreements to conflict. To see the structure more clearly, start with a simpler case. Suppose that you and I are sworn enemies and we are each committed to killing the other. Suppose, however, that, recognizing this, and recognizing some common interests despite our conflicting goals, we agree not to use poison and not to harm each other's family. We do not share an intention in each of us intending to kill the other. We are not playing a game, and we do not have a joint goal: mutual destruction. But we have cooperated in one respect, in placing limits on what each of us will do. We thus cooperate on constraining the nature of the conflict between us. This comes to our cooperating on ruling out certain determinate forms of conflict. Thus, we share an intention that, given the condition that we are engaged in a life-and-death struggle, certain forms that struggle might take are excluded. The cooperation takes place against the background of the acknowledged conflict, which is not a form of cooperation. We are not cooperating in making that the background. But for the conflict, there would be no ground for any cooperation. But the cooperation is not on engaging in the conflict, or some common aim in pursuing it, but on moderating it in certain ways. In this case the moderation comes in the form of restraints on behavior. So what we aim to do jointly is negative: we aim to make it the case that neither of us poisons the other or harms the other's family. Let us call using poison or harming families in a conflict *inhumane*. So what we are cooperating on doing together is *not engaging in inhumane conflict*. Cooperating on not engaging in a certain form of conflict does not entail that whatever form of conflict we do engage in involves our cooperating with one another on engaging in it. This gives us the structure of cooperation in war on observing certain prohibitions, e.g., against targeting civilians or the use of chemical or biological weapons.

[14] Cannot two Kings say: Let's go to war! We can imagine circumstances for that, but they would be circumstances in which this is simply an agreement that they cannot settle their differences by negotiation, so that cooperation on a resolution is at an end. In this connection, we might reflect on Clausewitz's famous aphorism that "War is a mere continuation of policy by other means." Indeed, when you cannot get what you want by negotiation (a form of cooperation on conflict resolution), then if you do not give up your end, you shift to other non-cooperative means. If you want a farmer's cow, and he asks more than you can pay for it, robbery is a continuation of your policy of possessing the cow by other means.

[15] The concept of war does not include in it anything that would make pursuing it successfully contingent on an opponent mustering effective opposition. But does it exclude the idea of cooperation? Could we share an intention to make war on one another? Would it be war if one party, recognizing that the other was not able to effectively make war, provided it with the means of doing so? Even making provision for vague boundaries and family resemblance, it seems to me that this would not be war as we understand it.

Let's enlarge slightly the scope of our cooperation, again with an eye to further applications to the laws of war. Suppose in addition to the two prohibitions we have already agreed to, we also agree that should any members of the other's family (who are not involved in the conflict—see below) come into our power, they will be treated humanely, not only not harmed, but provided with adequate shelter and food; they will not to be used as hostages in the conflict; and they will be provided with the ability to communicate in order to make arrangements, under a truce, to be returned. Here what we have agreed upon is not merely not to engage in certain sorts of activities but, on certain contingencies, to take certain positive actions, some of which involve active cooperation with our opponent. What we have agreed is to cooperate positively in certain tasks which are not constitutive of our conflict, but which may arise ancillary to it. Thus, again, we do not cooperate in our conflict. But we have a shared conditional intention to cooperate in certain circumstances, on distinct tasks, that we anticipate may arise in the course of it. This corresponds to provisions for the humane treatment of non-combatants in the laws of war.

Finally let us suppose that in our conflict we anticipate enlisting some members of our families, who then participate in the conflict as our agents, and whom we call combatants. Suppose that we come to an agreement that if any combatants of the other's family are in a position in which they can no longer effectively fight and are faced with being killed, they may signal (by saying "yield" or waving a white flag) that they lay down their arms and consent to pass into the control of the other, i.e., they surrender, and that once they have surrendered, they may be held so that they cannot reenter the conflict, but that in being so held they must be afforded necessary medical care, food, and shelter, and not be harmed, and that they can be released if they agree to no longer participate in the conflict (give their parole). This represents a further stage of cooperation because it is not merely cooperation on what not to do, or cooperation on what to do in contingences that are ancillary to, but anticipated to be a consequence of, the conflict, but cooperation on how particular episodes in the larger conflict may be brought to a close, by a formal surrender, and then how the status of those who surrender changes so that they cease to be treated as active combatants, i.e., as involved in the conflict, and pass to the status of prisoners, conceived in terms of the provisions of the agreement, and then at a further stage, on an agreement, to the status of being paroled. The agreement on arrangements for surrendering concerns the nature of the conflict itself. It is an arrangement on how to bring a certain portion of it to an end. The shared intention is a universally quantified conditional intention. It has the following form: we intend that whenever anyone involved in the conflict signals a surrender to the other side, conflict ceases (between those surrendering and those accepting it) and those who surrender pass into the control of those to whom they surrender, the victors, and cease to be treated by the victors as active combatants. This conditional shared intention does not extend to a shared intention to engage in the conflict, and does not entail that there is any shared goal those involved are pursuing. The further agreements about the treatment of prisoners and their release on parole concerning the aftermath of surrender likewise entail no shared intention to fight, though they presuppose, like the arrangements for

surrender, that the two sides are engaged in a conflict, each intentionally seeking to prevail over the other. These simplified arrangements between us are analogous to the provisions for surrender and treatment of prisoners of war in the laws of war.

To summarize, the laws of war make provisions for three sorts of cooperation:

(1) Cooperation on exclusion of certain forms of conflict, e.g., what weapons are to be used, or certain ways of pursuing the conflict, e.g., destruction of food crops, sewing fields with salt;
(2) Cooperation on how to handle situations that arise as a concomitant to the conflict but which are not constitutive of pursuing it, e.g., treatment of those displaced by the conflict falling into one's area of control or subject populations;
(3) Cooperation on regulation of the conflict itself, e.g., on a mechanism for surrender and treatment, and parole or exchange of those who surrender, or a mechanism for a truce, e.g., to recover wounded or for negotiations.

In the first case the parties share a negative intention and in so doing cooperate on ensuring the conflict does not take certain forms and so share a plan to that effect. Their refraining jointly from such activities is something that they do jointly intentionally. This does not entail that they share an intention with respect to the conflict itself. In the second case the parties share conditional intentions to cooperate on providing shelter and sustenance to those displaced by the conflict and in their power. Though typically their spheres of actions are distinct, yet the goal they aim at is general, that those displaced or falling into the control of either party be cared for. Here what they cooperate on is mitigating the effects on those not directly party to the conflict who are affected by it. This does not entail that they share an intention with respect to the conflict itself. In the third case they share an intention with respect to mechanisms for suspension or conclusion of conflict, that is, they intend to share a plan with respect to how conflict may be suspended or concluded. This is not a commitment to execute the mechanism on certain conditions, but rather a commitment to respond in a certain way to an overture by the other. This is similar to resigning in chess before checkmate, except that the agreements on the mechanism are not constitutive of the type of activity. Since the cooperation is limited in scope, it does not extend to making war itself.

Thus, the forms of cooperation that are involved in the laws of war are compatible with war itself not being in any sense a cooperative activity. The laws of war define cooperation on matters ancillary to what defines war, but which are not separable from that context.[16] Where there is cooperation, there is also a shared intention, and specifically a shared intention to bring something about in accordance with a shared plan. The laws of war can be contrasted with Robert's Rules of Order (RRO). RRO govern meetings at which people aim to make decisions. RRO are regulative rules,

[16]This is why cooperation between the warring parties in war is not a matter of fully cooperating and leaving inevitable conflict for a future date. The war is on. Cooperation moderates it, but the tasks involved are not tasks that would exist but for the background of conflict.

like the laws of war. But they govern what is by its nature antecedently a cooperative activity. In contrast, the laws of war govern what is by its nature a non-cooperative activity.[17]

2.6 Conflict Supported by Cooperation

Some forms of conflict presuppose forms of cooperation. This is true for competitive games where competing is built into the concept of something that the participants in fact cooperate on. But in this section I have in mind cooperation supporting conflict where the fact that the parties are in conflict is not part of what they are cooperating on. An example is a couple quarreling over something, e.g., how to discipline a child, or how to spend money. Here the conflict is verbal, and so it requires communication, but communication requires cooperation. Each party to the dispute wants to communicate with the other and so uses expressions for which they share meaning conventions, and so uses them in accordance with those shared conventions. They cooperate on maintaining a position in which they can talk to or yell at the other. But in quarreling over something they are not (or need not be) cooperating. They share an intention with respect to the use of a communicative system that allows them to talk, but they do not share an intention with respect to the resolution of what they are fighting about.

Not all arguments are fights in this sense. Two philosophers or scientists or journalists, etc., may be arguing about the merits of some analysis, each giving arguments to the other. But they share an aim in this, which is to arrive at the correct view of it, and they are cooperating on pursuing that aim. What constitutes the cooperation in this case is their both having the goal of accepting good reasons for conclusions and sifting through what are good or bad reasons by offering and considering reasons the other offers, probing for weaknesses, and so on. Thus, each is aiming to accept claims for which the other provides good reasons, and to give up claims which the other shows not to be well supported or to be false. They cooperate

[17] We can define a meeting held in accordance with RRO as a parliamentary meeting. Relative to that action type, RRO are constitutive rules. When we think of RRO as constitutive of parliamentary meetings, we must think of the rules as including what is only a precondition on their being used if they are conceived of as regulative rules, namely, that one hold a meeting—where it is conducted in such and such a manner. Then we bring into existence this type of activity constitutively in following the rules together intentionally. Can the laws of war be turned into constitutive rules by a similar maneuver? We *can* define a war governed by the laws of war as *regulated war*. What we cannot do is add to the rules that we engage in a war, because war excludes sharing an intention to engage in it. So though we have the concept of an activity that is governed by the laws of war, so that it is not that activity unless those laws are followed intentionally for the most part by the parties to it, we cannot add what would be required to provide a set of rules the intentional following of which would bring into existence that type of activity because its nature precludes sharing a plan to instantiate it.

on a process of giving and responding to reasons and objections that has this as its aim. (There is plenty of scope here for falling short of the ideal.)

Another case of conflict supported by cooperation is a jury trial in which the prosecution and defense have opposing aims. The defense attorney aims to get his client off the charges. The prosecutor aims to secure a conviction. The jury constitutes the audience to whom arguments are addressed. The defense attorney and prosecutor each aim to convince the jury of claims that are not consistent with the other's. But this particular form of conflict requires an institutional setting to take place. The participants occupy status roles in an institution (Ludwig 2017, chs. 10–11). Their roles are governed by constitutive rules for an essentially intentional collective action type, the conduct of a jury trial, with roles for others, the judge, the members of the jury, witnesses, the bailiff, court recorder, and so on. Thus, they are cooperating at the same time as they have opposing aims. To what extent is this like the case of cooperation on a conversation designed to arrive at the correct conclusion? The overall purpose of the system that sets up a conflict between defense and prosecution is to arrive at a correct legal judgment. However, while this is a feature of the design of the institution, this does not erase the conflict in the way it does when two people are discussing what the right view is. For the mechanism is to set two sides against each without it being the goal, at least of the defense, that it arrive at the correct legal judgment, on the assumption that this is the best method for insuring, inter alia, that the state proves its case adequately. For the defense attorney may know that his client is guilty because his client has told him, yet try to get him off because that is the role he plays in the administration of justice. To what extent is this similar to competitive games which build into their concept a conflict? One difference is that conflict is not essential in a trial. The defendant may be advised to plead guilty by his attorney, or advise his attorney that he wishes to do so. So there is no essential cooperation on conflict between prosecution and defense in the conduct of a trial. It is rather that the trial, which requires cooperation, is the setting for a certain type of conflict which it may, or may not, occasion.

Physical fights, outside the context of competitions, can also be supported by cooperation. If each of two people are intent on beating up the other in a fight, and they know this, then they each have an interest in getting close enough to the other to do it. The fight is a non-cooperative activity. But they may cooperate on fighting because it serves their ends.[18] This can be expressed in their finding an opportunity to fight, even to the extent of arranging by an exchange of notes or a phone conversation when and where to meet for a fight. The code duello formalizes cooperation for a fight, which typically would end with one of the parties being unable to continue due to injury or death. Some duels were explicitly to death. The popular movie motif of Western films, the quick draw duel, rare in actual fact, is another example.

[18] Why then may countries not cooperate on making war on each other? There may be similar arrangements in the conduct of war, two armies agreeing to meet on a certain field of battle for example, each confident of victory. But simply making war requires no such arrangements, as it simply involves general commitment to the subjugation of the other side. No cooperation is needed to facilitate making war as such.

In these cases we admit there is some cooperation but it is not full cooperation because the cooperation supports conflict. To what extent is it joint intentional action, and in what sense do the participants intend to act in accordance with a common plan? In these cases the form of the conflict depends upon cooperation. In this respect it is like competitive games. But the conflict is not constitutive of what they cooperate on in the sense that, e.g., carrying on a conversation (the couple fighting) or inquiry into facts (the jury trial) does not require conflict. It is rather that the kind of conflict is made possible by forms of cooperation that set the stage for it. What the parties cooperate on, though, is not what they are in conflict about. In the case of a couple arguing, irreconcilably, about spending money, they cooperate on using a common language, on carrying on a conversation, but not on where it is to take them. They share and intend to share a plan about having a conversation and speaking the same language. They do not share or intend to share a plan about where it takes them. So where there is cooperation, there is shared intentional action, and they intend to act in accordance with a common plan. But the shared intention does not extend to a common end with respect to the activity it supports.

2.7 Failure of Full Cooperation on an Intended Goal

The ideal of full cooperation includes not just that one do one's own part, but that (i) one not interfere with others doing theirs (unless in the light of provision (ii)) and that (ii) one render aid to others (which can include correcting mistakes) where one can compatibly with doing one's own part, while respecting (i). We can still cooperate, though, when one or the other of the conditions is not met. We can call failures of the sort described in (ii) examples of non-responsiveness, and failures of the sort expressed in (i) examples of partial sabotage. Failures to give aid as needed (ii) are a less radical departure from full cooperation, so we deal with these first.

(a) Non-responsiveness. Non-responsiveness is not just the failure to help. In some cases agents cooperating with each other are just not in a position to help others. For example, if we are members of a team of deep cover spies infiltrating another country's institutions, for operational security we may not be given information about who the other members of the team are. Thus, we are in no position to help them do their parts in the plan developed by the spy agency in accordance with which we are all acting. Non-responsiveness occurs when we have the knowledge and the means to aid another without compromising our own role but fail to provide aid. An example is Bratman's case of the Ayn Rand singers (2014, p. 120), who sing together intentionally but are not prepared to the aid any others if they falter, e.g., by slowing the tempo, or adjusting to a missed note, or synchronizing intonation. Each thinks, "I am not a servant to their needs." This can of course be a matter of degree, from complete unwillingness to render aid to being willing to render some aid if the success of the enterprise depends upon it, to being willing to render a lot of aid but not as much as one could. Participants still cooperate, they still act together inten-

tionally, they still intend to act in accordance with a common plan, even when they will give no aid, but their commitment to carrying out the plan is limited to their doing just what is strictly required of them, with the result that the enterprise becomes brittle because it is liable to failure due to small deviations from the canonical plan.

(b) Partial sabotage. The case of partial sabotage is more complicated. Someone who pretends to be part of a group that is doing something but is instead seeking to undermine their success is not in any sense cooperating with them. Seeking to sabotage what a group is doing seems to be incompatible with cooperating with them, and incompatible with sharing an intention with them (regarding the task). How can partial sabotage be any different? Examples help to show how. Partial sabotage is not uncommon. Suppose that two brothers, Jeb and Earle, are tasked with raking up all the leaves from the lawn, but each wants to have more of the leaves in his pile, to show he has worked the hardest, and so each, when the other is not looking steals leaves from the other's pile, but both also intend that in the end they rake up all the leaves into two piles together. When they finish there are more leaves in Jeb's pile than in Earle's. Still, they raked up the leaves together intentionally, despite the fact that Jeb not only did not aid Earle in doing his part but undermined (some of) his contributions—and vice versa. It is clear that this occurs in many ordinary contexts where a group has to cooperate to get something done but some members of the group also do things to interfere with the contributions of others, often in pursuit of more recognition, praise, or attention, though not to the extent that it risks failure in the joint enterprise. This is not modest sociality or shared cooperative activity in Bratman's sense, and it may seem puzzling how it could conform to the shared plan account of joint intention. It is clearly not part of their plan that, in our example, Jeb should be undermining Earle's contributions, or that Earle should undermine Jeb's, and it seems, moreover, that it is a violation of the plan that they are putatively executing together.

The solution is to appeal to the idea that the plan that they share is characterized by a margin of error around the canonical plan conception, which was mentioned at the end of Sect. 2.3. In individual intentional action things often don't go exactly according to the plan that we would have described prior to the undertaking. I intend to shoot someone between the eyes to kill him, but my bullet goes through his forehead. I still shot and killed him intentionally. The reason is that this is within the intended margin of error of my canonical plan for killing him. The same idea extends to shared intention. If the way something comes about is within the margin of error of our canonical plan conception, it counts as coming about in accordance with the plan we had in mind for the purposes of doing something together intentionally. The plan is the canonical plan plus the margin for error. In the case of Jeb and Earle, the reason we judge that they the raked the leaves up together intentionally is because we think that what Jeb/Earle does, while it would be frowned upon by Earle/Jeb, is still within the margin of error of their canonical plan. How do we know this? What counts as within the margin of error is determined by those who intend, and so it is revealed in what they say about when things go close enough to

the canonical plan. Our reactions to cases rest on our understanding of what people in the sorts of circumstances described typically allow as within the margin of error.

Thus, the shared plan account of joint intention can accommodate these cases as ones in which the participants intend to bring something about together in accordance with a shared plan. Cooperation in cases of partial sabotage is cooperation on working together enough to make it the case that the intended goal is reached.

It can serve as partial confirmation of this that we can imagine that both Jeb and Earle are trying to undermine the other and that they know this. We do not lose the sense that they are also aiming to cooperate, and if it were crucial to aid the other to get it done, they would do so. So clearly it is not incompatible with their sharing a plan to do it together that they also seek to set the other back, if not completely.

2.8 Halfhearted or Grudging Effort

Different from either failures to render aid, as needed and where possible, and active interference with what others do, is halfhearted or grudging engagement. This too is a very common feature of cooperative agency. Full cooperation requires that each participant in a joint activity engage with a degree of effort appropriate for the task. What counts as appropriate is likely a matter of standard expectations about a task and can be expected to be subsumed in the canonical plan conception.

An example of halfhearted engagement is the work slowdown at a factory. Workers do not cease to cooperate at all with management. They do not strike or quit, but they cease to engage with the degree of effort which is normal for the activity. In this case it is a strategic decision designed to put pressure on management in negotiations. But it need not be engaged in as a matter of strategy by everyone participating in an effort. Some may not give full effort in a task while others do. When some do not, they are not fully cooperating, and the activity is not a fully cooperative activity.

There are many examples of lack of full engagement. Passive-aggressive behavior in the work place or in family life is lack of full effort in a joint activity. Simple laziness or doing only the minimum that one is required to do as a part of one's job are also examples of failure to cooperate fully in the activity. A particularly striking example is disaffection with an enterprise. When workers are demoralized by working conditions, they may cease to be invested in their work and so deliver less than the expected level of effort. Citizens in a country in which the power structures are widely reviled may cooperate with its institutions but never give full effort, only the minimum required to satisfy necessary forms.

In these cases the sense in which they still cooperate is captured by the fact that they intend to contribute to bringing something about in accordance with a shared plan. The sense in which it is not full cooperation is captured by the fact that the canonical plan conception requires more effort and the fact that their failure to give more effort is not because it is not within their power to do so. If they fail to give the degree of effort expected because of factors that are not in their control, it does not

count as a failure to cooperate fully. But if they can put in the expected effort (and know what is expected) but do not, then it is a failure to fully cooperate.

2.9 Cooperation and Game-Theoretic Cooperation

So far we have looked at cases that combine a failure to cooperate fully with some degree of cooperation and a joint intention to do something together. But can there be cases we would call cooperation in which there is no shared intention?

In decision theory strategic reasoners are often called cooperators if they opt for strategies that are Pareto optimal (see note 4). We can illustrate this with the Prisoner's Dilemma, given in Table 2.1. The first number represents the utility for Agent 1 and the second for Agent 2. The numbers represent relative payoffs, depending on what the other does, for the two actions open to each: to confess to a crime they committed together, or to remain silent. If each confesses, they each get a moderate fine, if each remains silent, they get a light fine, but if one confesses and the other is silent, the one who confesses gets off and the other gets a prison sentence.

If they both remain silent, they are called cooperators. We can't read too much into this, because this is technical terminology whose meaning is given just by the structure of the payoffs and choices. But it might be argued that this captures a minimal ordinary sense of cooperation as well. It is particularly natural, perhaps, to think of cooperation as involved in cases in which a group of people face a coordination problem, and settle into a practice or regularity which de facto solves the coordination problem, on the basis of individual strategic reasoning

This is roughly David Lewis's theory of conventions (Lewis 1969). According to Lewis, a convention is a regularity in a community supported by the preferences and beliefs of its members that solves a coordination problem. Let us call a set of assignments of actions to a group of agents relative to some circumstances C a C-assignment. Suppose there is a subset of the C-assignments open to a group, which we call the optimal assignments, C_1, C_2, … C_n, for which the payoffs for the members of the group are roughly the same, and better for all of them than any of the C-assignments not in that subset. Each member has a preference that they all adopt one of C_1, C_2, … C_n, but each has a motive to act in accordance with the assignment for any one of those C-assignments only if all of the others do as well. In these circumstances they face a coordination problem. Call an optimal assignment a solution. All members of the group have a preference that they adopt a solution to

Table 2.1 Prisoner's dilemma

		Agent 2	
		Confess	**Remain silent**
Agent 1	**Confess**	−2,−2	0,−3
	Remain silent	−3,0	−1,−1

Table 2.2 Call drop coordination problem

		Called	
		Call back	**Wait**
Caller	**Call back**	0,0	1,1
	Wait	1,1	0,0

the coordination problem, but they have no preference between the solutions as long as all the others adopt it as well.

For example (from Lewis), a coordination problem arises when one person, A, calls another, B, and the call is dropped. If they both attempt to call back, then they each receive a busy signal. Thus, to reconnect, either A must call and B must wait for A's call, or B must call and A must wait for B's call. The preferences of A (caller) and B (called) are shown in Table 2.2.

The first number represents the caller's payoff and the second the called person's payoff. There are two optimal C-assignments (the circumstances being that they are in a call and it is dropped). They are indifferent between them.

Suppose that they settle into a regularity of the caller calling back and the called waiting. They each prefer that they adhere to that regularity provided that the other does (or in the general case that everyone else does). That this is so is common knowledge. In these circumstances Lewis says that there is a convention among them. If there being a convention entails that those who are party to it are cooperating with each other, and Lewis is right about convention, then, since these circumstances do not ipso facto entail that they share an intention to act in accordance with the rule that the caller call back when a call is dropped, the account I have given of minimal cooperation is incorrect. This is, I think, the strongest case for a decision-theoretic structure that might count as cooperation without shared intention.

There are two separable issues. The first is whether Lewis is right that in these circumstances a convention is present. But aside from this, even if he is wrong about this, there is the question whether in these circumstances, if they do not share an intention to follow the rule, but they each follow it, they are cooperating with each other.

On the question of whether they have adopted a convention, I think the answer is 'no', but I will not argue for this here, as I have written about it elsewhere (Ludwig 2017, ch. 9 sec. 1). Conventions of the sort that Lewis was interested in, on my view, are present in a community when they share an intention to act in accordance with a rule expressing a solution to a coordination problem whenever any of them are faced with it. Thus, to say that the callback convention has been adopted in a community is to say that its members share an intention to implement the callback rule whenever any of them are in a situation in which one is the caller and the other the called and the call is dropped. Conventions of this sort then do entail both cooperation and joint intentional action. But even if this is the right way to think about conventions of this sort, there remains the question whether, when a regularity that is a solution to a coordination problem is not supported by a shared intention, those who act in accordance with it are cooperating.

I think that the answer is that they are not. To see this, we need to distinguish between coordinated behavior and cooperative behavior. Cooperative behavior entails some degree of coordination in behavior, sometimes only by its aiming at the same goal (e.g., fund raising). But coordinated behavior does not entail cooperation. Coordinated behavior is collective behavior in which the actions of each or most of the members of a group are responsive to those of the others. While we often think of coordinated behavior as being intentional, or as having a goal, this need not be so. For example, the phenomenon of entrainment involves coordinated behavior. If everyone in a room practicing breathing meditation comes to breathe in and out at the same time, their breathing becomes synchronized. Synchronized behavior is a form of behavioral coordination. In this case it need not be intentional. It need not be that individuals share an intention to coordinate their breathing with others, and it need not be that each individually even has an intention to synchronize his breathing with the others. The processes involved can be below the level of intention or conscious awareness. Even if, however, each one intended to coordinate his breathing with the others, there would not need to be a shared intention among them to do so. For example, each may have read an article on entrainment, and expect there to be a tendency for everyone to settle into the same pattern of breathing, and aim to coordinate his own behavior with what others are doing, both to ensure that his breathing patterns conform and to help the entrainment process. In some cases coordination of behavior can occur in the face of conflicting aims. If lions from different prides are each stalking the same antelope, they will adjust their behavior in light of the behavior of the lions from the other pride, approaching, e.g., from different directions, balancing stealth against the speed of the competition, each group hoping to claim the kill before the others. Prey and predator coordinate behavior without cooperation as well. Cape buffalo turn to face lions, and turn as the lions circle them. The coordination of behavior is of the same sort one might expect in a dance, but it is not per se cooperative behavior.

What must be added to coordination to get cooperation? Cooperation implies a shared goal, even if minimal. But it is not enough simply that the members of the group have the same goal and their various behaviors are coordinated toward that goal. Return to the case of synchronized breathing. If each member of a group engaged in breathing mediation has the goal of their coming to breathe in the same rhythm, but they are unaware that this is a goal each of them has, they are not cooperating, even though, in the distributive sense, they share the same goal with respect to coordinated group behavior. Suppose we add that they all know that the others have the same goal. Does this entail that they are cooperating with each other? First, all that happens is that they become aware of something that is not sufficient for cooperative behavior. Becoming aware of something insufficient could hardly make it sufficient. Second, it is true that if they are all aware that all of them have the same goal, then cooperating would be a *reasonable* thing for them to do. But the fact that it would be reasonable for all of them to cooperate given that they know that they all share the same goal shows that knowing that they share the same goal is not ipso

facto to cooperate. And in this case it is not clear that there is any sufficient reason for them to do more than what they would have done otherwise, for they can all see that that will be sufficient for the goal they each have. They can each then have the intention to synchronize their breathing with the others, come to know that the others have the same intention, but not come to have the intention to cooperate with them. The extra step is that they should come to share, collectively, the goal of synchronizing their breathing and then act on that, but that is simply for them to come to share the intention to synchronize their breathing. The crucial step then from coordination of behavior, even toward a goal pertaining to the group that each member of the group has individually, to cooperating is to form a joint intention to pursue that goal, and, hence, to each intend to contribute to their bringing it about in accordance with a shared plan.

In the case of Lewis's account of a convention, it is natural for those faced with a coordination problem, in conditions under which it is common knowledge, to come to share an intention to solve it and then to come to intend to act in accordance with rules appropriate for one of the ways of solving it. It is difficult to imagine ordinary cases without the overlay of a shared intention. Consequently it can be difficult to judge, because it is difficult to isolate in our thinking, whether in the absence of a shared intention, those engaging in a practice that solves a coordination problem are cooperating or merely coordinating their behavior. However, the case is not in principle different from the case of synchronized breathing. In principle two people could coordinate their behavior, e.g., to make a hollandaise sauce, to borrow an example from Searle (1990), one pouring melted butter into beaten egg yolks, while the other stirs the mixture, without exactly cooperating, if each is merely taking advantage of what the other is doing, even if this is out in the open between them. When each simply treats the other as a state of nature, rather than an agent with whom he is acting intentionally, then they fall short of what is required for cooperation, which requires that each see the other, to put it circularly, as cooperating with him.[19] The trouble with the conditions that Lewis provides is that they are compatible with everyone in the group regarding the others as if they were states of nature. But cooperation requires acknowledging the others as co-participants, intending, as one does oneself, that the group follow a particular practice jointly intentionally—in this case, to solve a coordination problem. The game-theoretic conception of cooperation is, as it were, the third person perspective on rational cooperation among essentially self-interested individual strategic reasoners. It treats cooperation as a matter of patterns of behavior directed toward a structure of individual ends that are group optimal (in one sense). Here the behavior is called cooperative because the end is conceived of as a pro-social cooperative goal. The game-theoretic investigation of cooperation is largely focused on when it makes

[19] Each treats the other as one would a machine that poured/stirred while one stirred/poured in a way that took into account feedback on one's rate of stirring/pouring.

sense for individual strategic reasoners to coordinate on the group optimal goal.[20] But this confuses the ground for rational cooperation with what it is a ground for. Of course, since these ideas were developed prior to there being an adequate theoretical framework for understanding joint intentional action, it is not surprising that cooperation was characterized in terms of the available tools.[21] This is an instance of Maslow's Law of the Instrument: If all you have is a hammer, everything looks like a nail.

2.10 Cooperation and Coercion

Surely it is a problem for the view that joint intentional action is per se cooperation that joint intentional action can be coerced. For surely coercion and cooperation are opposed to one another. Thus, if I coerce others to help me carry a piano upstairs by threatening to torture them if they do not, while we carry the piano upstairs together intentionally, we do not cooperate in doing so.

I think this is a mistake.[22] Coercing others to cooperate with you or with each other is not an oxymoron. In these circumstances we would deny the others were cooperating with you or each other *voluntarily*, but we would not deny that they were cooperating. Coercion provides reasons for doing something by placing a cost on not doing it with the intention that one (as the coercer) should be recognized as the conditional agent of that and that it is done with the intention that you do the thing to avoid the cost. This provides others with reasons for cooperating, or for

[20] The situation is not different if we shift to the Team Reasoning framework (Bacharach 2006; Gold 2018). Team reasoners are still individual strategic reasoners. It is just that their goal is the team optimal outcome (or some outcome defined in terms of the group). But this isn't enough to ensure that we see each other as co-participants sharing an intention to bring about the goal, contra (Gold and Sugden 2007). Searle's business school graduates who intend to promote general welfare by pursuing their own selfish ends without reference to anyone else (accepting the theory of the "invisible hand") are not cooperating and do not share an intention (1990). But they do have (what they think of as) the team optimal goal in mind and each acts to promote it as their ultimate end rather than merely individual goals. Since this is also a case of a mutually beneficial practice (as they believe) it appears to fit the Schema of Cooperative Team Reasoning in (Sugden 2015, pp. 161–162) as well.

[21] It might be objected that even if I have captured the ordinary notion, or at least one ordinary notion, of social cooperation, the game-theoretic notion is theoretically more fruitful because it is broader and subsumes more phenomena. It depends, though, on what one is theorizing about. For certain purposes, the game-theoretic notion may be exactly what is required. But there is a richer notion, which plays a role in our self-understanding, and to mistake *it* for a different notion would not be a theoretical advantage. It would instead miss an important distinction that is actually at play in our thinking, which plays a role in other concepts we deploy in thinking, planning, and acting, in relation to the social world. And since how we conceptualize the social world is constitutive of its nature, we should not elide this category in our theorizing if we are interested in understanding human social reality.

[22] Here I agree with Tuomela (1993, p. 99).

doing something with you or with each other intentionally. But the reasons why people cooperate are external to the activity on which they cooperate. We can cooperate for its own sake or for the sake of something that cooperating helps bring about. Typically we cooperate for the sake of something else. We may cooperate because we are each being paid to do so, or we may cooperate to get a greater benefit than we could get if we did not, or to avoid harm which we could not otherwise avoid. For the latter, suppose that it is a matter of drilling a well so that we have water during the dry season and that it is a matter of survival that we do so. The difference between this and coercion is that it is the natural world that provides a negative incentive for non-cooperation rather than an agent, who may or may not be a member of the group.

If this is granted, it might still be thought that coercion in favor of cooperation represents a deviation from full cooperation. Bratman, in his (1992) paper "Shared Cooperative Activity," required that no coercion be present as a condition on shared cooperative activity. It seems clear that there is something infelicitous about coerced cooperation. Those who are coerced would not do their parts but for being under threat. There is nothing improper about defining a form of joint intentional action that precludes any coercion. This is in some respects a happier form of joint intentional action. But is it more cooperative? I think the tendency to say that it is comes from thinking that typically cooperation is directed toward ends that are both mutually beneficial and beneficial for each in the same way. And that is missing in the case of coerced cooperation. But full cooperation does not require that everyone benefit in the same way. Some may join a cause because they are true believers, others for power, others for monetary gain, others to avoid being shamed. This does not entail that they are less cooperative than if they were all to benefit in the same way. It seems possible for someone to coerce another into cooperating by issuing a threat that will be carried out if they do not *cooperate fully* (we can suppose the coercer is a telepath and that the person being coerced knows this). This does not seem to be a self-defeating threat, as it would be if cooperating under threat precluded full cooperation.

2.11 Cooperation and Rationality

What about rationality, though? That rationality should be a requirement on cooperation is suggested by the game-theoretic conception of cooperation as coordination with others to arrive at a group optimal solution. I have already argued that the game-theoretical notion is not what I have called social cooperation because it is compatible with its absence. But the reason it is too weak also shows that it is too strong, for insofar as joint intentional action is ipso facto cooperation, cooperation is compatible with irrational action. If an individual can do something intentionally though it is not rational for him to do it, a group can do something intentionally together though it is not rational for them to do it. For example, if one can do something intentionally that one knows at the time is not the best thing to do (weakness

of the will), then a group can do something together intentionally while recognizing at the time that it is not the best thing for them to do (group-level weakness of will). For example, a group of people may organize to elect a candidate on the basis of his emotional appeal to them though they know that his policies will make them worse off. They cooperate to elect him, but they are not acting rationally in doing so, and they know that they are not. A mob may attack a police station in anger over the shooting of an unarmed motorist at a traffic stop, even though they know, and it is common knowledge among them, that it is not a constructive response and that it will only make things in their community worse. Two people may know that it is bad idea to pursue a relationship, and yet they do it anyway, fully intentionally. Every form of irrational intentional behavior open to individuals is open to groups of individuals because all it requires is that the individuals share the relevant attitude structures directed toward a joint end. Since intentional coordination toward a mutually desired end (even if irrational) that requires contributions from all is a paradigmatic form of cooperation, cooperation does not require rationality.

2.12 Conclusion

The goal of this paper has been to identify minimal conditions for cooperation. The thesis is that the minimal conditions for cooperation are also the minimal conditions for joint intentional action. A subsidiary thesis is that this takes a particular form: joint intentional action is action that results from a group successfully executing an intention to act in accordance with a shared plan. A group intention to act in accordance with a shared plan is reduced to each member of the group intending to contribute to their acting in accordance with a shared plan. The strategy of the paper has been to start with a characterization of full cooperation and consider various ways in which cooperation could depart from the ideal without ceasing to be cooperation and to show in each case that it involved joint intentional action and an intention to act in accordance with a shared plan. Full cooperation is characterized by (1) shared intention, (2) absence of conflict, (3) willingness to aid ((a) non-interference, (b) mutual aid, (c) attention to others), and (4) full effort, in the sense of a degree of effort commensurate with the standards expected for the type of action.

1. Violations of the absence of conflict condition (1) include competitive games governed by constitutive rules, conflicts like war governed by regulative rules, and conflicts that presuppose forms of cooperation.

   (a) Competitive games are essentially joint intentional actions. On the shared plan account of we-intentions, the key is that the plan that participants intend to follow (i) includes that they compete with respect to some goal but (ii) does not include anything about who wins.
   (b) For conflicts governed by regulative rules, like the laws of war, the key is that while the activity governed is by its nature non-cooperative, cooperation is possible in matters ancillary to the opposed goals of the warring parties: (1)

excluding weapons whose destructive scope cannot be limited, (2) dealing with circumstances that war occasions, such as the treatment of civilians, the sick and wounded, and the enemy dead, (3) provisions for taking and treating prisoners, negotiating surrenders, and so on. These forms of cooperation involve shared intentions, shared plans, and collective intentional action. In some cases this involves negative action directed toward not doing things—not using indiscriminate biological weapons, for example. In some cases it involves active cooperation with the opposing side. But the areas of cooperation presuppose a non-cooperative activity as their setting. In contrast to competitive games, the conflict is not embedded in the concept of what they cooperate on, but rather the non-cooperative activity is presupposed by the regulative rules.

(c) Some forms of conflict presuppose cooperation. Any form of conflict that involves communication essentially is an example, such as a verbal fight where the participants have no common goal. Other examples are jury trials in which the defense and prosecution have opposing aims, and dueling governed by a code duello. In these cases cooperation puts the parties in a position to engage in a non-cooperative activity. In the matters where they cooperate, they do so intentionally and intend to act in accordance with a common plan in the relevant matters.

2. Distinct from cooperation surrounding conflict in one way or another is failure of full cooperation on an intended goal. The main varieties are violations of (3), either unwillingness to aid or active interference, on the one hand, and violations of (4), halfhearted or grudging effort on the other.

   (a) The two main sorts under the first heading are non-responsiveness in providing aid and partial sabotage of what others engaged in the shared task are doing.

   (i)The former is straightforward: the parties share an intention and intend to act in accordance with a common plan, but will not help if needed.
   (ii) The latter is more puzzling, for interference seems inconsistent with the intent to act in accordance with a common plan, but is resolved when we take into account the allowed margin for error around the canonical plan conception.

   (b) Halfhearted or grudging effort, as in the work slowdown, is a matter of deliberately providing less than the expected effort in one's role in a joint task. It is still in accordance with the plan but does not conform to the canonical plan conception. It still counts as cooperation because it falls within the margin of error of the canonical plan conception.

3. A final challenge is the claim that certain sorts of coordinated behavior patterns that are stable because of interlocking beliefs and preferences of those engaged in it, without their sharing an intention, count as cooperation. We distinguished

between coordinated behavior and cooperation, however, and showed through an examination of cases that the crucial difference comes down to whether or not those participating share an intention to act toward a common interest via a common plan.

Minimal cooperation is consistent with coercion, as this pertains to the reasons for it, and does not require the group be acting rationally, any more than intentional action requires this in the individual case. Thus, minimal cooperation consists in meeting the minimal conditions for collective intentional activity, which is doing something together with the intention that those acting should do it in accordance with a common plan, defined by a margin of error around a canonical plan conception.[23]

References

Bacharach, M. (2006). *Beyond individual choice: Teams and frames in game theory*. Princeton: Princeton University Press.

Blomberg, O. (2015). Common knowledge and reductionism about shared agency. *Australasian Journal of Philosophy, 94*(2), 315–326. https://doi.org/10.1080/00048402.2015.1055581.

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. (2014). *Shared agency: A planning theory of acting together*. Oxford: Oxford University Press.

Fiebich, A. (2017). Social cognition, empathy and agent-specificities in cooperation. *Topoi, 38*(1), 1–10. https://doi.org/10.1007/s11245-017-9480-x.

Gilbert, M. (2006). *A theory of political obligation: Membership, commitment, and the bonds of society*. Oxford: Clarendon Press.

Gilbert, M. (2013). *Joint commitment: How we make the social world*. New York: Oxford University Press.

Gold, N. (2018). Team reasoning. In M. Jankovic & K. Ludwig (Eds.), *The handbook of collective intentionality*. New York: Routledge.

Gold, N., & Sugden, R. (2007). Collective intentions and team agency. *Journal of Philosophy, 104*(3), 109–137.

Lewis, D. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard Univ Press.

Ludwig, K. (2007). Collective intentional behavior from the standpoint of semantics. *Noûs, 41*(3), 355–393.

Ludwig, K. (2014). Proxy agency in collective action. *Noûs, 48*(1), 75–105. https://doi.org/10.1111/nous.12013.

Ludwig, Kirk. 2015a. Methodological Individualism, the We-mode, and Team Reasoning. In: *Social ontology and collective intentionality: Critical essays on the philosophy of Raimo Tuomela with his responses* (pp. 3–18). Cham, Springer.

Ludwig, K. (2015b). Shared agency in modest sociality. *Journal of Social Ontology, 1*(1), 7–15.

Ludwig, K. (2016). *From individual to plural agency: Collective action 1*. Oxford: Oxford University Press.

Ludwig, K. (2017). *From plural to institutional agency: Collective action 2*. Oxford: Oxford University Press.

[23] My thanks to Daniel Creech for comments and proofreading.

Nguyen, C. T. (2017). Competition as cooperation. *Journal of the Philosophy of Sport, 44*(1), 123–137.
Paternotte, C. (2014). Minimal Cooperation. *Philosophy of the Social Sciences, 44*(1), 45–73.
Searle, J. R. (1969). Speech acts: An essay. In *The philosophy of language*. London: Cambridge University Press.
Searle, J. R. (1990). Collective intentions and actions. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication*. Cambridge, MA: MIT Press.
Sterelny, K., Joyce, R., Calcott, B., & Fraser, B. (2013). *Cooperation and its evolution*. Cambridge, MA: MIT Press.
Sugden, R. (2015). Team reasoning and intentional cooperation for mutual benefit. *Journal of Social Ontology, 1*(1), 143–166.
Tuomela, R. (1993). What is cooperation? *Erkenntnis, 38*, 87–101.
Tuomela, R. (2000). *Cooperation: A philosophical study* (Philosophical studies series) (Vol. 82). Dordrecht/Boston: Kluwer Academic Publishers.
Tuomela, R. (2005a). Cooperation and trust in group context. *Mind and Society, 4*(1), 49–84.
Tuomela, R. (2005b). We-intentions revisited. *Philosophical Studies, 125*, 327–369.
Tuomela, R. (2007). *The philosophy of sociality: The shared point of view*. Oxford: Oxford University Press.
Tuomela, R. (2010). Cooperation as joint action. *Analyse & Kritik, 33*(1), 65–86.
Tuomela, R. (2013). *Social ontology: Collective intentionality and group agents*. New York: Oxford University Press.
Tuomela, R., & Miller, K. (1988). We-intentions. *Philosophical Studies, 53*(3), 367–389.
Zwicky, A., & Sadock, J. (1975). Ambiguity tests and how to fail them. In *Syntax and semantics* (pp. 1–36). New York: Academic.

Chapter 3
Joint Action: Why So Minimal?

Cédric Paternotte

Abstract The repeated attempts to characterise joint action have displayed a common trend towards minimalism – whether they focus on minimal situations, minimal characterisations, cognitively minimal agents or minimal cognitive mechanisms. This trend also appears to lead to pluralism: the idea that joint action may receive multiple, equally valid characterisations. In this paper, I argue for a pluralist stance regarding joint action, although one stemming from maximalism. After describing three cases of "maximal" joint action – demonstrations, deliberations and free collective improvisation – that stretch our conceptual characterisations of joint action, I introduce and defend contextual minimalism, which focuses on joint actions occurring in contexts from which the factors that typically favour successful cooperation are absent. Although maximalist as compared to the other forms of minimalism, contextual minimalism does fit the minimalist trend and its recent emphasis on specific cognitive cooperative mechanisms.

Keywords Joint action · Minimalism · Pluralism · Contextualism · Mass action · Improvisation

Doing things together is often easy. Each of us cooperates hundreds of times with others on a weekly basis. A conversation requires many small acts of successful coordination; working for a salary involves taking part in a collective enterprise; and so on. For sure, there are also social dilemmas, seized opportunities to free ride, misunderstandings, awkward dialogues, scattered gatherings, deserted demonstrations, messy parades. However, by and large, successful cooperation is our norm.

That cooperative issues are rare has influenced our understanding of joint action in general. If acting together is simple, then we may hope that it proves relatively easy to understand and is characterised in a somewhat straightforward way. Of course, the ubiquity of a phenomenon offers no guarantee for its conceptual simplicity; and many naive, familiar concepts have proven elusive under philosophical

C. Paternotte (✉)
SND Research Team, Department of Philosophy, Sorbonne Université, Paris, France

A. Fiebich (ed.), *Minimal Cooperation and Shared Agency*, Studies in the Philosophy of Sociality 11, https://doi.org/10.1007/978-3-030-29783-1_3

scrutiny. But we also do not tend to find cooperation particularly mysterious. In fact, it often takes a contact with some literature on the topic to start being puzzled by the very possibility of joint action. Overall, the alliance of ubiquity and straightforwardness of joint action lets us think that philosophers too could have it easy.

Indeed, most approaches to joint action – even recent ones – have been striving to provide *minimal* characterisations. They are based on simple, intuitive cases. They tend to focus on normal individuals in normal contexts. They involve but a handful of concepts. They strive to identify a limited number of necessary conditions or of underlying mechanisms. They aim for frugality.

This raises a number of worries though. Is minimalism the right choice? How do we choose between competing minimal standards? May these extant forms of minimalism not lead us to neglect legitimate, crucial aspects of joint action? And how can we tie up these minimal standards with more complex cases of joint action? In short, can minimalism be defended rather than assumed?

In what follows I question the quest for minimalism and argue for a pluralist stance regarding joint action that embraces a kind of maximalism. I argue that minimal standards are multifarious (Sect. 3.1); and that each standard is in turn compatible with multiple characterisations (Sect. 3.2). I then discuss the links between minimalism and pluralism about joint action (Sect. 3.3), before describing a few cases of "maximal" joint action that stretch and should enrich our conceptual characterisations (Sect. 3.4). I finally argue for contextual minimalism, which takes such cases seriously and mixes both minimalism and maximalism, although in different senses, by focusing on the specificity of human joint action (Sect. 3.5). More specifically, contextual minimalism focuses on cases of joint action in which factors that typically favour it are absent.[1]

3.1 Kinds of Minimalism

The search for minimal characterisations of joint action is far from being a unified enterprise, and there is no hope of assessing it without describing it in all its variety. While it is fair to say that most influential accounts of joint action have embraced minimalism (as we will see), it comes in various guises, and stems from a variety of motivations.

Historically, minimalism has been present even in early attempts to characterise joint action.[2] It first appeared with the emphasis put on the analysis of *small-scale joint actions*, that is, joint actions that involve a small number of interchangeable agents and of individual actions (as well as other occasional restrictions). *Scale minimalism* focuses on such small-scale activities, famous instances of which are

[1] Such factors include commitments, promises, reputation, publicity, shared goals with clear conditions for success, common background values, beliefs or interests; see below.

[2] Although not in the very first analysis of joint action, namely Tuomela and Miller (1988). Tuomela's is the main body work that bucks the trend of minimalism.

Gilbert's (1989, 1990) example of walking together, Bratman's (1992, 1993) cases of singing a duet or painting a house together and Searle's (1990) preparing a hollandaise sauce together. They are best captured by Bratman's more recent description of 'modest sociality':

> Cases of modest sociality are cases of small scale shared intentional agency in the absence of asymmetric authority relations. Examples include: our singing a duet together, but not the activities of an orchestra with a conductor; our going to NYC together, or (to use Margaret Gilbert's example) our walking together, but not a school trip led by a teacher; our having a conversation together, but not an exchange in which a language teacher guides a novice. Modest sociality is a deep feature of our lives, something we frequently value both for its own sake and instrumentally. (Bratman 2009: 150)

Such simple cases work well as intuition pumps or pedagogical devices. But their recurrence also stems from the hope that understanding simple cases of joint action will pave the way for more complex cases. Simple cases may be considered as situations in which only a few factors are at work and so as a way to isolate them and pinpoint their role, which would be obscured in socially richer contexts. Bratman reveals precisely such a hope: "Reflection on the underlying structure of such modest sociality may also help us think about larger scale cases" (Bratman 2009: 151). A still more ambitious hope may be that nothing fundamental will differ between the analysis of small-scale and large-scale cases. Gilbert explicitly defends this view by developing her analysis of walking together before claiming that "a plausible account of social groups in general can be given in similar terms." (Gilbert 1990: 1). This claim, however, is only supported by analogies between the vocabulary used in her account of small-scale joint action and the terms found in Hobbes and Rousseau's writings about association and commonwealth (Gilbert 1990: 10–11).

A second kind of minimalism consists in striving for economy regarding the number of concepts or mental attitudes involved in a characterisation of joint action. *Conceptual minimalism* aims for simple characterisations, according to which joint action boils down to a simple conceptual core. For instance, Searle (1990) understands joint action as constituted by individual intentions only (although particular ones); Gilbert (1989) bases her account of joint action on a concept of joint commitment. Conceptual minimalism may seem like a default stance: when we define something, don't we always try to be as economical as possible? The definition of a concept tends to become less convincing as it resorts to additional concepts or additional clauses so as to cover all possible cases. Complex definitions often look artificial or *ad hoc*, especially when their complexity is at odds with the simplicity of the *definiendum*; and joint action seems particularly simple at first glance. Even if nothing ever guarantees that a simple definition of anything may ever be obtained, conceptual minimalism may be a reasonable heuristic strategy when looking for reasonable characterisations.

Ontological minimalism constitutes a third option, according to which the number of potentially problematic entities needed in a characterisation should be kept to a minimum. In the case of joint action, this amounts to a commitment to individualism – the refusal to refer to intrinsically social entities or properties in order to

define joint action. This is because the existence of individuals is never doubted, and they are the only entities in this case; so individuals constitute the minimally acceptable ontological basis. Ontological minimalism is less widespread kind than the previous ones. In fact, only a few authors manage, or even try, to provide wholly individualistic accounts. One example would be Miller's (2001) definition of a joint action, based on "the notion of a collective end [, which is] an individualist notion" (2001: 58). While some authors embrace holism and are happy that their characterisations display the irreducibly social nature of joint action (e.g. Gilbert 1989, Tuomela 2013), others strive to stay as close as possible to individualism (e.g. Bratman 1992, 1993). The main motivation for ontological minimalism is a classic one: it aims to avoid explanatory circularity when dealing with social phenomena and presupposes a desire to provide an analysis of joint action – a decomposition in more fundamental elements (hopefully necessary and sufficient conditions for it).[3]

The fourth kind of minimalism is also its most conspicuous recent instance. *Cognitive minimalism* holds that a characterisation of joint action should fit the cognitive abilities of the less sophisticated agents that are able to act jointly. It stems from a number of distinct ideas. First, if cognitively bounded agents are able to act jointly, then characterisations of joint action should not assume abilities that they do not possess. The usual example is that of children, who it is argued, are able to display a range of cooperative behaviours from the end of their 2nd year (Warneken et al. 2006). As a consequence, characterisations of joint action should only invoke mental attitudes or abilities that are present in young children. Such arguments are typically provided in a criticism of traditional characterisations, typically Bratman's, as they typically involve higher-order beliefs about the intentions of others. As a consequence, simpler characterisations are offered, which avoid the use of demanding mental attitudes (Tollefsen 2005; Butterfill 2012; Pacherie 2013). A second motivation for cognitive minimalism stems from the focus on coordination and cooperative mechanisms in experimental social psychology (Knoblich et al. 2010). Philosophical characterisations of joint actions have typically abstracted from such mechanisms; however, their variety, efficiency, and the fact that a number of them operate at the motor level and do not involve propositional attitudes, casts doubt on any characterisation of joint action that neglects the motor level. Here, traditional accounts are not criticised for being too cognitively demanding, but, similarly to the children case, for focusing on a limited subset of elaborate joint actions and so for providing at best partial accounts.[4]

This list, which is not intended to be exhaustive, already leaves us with four kinds of minimalism: scale minimalism, conceptual minimalism, ontological minimalism, cognitive minimalism. Although not mutually exclusive, they are also

[3] Of course, this does not prevent analyses from being entirely compatible with holistic approaches as well.

[4] Using the vocabulary of constraints rather than of minimalism, Paternotte (2014b) labels these two strands of cognitive minimalism as the 'developmental constraint' and the 'motor constraint' for joint action.

rarely all adopted simultaneously. As a result, the same authors may be labelled as minimalists or maximalists depending on one's favoured sense of the term.[5] Bratman is a scale, maybe ontological, but not conceptual or cognitive minimalist. Gilbert is a scale, conceptual but not ontological minimalist. Kutz (2000) and Paternotte (2014a) call their approaches minimal but are not scale minimalists, as they focus on large-scale actions. Tuomela, who understands all joint actions in terms of a specific perspective (the we-mode), may only be a minimalist in the conceptual sense. Several approaches labelled as 'minimal' may be at odds. And as we will see, there are also discrepancies among minimalists of the same kind.

3.2 The Looseness of Minimalism

Minimalism is not a unified enterprise for at least two reasons. First, as just established, one can be minimal according to at least four stances. Second, each minimalist stance allows for various minimal characterisations. In mathematics, an element is said to be minimal if there exists no distinct smaller element in the same set.[6] But this does not entail uniqueness: a set may contain several minimal elements. For instance, suppose that the minimum number of voters for a petition to be considered by a government agency is 1000, and that there are 5000 possible voters. Then there are as many minimal sets of voters – sets such that removing any member would nullify the petition – as there are ways to pick out 1000 individuals among 5000.

This point carries over to our four minimalist stances. Scale minimalism is compatible with many paradigmatic cases. Setting a small number of agents (two) and a set of simple individual actions that compose the joint action leaves open a wide range of possible situations. Walking together, painting a house or singing a duet together, pushing a heavy object together, rowing a boat together, talking together, cleaning a courtyard together… seem markedly different. Some are instantaneous, others are extended in time. Some involve simultaneous actions, others alternate ones. Some involve delicate coordination, others broadly complementary tasks. Some involve difficulty and obstacles, others not. Some can emerge naturally, others need some prior interaction. As a result, that scale minimalism has in turn led to otherwise very different characterisations should come as no surprise. Gilbert's joint commitment has little to do with Bratman's conditions of dynamic consistency (see below).

Conceptual minimalism too allows for different options. Let us just consider two examples. Gilbert understands joint action as stemming from shared intention, defined as follows: "Members of some population P share an intention to do A if and

[5] This was already noted by Heinonen (2016), although he considers only two strands of what I have dubbed cognitive minimalism.

[6] To be formally accurate, minimality (and maximality) can be defined as soon as a set is endowed with a partial order, that is, a reflexive, antisymmetric and transitive binary relation.

only if they are jointly committed to intend as a body to do A." (Gilbert 2009: 179). Most of the work is accomplished by the only concept of joint commitment, which captures the normative dimension of joint action, that is, the fact that agents who act jointly are typically under an obligation towards the group to do their part. Joint action is reduced to a single concept, that of joint commitment, which renders the characterisation conceptually minimal. Now consider Searle's take on joint action, which is famous for being non-reductive. For him, in order to understand joint intentions (which he calls we-intentions, and Gilbert shared intentions), we just have to "recognize that there are intentions whose form is: We intend that we perform act A; and such an intention can exist in the mind of each individual agent who is acting as part of the collective case." (Searle 1990: 106). Although Gilbert can be said to provide an analysis of joint action, Searle cannot: the claim that shared intentions amount to several individual intentions and nothing else can hardly be seen as the identification of more fundamental elements. This does not prevent his account from being conceptually minimal, in the sense that substracting from it any mental attitude could not possibly leave us with a characterisation of joint action.

Ontological minimalism is a special case. In principle, different accounts of joint action may equally count as purely individualistic – it would suffice that they rely on distinct individual properties or mental attitudes. Unfortunately, there hardly exist any such account (Miller's is an anomaly). There are, however, nearly individualistic characterisations, that only include one social element, the nature of which may significantly vary. Consider for instance Bratman's classic definition:

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

At first glance, this may seem individualistic, as only individual mental attitudes are involved: individual intentions, compatible individual plans of action, and common knowledge (which can be characterised on the basis of individual knowledge). However, the individual intentions contain a social element, as they are intentions 'that *we* J'. A similar stance is found in one of Tuomela's old definition of intentionally acting together, one of the clauses of which is: "I intend *us* to perform X *together*, and I perform my part of X or participate in the performance of X in accordance with and partly because of this intention." (Tuomela 2000: 74). Here too, the content of individual intentions has a social dimension. For other authors, the locus of the social element may also be the *subject*, as for Gilbert's, who considers that joint commitments are taken by and with respect to plural subjects.[7,8]

[7] The claim that the possible location of social elements in joint action may be the content, the subject or the mode of the intention is made by Schweikard and Schmid (2013). I do not mention the mode here, as it is a more complex notion, found mostly in Tuomela's recent work, and because it in turn involves several social elements. So it cannot be part of an ontological minimal or nearly minimal account.

[8] Note that any characterisation that is nearly individualistic in this sense can be suspect of circularity, as it aims to analyse a social concept by resorting to at least another one.

Finally, cognitive minimalism– the most active kind of minimalism in the recent literature – can be pursued in multiple ways as well, and is compatible with different aims. Some authors have emphasised that some agents with limited cognitive abilities (notably children) nonetheless engage in successful joint actions and have strived to build definitions that would fit such agents by avoiding high-level mental attitudes such as meta-representations (Tollefsen 2005; Butterfill 2012; Pacherie 2013). Others have focused on a host of low-level cognitive mechanisms involved in various forms of cooperation and coordination and are typically not reducible to high-level propositional attitudes (Knoblich et al. 2010; Vesper et al. 2010). Such authors argue that their role in the success of joint actions has been underestimated, but without trying to formulate new definitions. As Heinonen (2016) puts it, these two strands of research correspond to a distinction between top-down and bottom-up approaches, which endorse different heuristics: one may start with a functional notion of joint action and explore the cognitive mechanisms that may realise it; alternatively, one may start by investigating the low-level mechanisms involved in cooperative or coordinated activities and general, without aiming for a definition of joint action.

Interestingly, minimalist approaches display diversity even within each of these two strands. Mechanism-based approaches may focus on a host of interaction contexts, ranging from mere synchronization to planned action, and accordingly identify many distinct mechanisms. Definition-based approaches too may favour various key components, even when they aim to cover the very same cases. For instance, while sharing a similar aim of characterising joint action among children, Tollefsen (2005) emphasises the role of joint perception (as opposed to common knowledge); Butterfill (2012) focuses on expectations and identification; while Pacherie (2013) highlights the importance of team reasoning. As a result, although they share the commitment not only to minimalism but to its cognitive version, these three authors end up with three markedly different characterisations of an allegedly similar phenomenon.

3.3 Pluralism or Monism?

The variety of extant minimalist stances as well as of the solutions available for each stance seem to suggest that joint action should receive a pluralist characterisation. I now examine several possible objections to this claim; later sections will advocate for an extension of such pluralism that would be both minimalist and maximalist, although in different senses.

The multiplicity of joint action characterisation may be thought to shroud a deeper unity, for several reasons. First, existing characterisations may be specifications of a unique one, that is, one of them may entail the others. This would be the case if the particular elements emphasised by a number of characterisations followed as a matter of fact from elements of another one, or if they merely expressed some of them differently. Such links may hold between Bratman's and Gilbert's

accounts. Indeed, when replying to some of Bratman's earlier objections to the necessity of mutual obligations, Gilbert notes that all of Bratman's alleged counter-examples are actually compatible with the presence of such obligations (Gilbert 2009: 177–8). This leaves open the possibility that both accounts are compatible. However, she later distinguishes between the commitments involved in Bratman's account and in hers. The former are individual commitments to act according to one's plan, which can be rescinded unilaterally, while the latter are personal commitments derived from joint ones, which can only be collectively rescinded. As a result,

> the derived commitments involved in shared intention on the plural subject account have a greater stability in terms of revisability and rescindability than both the derived and non-derived commitments involved in any personal intentions account. (Gilbert 2009: 184)

Moreover, Bratman considers the two accounts as disagreeing regarding the characterisation of joint action:

> I agree that mutual obligations and entitlements are extremely common in cases of modest sociality […] But, first, I am not convinced that such obligations are essential to modest sociality. (Bratman 2009: 151)

He later adds about Gilbert that "Our disagreement is […] about how precisely to understand the interrelations among participants that constitute modest sociality." (Ibid.: 164). Overall, the main two classical accounts of joint action, which both embrace scale minimalism and possibly conceptual minimalism, are fundamentally at odds.

Similarly, it is hard to see how cognitively minimalist approaches may be combined or even made consistent. Compare again the characterisations of Tollefsen (2005), Butterfill (2012) and Pacherie (2013): while some share a number of common elements, they disagree regarding others. Tollefsen (2005: 92–3) holds that a definition that fits child joint action may be obtained by replacing, in Bratman's definition (provided in Sect. 3.3), the requirement of common knowledge by that of joint attention or of Peacocke's (2005) open-ended perceptual availability.[9] By contrast, Butterfill (2012: 40) considers joint action as based on shared goals, which are in turn composed of: a single goal, an identification of the other agents, expectations that others will perform their part and expectations of a common effect. This involves no reference to common knowledge or to any related concept (nor does it mention subplans). Finally, Pacherie (2013: 1833) defines shared intentions in terms of agents framing themselves as team members, believing that all so identify with the team, reasoning from a group perspective and thus forming relevant individual intentions. Again, common knowledge is not deemed necessary. The 'identification' refers to identification *with* a group rather than Butterfill's identification *of* other participants. And team reasoning does entail meshing subplans of action, but which must be determined, whereas they can be adapted and modified on the fly according

[9] See Paternotte (2015) for an argument that Tollefsen's and Peacocke's alternatives to common knowledge are actually equivalent to it, since they are only weakenings of an artificial, unduly strong concept of common knowledge.

to Bratman's view.[10] Overall, there is no obvious way to combine any pair of these three accounts.

Are we not asking too much though? Maybe such accounts should not be seen as competing characterisations to be reduced to a unique one but rather as complementary ones. Although Gilbert and Bratman do identify disagreements between them, suggesting that they have similar descriptive and explanatory aims, the cognitive minimalists at least may be collectively salvaged by appealing to complementarity. Pacherie (2013: 1818) explicitly assumes so when she describes her approach as one of 'two broad strategies' for dealing with the 'gap between the dominant philosophical approaches and the empirical investigations of joint agency'. These two kinds of approaches correspond to the two strands of cognitive minimalism mentioned in Sect. 3.2 and labelled, following Heinonen, as top-down and bottom-up approaches. Pacherie adopts a top-down approach, by providing a 'less demanding account of what counts as shared intention' (Ibid.: 1818), while she sees Butterfill (2012) as belonging to the bottom-up strand, according to which children, for instance, display joint activities that are different from and weaker than joint actions.

The idea of complementarity first surfaces in Butterfill and Sebanz (2011: 145), who contend that

> the objective of the [bottom-up] strategy [is not] to eliminate appeal to sophisticated notions of shared intentions altogether. The aim is rather to explore whether the most sophisticated cases could emerge from more basic forms of join action and, if so, how they might emerge.

This talk of emergence can be interpreted in a pluralist as well as a monist light though. Phenomena of sophisticated joint actions may *dynamically emerge* from that of simpler ones, for instance if motor processes trigger the appearance of certain mental attitudes. For instance, a successful mechanism of automatic synchronization may trigger team reasoning in participants. Such complementarity is compatible with a complete conceptual heterogeneity among the top-down and bottom-up approaches. However, emergence may also be understood in the sense that complex joint actions are *constituted* by a number of functional elements, some of which are realized by low-level mechanisms. In this case, one would expect that the elements identified by both approaches be somewhat homogeneous, as some would have to be able to instantiate others.

By and large, the joint action literature seems to have adopted the latter (monist) view. Maybe this is for diplomatic reasons. Complementary approaches can peacefully coexist; one's relevance does not preclude another's, at least not if each of us is making a step in our collective march towards a common goal. The possibility of pluralism, while mentioned, is never quite explicitly embraced. For instance, Heinonen (2016) concludes on the claim that mechanisms and functions are both needed, and that as a consequence philosophers should become more prone to

[10] Bratman's definition only requires action according to meshing subplans, regardless of whether they have been determined prior to the interaction. By contrast, team reasoning consists In doing's one part of what has been determined (in one way or another) as the best strategy from the team's perspective (Bacharach 1999, 2006).

revising their accounts when new empirical data comes to light. Pluralism is temporary, integration is the horizon.[11] However, the inconsistency among extant scale or cognitive minimalist accounts makes it difficult to imagine such integration unless some of them are abandoned (but which ones then? Nowhere is this possibility explored).

One last possible defense of complementarity consists in claiming that different accounts focus on different aspects of joint action. For instance, for Butterfill (forthcoming)

> joint action raises a tangle of scientific and philosophical questions. Psychologically and neuroscientifically we want to know which mechanisms make it possible… Developmentally we want to know when joint action emerges, what it presupposes and whether it might somehow facilitate socio-cognitive, pragmatic or symbolic development… Phenomenologically we want to characterise what (if anything) is special about experiences of action and agency when collective agency is involved… Metaphysically we want to know what kinds of entities and structures are implied by the existence of joint action… And normatively we want to know what kinds of commitments (if any) are entailed by joint action and how these commitments arise.

So maybe the various accounts of joint action could be seen as participating in a division of labour by highlighting different aspects of joint action. However, the existence of multiple intuitive questions does not prove that they are equally relevant or that answering one cannot provide answers for others. More importantly, the plurality of questions regarding joint actions does not entail a plurality of acceptable accounts. Indeed, Butterfill strives to identify the features that characterize joint action and would provide answers to most of these questions at once; hence his own 'minimalist approach', which betrays a monist aim.

Overall, minimalism appears to aim for a monist account. But its accomplishments smack of pluralism. In what follows, I put some further stress on both ideas of minimalism and monism, by providing maximal cases that escalate the variety of joint action phenomena.

3.4 'Maximal' Cases

I now turn to three examples of joint action that are, or so I later argue, as fundamental as the 'minimal' cases mentioned so far. This section introduces and motivates the cases, while the next one defends their relevance to the pluralism and minimalism discussion and assesses their consequences for the nature of joint action in general.

Demonstrations In politically loaded contexts at least, demonstrations are commonplace – hardly a day without one occurring somewhere in the world. Yet, as

[11] Also note that this convergence presupposes at least some shared background assumptions regarding which empirical evidence is relevant to which account or to which research question (see below).

compared to most joint actions, they are of a unusual kind. They are mass actions, involving a multitude of agents, often from diverse backgrounds. They hardly seem reliable, for failed demonstrations are legion as well. They feature particularly weak epistemic links between agents, most of which necessarily know neither each other's identities nor each other's motivations for participating.

Demonstrations do not entirely defy characterisation though. They have received several tentative definitions (Kutz 2000; Paternotte 2014a). They may share characteristics with more classical forms of joint action, such as their dependence on a form of common knowledge, even if a weakened one (Paternotte 2015). However, they help point at aspects of joint action that are typically underestimated, or even neglected. In some demonstrations, the collective goal is particularly *difficult* to reach, due to the general uncertainty regarding the possible participants' identities and intentions – it is ripe with *epistemic obstacles*. As a consequence, it must be reached by means that are not at work in other joint actions. That enough participants gather doesn't stem from previous joint commitments or plural subjects, from meshing subplans or mutual help – which only happen after enough people have gathered. Similarly, synchronisation or entrainment mechanisms cannot be at work before the interaction starts. So if nothing else, demonstrations stretch the variety of joint actions still a bit further.

Sometimes, the links between the participants of a demonstration are tighter in some regards– they have a better grasp of their general motivation, their expected number – but looser in others. Demonstrations may follow from earlier gatherings during which the general idea of a possible common goal emerges. Consider for instance the Nuit Debout movement born in France on March 2016 after a first demonstration against a government bill. The success of this first event led the participants to consider the possibility of a general, longer movement – the idea that there was a more impactful and lasting action within reach. Although still alive on the Internet, the movement lost in momentum and somewhat fizzled out. This, however, constitutes a case in which some conditions for successful joint actions, including but not limited to demonstrations – emerged from a prior joint event. Such cases of goal emerging from a prior successful demonstration are intriguing in their own and should shed light on the mechanisms that underlie mass joint actions.

Deliberation Individuals routinely have to take decisions together or make collective choices. From jurys choosing the best candidate for a job, to company boards determining their next financial or commercial strategy, to scientists writing common reports to assess a research program, deliberation-based consensus is sought and valued in a variety of contexts. During deliberation, a number of individuals strive to issue a collective opinion or decision by discussing and exchanging arguments – to decide something *together*.

Deliberation is seldom mentioned in the joint action literature. Hakli et al. (2010) discuss it at length, although not as a joint action per se but as a preliminary step for team reasoning. Team reasoning, in the form of the we-mode, is integral to Tuomela's (2007) characterisation of joint action. But it operates only if a collective or team

preference exists (as well as common knowledge about everyone's tendency to team reason). Deliberation is one way by which collectives may form such a preference; through discussion, team members mutually influence one another's preferences and beliefs, which may result in a convergence and so in common collective attitudes that would enable a further joint action.

However, there is no reason not to consider deliberation itself as a joint action. Even if the establishment of collective preferences and beliefs may be a preliminary step before the adoption of a collective goal and its accomplishment through a joint action, the very collective effort of adopting common mental attitudes may be seen as an intentional collective behaviour. Collective adoption of a goal can be a goal in itself.

Deliberation may be seen as a mere collective action rather than a joint one. Just as individuals who walk among a crowd mutually adjust their individual paths depending on the others' movements, participants in a deliberation seem to mutually adjust their beliefs and preferences – they are engaged in mutual epistemic and motivational adjustment. This is not a fair description though. During a collective deliberation aimed at a collective decision, individuals do not just happen to navigate while submitted to the influence of others; they have a goal that is collective in nature – they are acting so that the group may reach a consensus. For sure, there are deliberation that have no aim – for which dissensus may be as acceptable as consensus as a final outcome. I only claim that there are stronger forms of collective deliberation, which constitute joint actions.

Deliberation puts the traditional account of joint action under pressure for at least three reasons. First, its goal is not clearly defined in terms of component actions. Typical accounts contain a clause of the type: agents act jointly only if each agent does her part of the action, etc. Here, parts are not clearly established, as there are many ways to participate in a deliberation. For instance, staying silent and not actively contributing is still compatible with being part of the final consensus. Second, although deliberations are extended in time, the usual dynamical conditions for joint action do not apply. For instance, Bratman's conditions of mutual help between agents are not met, as participants in a deliberation are not trying to help others to defend their own opinion; worse, they typically have partly conflicting interests and may oppose one another – this is precisely what is expected of a critical discussion. Consensus may stem from compromises, which are usually not considered in joint action. In other words, deliberation involves specific *strategic obstacles*. Third and finally, deliberation clashes even with the bottom-up, mechanistic approach to joint action, because we have good reasons to think that it amplifies the negative effect of individual biases; our cognitive abilities seem to increase the odds of deliberation failure (Solomon 2006; Sunstein 2006).

Collective Free Improvisation Since Bratman's early, seminal example of two agents singing a duet together, collective music performance has always been present in the joint action literature. It has become increasingly explored recently (Keller 2008; Loehr et al. 2013), partly because of minimalist motivations (Michael 2017). Musical performances are interesting partly because their rich dynamic structure

makes coordination issues and results more salient. This is even more so in the case of improvisation, in which the absence of a leader and of shared musical parts still complicates the interaction. But nothing stresses coordination like collective free improvisation (Bailey 1993), which consists in a musical performance that is accomplished without any underlying background musical work by performers who may have never player together before and favour different styles. Typical joint improvisation already challenges coordination because it offers an important number of ways to coordinate. Still, it allows for a number of coordinating cues:

> Sharing a referent ('calling a tune') fulfils an essential function for musicians who improvise together: it facilitates their coordination by establishing explicit common knowledge […] From the moment a band chooses for example *My Funny Valentine* as a basis for improvisation, they know (and know that each of them knows) that the style of music is more likely to be jazz than twelve-tone atonality; the musicians will have to follow a given chord progression and go into cycles around it; the piece is usually played in a moody and relaxed atmosphere, and so on. (Canonne and Aucouturier 2016: 545)

However, collective free improvisation offers even more of a coordination challenge, because, by suppressing such a referent as well as cues from past interactions, it "provides a nearly infinite number of "musical strategies" to each player coupled with the difficulty of precisely defining what it is to coordinate in such a situation." (Canonne 2013: 43) The last part of this quote highlights that collective free improvisation does not just make coordination difficult by multiplying the coordination options, but also because *standards for a successful coordination may not be shared by the performers*. It is a fundamentally dynamical process during which emerges not a common goal but the criteria for a possible common goal; whether it has been reached may only be assessed during and after the performance, usually by the performers themselves. In other words, collective free improvisation displays particularly stringent *coordinative obstacles*. Classical characterisations of joint action may provide some leeway regarding the sets of individual actions that would constitute the accomplishment of a goal; but they always presuppose the existence and collective recognition of such a goal. By contrast, in collective free improvisation, the goal itself is not identified beforehand – not any more than the criteria for determining such a goal. This at least suggests that human beings are able to act jointly on the basis of mechanisms that have more coordinating power than is usually thought.

3.5 Contextual Minimalism

What do the cases of demonstration, deliberation and free collective improvisation (FCI) bring to the characterisation of joint action? I have argued that they are bona fide cases of joint action; but note that they are not minimal cases. Demonstration and deliberation are not scale minimal because they involve a high number of

agents; FCI is not because it involves a high number of actions, most of which not predetermined.

Although only a definition of these forms of joint action would show, we also have no reason to assume that our three examples are conceptually minimal, that is, that they may be characterised by one kind of concept. One reason has to do with the role of the concept of a collective goal, which is, either explicitly or implicitly, presupposed in most extant accounts. In its weaker, implicit form, it only features as a list of predefined sets of individual actions that would constitute the joint action. When made explicit, it is motivated by the analogy between individual and joint intention; just as individual intention is goal-directed, so should joint intention. However, our three examples stretch the notion of a goal in joint action. In demonstrations, collective goals may emerge from heterogenous motivations. In deliberation, collective goals have vague contents. In FCI, collective goals do not enjoy pre-established conditions for their accomplishment.

Finally, our three cases do not appear to be cognitively minimal. Success in such complex joint actions would be out of reach for cognitively limited agents. Moreover, their peculiarity makes it unlikely that whatever cognitive mechanisms allow agents to successfully demonstrate, deliberate or collectively improvise would constitute building blocks of joint action in general or would be at play in simpler, more traditional forms of joint action.

Demonstrations, deliberation and free collective improvisation can thus be labelled ‘maximal’ cases of joint action. But there are only maximal with respect to the senses of minimalism considered so far and favoured in the literature, and may be labelled minimal in another sense. Note that in the intuitive cases of joint action on which classical accounts are based, cooperation is *easy*. It occurs in conditions that favour its success, or at least make it unsurprising. As we remarked in the outset of the paper, that cooperation is both easy to accomplish and difficult to pin down is one source of motivation for its study. But this does not mean that we should only or mostly be interested in which the success of joint action is easy to understand. If anything, we should focus on cases in which success is frequent enough.

Our three examples are particular because they constitute cases of frequently successful joint action in particularly difficult conditions – conditions in which the factors that favour cooperation have all but vanished. Some of them are one-shot joint actions, in which the incentives to cooperate born from duration and/or repetition (such as promises, commitments, reputation concerns, etc.) are absent. None of them involves a fully formed shared goal – it is either vaguely defined or has no clear conditions for success. All of them feature obstacles to the formation of expectations about other participants’ cooperative behaviour, either because they are anonymous (demonstrations), have conflicting interests (deliberation) or because they share different background values and beliefs (free collective improvisation).

Still, our test cases may still be labelled as minimal, although in an unusual sense – the minimality of the set of their favouring factors. More precisely, they may be described as joint actions that are *contextually minimal* – they occur in contexts

from which the most factors that intuitively favour cooperative success are absent.[12] This does not prevent them from being maximal in the more traditional senses reviewed so far. Contextually minimal cases of joint action are a strange ilk. They are conceptually difficult to characterise, because they involve concepts that are possibly related to but more complex or more exotic than traditional ones. For instance, in a characterisation of minimal cooperation that includes demonstrations, Paternotte (2014a) appeals to indirect common knowledge – which plays the same functional role as classical common knowledge but appears in different ways. But the very interest of contextually minimal cases lies in what distinguishes them from other cases. Although cognitively non-minimal, the fact that they arise in specific contexts mean that the cognitive mechanisms at work have to be different. Synchronisation, trust or expectations about others' actions and opinions have to be generated in particular, unique ways.

As a consequence, *contextual minimalism* – the approach that aims to explore contextually minimal joint action – should appeal to some of the cognitive minimalists. After all, they strive to identify specific cognitive mechanisms that participate in cooperation in general, in a variety of contexts. Minimal contexts may be included in this list, even if the cognitive mechanisms involved in such joint actions would arguably not be simple or elementary ones.

Contextual minimalism should also appeal to joint action theorists in general, who share the common interest for joint actions accomplished by human beings, or agents that share enough cognitive properties with them. For instance, joint as some minimalists are inspired by joint action among children, others may draw insights from, say, division of labour among chimpanzees. But although eusocial insects are amazing co-operators, no one would argue that their example should impinge on our understanding of joint action. Of course, there is no clear limit as to what counts or doesn't as a *relevant* case of joint action. My point is only that the focus of the joint action literature has always been human joint action, for which any example of frequently successful cooperation among human-like enough agents may be relevant. Contextual minimalism fares well in that respect, because it is interested in joint actions that are typically human, that is, in which cooperative success arises in conditions that would make it unreachable by cognitively weaker agents. That we can act jointly in minimal contexts – which may be epistemically, motivationally or teleologically disadvantaged – is testimony to our specific powers of coordination and cooperation. A contextually minimal approach supports the view that the key ingredients of human joint action are not those that allow agents similar enough to human beings to cooperate in simple enough situations; rather, they are at play in taxing joint actions, which only human beings are able to perform, and the success of which depends on very specific cognitive abilities.

Contextual minimalism also appears to be more consistent with pluralism about joint action than with monism. It advises us to extend the range of the cooperation-prone contexts; it emphasises the variety of contexts in which joint actions may be

[12] Paternotte (2014a) talks of 'minimal cooperation' in this sense.

accomplished as well as of the factors that favour their success. It suggests to separate and isolate interactive contexts of interest before analysing the specific joint actions for which they allow. It makes no commitment towards a possible unity of joint action types in general. Nonetheless, monism is not excluded altogether. The various mechanisms that underlie joint action in different contexts may well accomplish the same set of functions, a set which would thus characterise or define joint action.[13] Contextual minimalism would then be seen as an effort to investigate specific ways by which this function can be realised, ways which occur in cooperatively deprived contexts.

Let me conclude by dispelling one possible objection to contextual minimalism. It is often thought that the easiness or difficulty of a joint action should not impinge on its definition. The set of intertwined mental states that constitute and define a joint intention is one thing (whatever that is); the ways by which such attitudes may or may not be formed, their relative strength or stability and the possible obstacles to their formation, are distinct things. But nothing I have said opposes such a view. My only claim is that difficult joint actions are the most informative regarding the nature and scope of our fundamental cooperative abilities – they best reveal them. Difficult joint actions are different from easy ones because their accomplishment may necessitate (or so a contextual minimalist would argue) particular cognitive abilities that would otherwise not be at play. The conceptual point is that the fact that problems can be of varying difficulty does not mean that all problems can be solved by the same method, applied with varying rigour. More difficult problems may require methods that differ from those suited to simpler ones.

3.6 Conclusion

The repeated attempts to characterise joint action have displayed a common trend towards minimalism – whether they focus on minimal situations, minimal characterisations, cognitively minimal agents or minimal cognitive mechanisms. This trend is also one that appears to lead to pluralism: the idea that joint action may receive multiple, equally valid characterisations. I have argued that the study of joint action would benefit from even further embracing pluralism by adopting a contextually minimalist approach, which focuses on types of joint action that occur in contexts that are cooperatively disadvantaged – in which most factors that typically favour cooperative behaviour are absent. Demonstrations, deliberation and free collective improvisation are a few examples of such situations containing epistemic, strategic and coordinative obstacles. In particular, they lack the kind of clearly defined goals that typically feature in more classical cases of joint action; by contrast, their goals are vaguely defined or even lack precise definitional criteria.

[13] The defense of this position, which I tend to favour, would go beyond the scope of this paper.

Contextual minimalism is not a quirk. It shares the spirit of several current approaches, in particular of the ones focusing on the basic mechanisms at work in joint action. It calls our attention to contexts of interaction in which the usual explanations for the success of joint actions do not apply, and so in which distinct cooperative abilities are likely to be at work. Our understanding of joint action has moved forward by focusing on elementary cooperative mechanisms that underlie simple interactions and that human beings may share with cognitively simpler agents. It would also move forward by considering contexts of interaction in which joint action often succeeds in spite of difficulties, which indicates the use of distinct cooperative abilities.

References

Bacharach, M. (1999). Interactive team reasoning: A contribution to the theory of cooperation. *Research in Economics, 53*(2), 117–147.

Bacharach, M. (2006). *Beyond individual choice*. Princeton: Princeton University Press.

Bailey, D. (1993). *Improvisation: Its nature and practice in music*. New York: Da Capo Press.

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

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

Bratman, M. (2009). Modest sociality and the distinctiveness of intention. *Philosophical Studies, 144*(1), 149–165.

Butterfill, S. (2012). Joint action and development. *The Philosophical Quarterly, 62*(246), 23–47.

Butterfill, S. (forthcoming). Joint Action: A minimalist approach. In J. Kiverstein (Ed.), *Routledge handbook on the social mind*. Routledge.

Butterfill, S., & Sebanz, N. (2011). Joint action: What is shared? *Review of Philosophy and Psychology, 2*(2), 137–146.

Canonne, C. (2013). Focal points in collective free improvisation. *Perspectives of New Music, 51*(1), 40–55.

Canonne, C., & Aucouturier, J.-J. (2016). Play together, think alike: Shared mental models in expert music improvisers. *Psychology of Music, 44*(3), 544–558.

Gilbert, M. (1989). *On social facts*. Princeton: Princeton University Press.

Gilbert, M. (1990). Walking together: A paradigmatic social phenomenon. *Midwest Studies in Philosophy, 15*, 1–13.

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

Hakli, R., Miller, K., & Tuomela, R. (2010). Two kinds of we-reasoning. *Economics and Philosophy, 26*(3), 291–320.

Heinonen, M. (2016). Minimalism and maximalism in the study of shared intentional action. *Philosophy of the Social Sciences, 46*(2), 168–188.

Keller, P. E. (2008). Joint action in music performance. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting intersubjectivity: A cognitive and social perspective to the study of interactions* (pp. 205–221). Amsterdam: IOS Press.

Knoblich, G., Butterfill, S., & Sebanz, N. (2010). Psychological research on joint action: Theory and data. In B. Ross (Ed.), *Psychology of learning and motivation* (Vol. 51, pp. 121–154). Burlington: Academic.

Kutz, C. (2000). Acting together. *Philosophy and Phenomenological Research, 61*(1), 1–31.

Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., & Knoblich, G. (2013). Monitoring individual and joint action outcomes in duet music performance. *Journal of Cognitive Neuroscience, 25*(7), 1049–1061.

Michael, J. (2017). Music performance as joint action. In M. Lesaffre, P.-J. Maes, & M. Leman (Eds.), *The Routledge handbook of embodied music interaction* (pp. 160–166). Routledge.
Miller, S. (2001). *Social action – a teleological account*. Cambridge, UK: Cambridge University Press.
Pacherie, E. (2013). Intentional joint agency: Shared intention lite. *Synthese, 190*(10), 1817–1839.
Paternotte, C. (2014a). Minimal Cooperation. *Philosophy of the Social Sciences, 44*(1), 45–73.
Paternotte, C. (2014b). Constraints on joint action. In M. Gallotti & J. Michael (Eds.), *Perspectives on social ontology and social cognition* (pp. 103–123). New York: Springer.
Paternotte, C. (2015). The epistemic core of weak joint action. *Philosophical Psychology, 28*(1), 70–93.
Peacocke, C. (2005). Joint Attention: its nature, reflexivity and relation to common knowledge. In N. M. Eilan, C. Hoerl, T. McCormack & J. Roessler (Eds.), *Joint attention: Communication and other minds*. Oxford: Oxford University Press.
Schweikard, D. & Schmid, H.B. (2013). Collective intentionality. *Stanford encyclopedia of philosophy.* https://plato.stanford.edu/entries/collective-intentionality
Searle, J. (1990). Collective intentions and actions. In P. R. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 401–415). Cambridge, MA: MIT Press.
Solomon, M. (2006). *Groupthink* versus *The Wisdom of Crowds*: The social epistemology of deliberation and dissent. *The Southern Journal of Philosophy, 44*, 28–42.
Sunstein, C. R. (2006). Deliberating groups versus prediction markets (or Hayek's challenge to Habermas). *Episteme, 3*, 192–213.
Tollefsen, D. (2005). Let's pretend! Children and joint action. *Philosophy of the Social Sciences, 35*(1), 75–97.
Tuomela, R. (2000). *Cooperation: A philosophical study*. New York: Springer.
Tuomela, R. (2007). *The philosophy of sociality-the shared point of view*. Oxford: Oxford University Press.
Tuomela, R. (2013). *Social ontology: Collective intentionality and group agents*. New York: Oxford University Press.
Tuomela, R., & Miller, K. (1988). We-intentions. *Philosophical Studies, 53*, 367–389.
Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks, 23*(8–9), 998–1003.
Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Development, 77*, 640–663.

Chapter 4
What in the World: Conversation and Things in Context

Shaun Gallagher

Abstract Conversation is clearly a form of social interaction and depends on the same kinds of dynamical processes found in interaction more generally. In conversation meaning emerges at the intersection of a set of semiotic resources that include social, cultural, material structures and their dynamical changes in the environment where action and interaction occur. Meaning, including what we can understand of the other's actions, is accomplished, not just in a linear set of speech acts but, by drawing on the various resources available in the environment and in whole body pragmatics. Alignment is an important concept in this context and I define it in wide terms to include the broad range of embodied, ecological and material processes integrated into such events.

Keywords Social interaction · Conversation dynamics · Meaning · Semiotic resources · Alignment

We sometimes find ourselves unintentionally caught up in events. That is, there are circumstances in which, without forming a prior or distal intention to engage in some activity, we find ourselves pulled into it. It may be, for example, that some artifact or work of art catches our eye, and we stand there lost in contemplation concerning its use, or caught up in aesthetic appreciation that we had not planned for. Most often, however, this kind of thing happens in social or intersubjective contexts, where other people (friends, or enemies, or just casual acquaintances) engage us and get us involved in processes that follow a trajectory we had not anticipated. Getting into a conversation is often like this; finding ourselves engaged in a joint action can also happen in this manner.

S. Gallagher (✉)
Department of Philosophy, University of Memphis, Memphis, TN, USA

Faculty of Law, Humanities and the Arts, University of Wollongong (AU), Wollongong, Australia
e-mail: s.gallagher@memphis.edu

A. Fiebich (ed.), *Minimal Cooperation and Shared Agency*, Studies in the Philosophy of Sociality 11, https://doi.org/10.1007/978-3-030-29783-1_4

In such cases it would not be correct to describe ourselves or our situation as being out of control. In some partial way, we may come under the control of others, or of normative expectations to do what we need to do in such cases; or we may engage in the process in a fully intentional way (even if not on the basis of a prior intention). For example, two friends may be standing around with others at an event; the music starts to play and in a completely spontaneous way they start dancing together. To do so, one has to coordinate one's movements with the other person. This is a coordination that emerges on the basis of some degree of controlled movement on the part of the participants, in connection with the music and perhaps some vague rules about an appropriate dance style. If we follow Elizabeth Pacherie's (2006) distinction between prior or distal (D) intention, present or proximal (P) intention, and motor (M) intention, our spontaneous joint action is not the result of a D-intention. But it certainly does take shape on the basis of an ongoing and developing P-intention (or what Searle (1983) calls 'intention-in-action'), which allows us to cope with our immediate environment (e.g., so that we dance without bumping into others on the dance floor), and an M-intention that gets expressed directly in the movements we perform such that we dance rather than walk or move in a non-dancing way. In this respect, we do not perform unintentional actions as we engage in this spontaneous joint action.

Another good example, and the one I will focus on in this paper, is the relatively frequent event of engaging in spontaneous conversation. Indeed, as in the example of spontaneous dancing, not only can the initiation of a conversation be unplanned, but, in addition, much of the ongoing performance of this joint communicative action is necessarily unplanned. A conversation, clearly, does not involve the delivery of prepared statements, even if in some cases a prepared statement sets the parameters of the conversation. Most of a conversation is necessarily unplanned because you do not know for sure what the other person is going to say in response to your words, or what you will say in response to her words. Certain conditions, of course, allow us to anticipate what the other may say, but not with 100% surety. In examining the case of conversation, we want to ask precisely what conditions and resources and processes are involved such that something like cooperation and shared agency emerge.

4.1 Communicative Actions and Interactions

I begin by taking three pointers from Merleau-Ponty's analysis of language. First, as Merleau-Ponty (1964, 40) puts it, we are born into a 'whirlwind of language'. From the beginning, the child hears speech, and is spoken to. Indeed, the whirlwind may even start prenatally, since newborn infants already have a preference for their mother's voice and the rhythms of her speech (DeCasper and Fifer 1980). Second, as Merleau-Ponty (2012) explains, language accomplishes thought. Consider the common experience of the speaker not knowing what he is going to say until he says it. We do not first think things out, and then simply export our pre-formed thoughts.

Rather, the expression is the thought. Through it we discover what we, and others, mean. The same applies to gesture; gesture, which is itself language, accomplishes thought (Cole et al. 2002). Finally, according to Merleau-Ponty language transcends the body. This does not mean that language can be accomplished without the body, or without movement, or completely independently of physical processes. But language, specifically in its capacity to generate meaning, is not reducible to the physical properties of speech or other forms of expression.

The idea that language transcends the body also reminds us of the hermeneutical principle that the text, or the conversation, like interaction itself (see De Jaegher et al. 2010), always goes beyond the author/agent/participant. As one is involved in conversation, one is involved in a process that is larger than any individual speech act, or collection of speech acts. Language not only accomplishes thought, it accomplishes other things too. Speech acts and gestural acts and communicative actions of all kinds get things done, and often have unintended consequences.

First-order communicative practices, that is, gestures (body language in general) and speech acts, are second-person practices (Gallagher 2008). They involve interaction; they are, in most cases, *for-others*, and in cases where they are not (e.g., in cases where I speak or gesture to myself) they are derivative from the social context. The infant responds to motherese and child-directed speech, where intonation and rhythm emphasize emotional meaning. And then the caregiver responds to the infant's response. Turn taking ('cyclic proto-conversation') emerges very early, by 6 weeks of age, and there is a certain rhythm to it. Engaging in vocalizations often involves whole body movements for the infant, in 'dance-like enthusiasm', and these movements can facilitate the mutual regulation of feelings (Trevarthen et al. 1998, 94). Imitation, attention, emotional expressions, hand gestures, movements of the lips and tongue are all part of this original "tango of communication" (Gallagher in press; 1996).

Conversation, in its complex details, is clearly a form of social interaction and depends on the same kinds of dynamical processes found in interaction more generally. This idea is very much in line with what Charles Goodwin has shown in his empirical studies on the situated pragmatics of conversational interaction. Goodwin (2000), for example, shows that meaning emerges at the intersection of a set of semiotic resources that include social, cultural, material structures and their dynamical changes in the environment where action and interaction occur. The factors involved include vocalization, gesture, postural orientation, and the use of items in the local environment. Meaning, including what we can understand of the other's actions, is accomplished, not just in a linear set of speech acts but, by drawing on the various resources available in the environment and in whole body pragmatics.

In one of Goodwin's examples, he provides a detailed analysis of a dispute between two young girls over a game of hopscotch. There is a complex, dynamical organization of various phenomena that have to be considered to understand the full encounter. 'For example, spoken language builds signs within the stream of speech, gestures use the body in a particular way, while posture and orientation use the body in another, etc.' (2000 p. 1494). Goodwin emphasizes the 'visible, public deployment of multiple semiotic fields that mutually elaborate each other' (Ibid). Factors

in these fields include the temporal flow/rhythm of high versus low, and hard versus soft vocal intonation of the speech, for example – some of which have a deontic rather than descriptive force. But also, the set of instituted norms involved in the game of hopscotch, i.e., the rules of the game, as well as reference to a completed action (throwing a marker on one of the squares). One girl intentionally moves and stands in the way of the other girl, interrupting the game. The bodily orientations of the two girls, which allow for eye contact and joint attention toward the hopscotch pattern on the ground, as well as the ongoing temporal modifications in those postures, create meaning as the encounter unfolds. In the communicative actions, hand gestures integrate with the speech, but also with body positions of both girls.

> Carla [one of the girls] has to use her body in a quite precise way while taking into account the visible body of her co-participant. She is faced with the task of using not only her talk, but also her body, to structure the local environment such that her gestures can themselves count as forms of social action.… Unlike talk, gestures can't be heard. [This means] Carla actively works to position her hand gestures so that they will be perceived by Diana [the other girl]…. Carla's hand is explicitly positioned in Diana's line of sight… thrusting the gesturing hand toward Diana's face twists Carla's body into a configuration in which her hand, arm and the upper part of her torso are actually leaning toward Diana. (2000, p. 1498)

The meaning of the gesture depends to some extent on how close it is to the other girl's face. Placement and proximity have meaning. If it were not a gesture, but a touch, how hard or soft, and where the touch occurred, would also have meaning. The gesture is meant to be attention grabbing, forcing the other to orient to the point being made in the speech, or to a point of joint attention towards something in the environment. Grabbing the other's arm could do the same thing. Importantly, this is not one-sided: the other girl attempts to finish her jump through the hopscotch squares in an attempt to ignore the other girl, and the accusation of cheating.

The interaction, the conversation, is not confined to vocalization and gesture. Reference is made to the physical environment, with pointings and glances to the hopscotch squares under discussion. The fact that joint attention is broken when one girl looks away is also meaningful and shows that the accomplishment of meaning involves two-way interaction which is not under the control of just one individual.

The meaning and social understanding involved in this kind of encounter builds on the complex integration of primary and secondary intersubjective capacities. Primary intersubjectivity consists of innate or early-developing sensory-motor capacities that bring us into dyadic, face-to-face relations with others and allow us to interact with them (Trevarthen 1979). These capacities are manifested at the level of action and perceptual experience involving the other's bodily postures, movements, gestures, facial expressions, gaze direction, vocal intonation etc. We respond with our own bodily movements, gestures, facial expressions, gaze, etc. On this view, in second-person interactions, the 'mind' of the other is given and manifest in the other person's embodied comportment.

Secondary intersubjectivity involves contexts of shared attention – shared situations. In the contexts of secondary intersubjectivity, which developmentally starts with joint attention around 9 months of age, we learn what things mean and what

they are for (Trevarthen and Hubley 1978). Peter Hobson succinctly summarizes this notion of secondary intersubjectivity.

> The defining feature of secondary intersubjectivity is that an object or event can become a focus *between* people. Objects and events can be communicated about.… [T]he infant's interactions with another person begin to have reference to the things that surround them. (Hobson 2002, 62)

These processes of primary and secondary intersubjectivity are situated within pragmatic and social contexts, supplemented with and supporting communicative and narrative processes. In such interactions, agents enter into dynamical relations that constitute a system extending beyond what each individual agent brings to the process (De Jaegher et al. 2010). The physical and social affordances presented by the pragmatic and social contexts are relational and depend on the possibilities opened up by interaction itself. One person's attempt to understand or communicate with the other, whom she knows, and with whom she is interacting, is not reducible to her observation of the other in her situation; rather, she is already part of the other's situation, as the other becomes part of hers. It's a shared situation; a shared context within which they encounter each other. In interaction, contexts are relational.

The communicative situation, then, contains a wealth of resources that an agent can draw on, beyond the vocalized words:

- Bodily movements, postures and proximity
- The gestures and facial expressions of the other person
- The intonation of voice
- The other's attention – the means to grab it for joint attention
- The temporal flow/rhythm of interaction
- Instituted norms
- Social rules, roles, and identities
- Knowledge of completed actions
- Knowledge of person-specific traits, preferences, attitudes, etc.
- The rich material environment

As Goodwin (2000) notes, 'this is by no means a fixed array of fields. Thus, on many occasions, such as phone calls, or when participants are dispersed in a large visually inaccessible environment (e.g., a hunting party, or a workgroup interacting through cyberspace), visual co-orientation may not be present' (p. 1500). Contexts change over time; they may be enriched or impoverished, but they always count towards the production of understanding or misunderstanding (Duranti and Goodwin 1992).

This kind of analysis shows that live communication is not reducible to a simple set of alignment processes, although they include such processes. Intersubjective alignment is a notion defined in different ways by different authors. Some equate it with simple matching, imitation or entrainment and distinguish it from other more intentional coordination processes (e.g., Rothwell et al. 2017); others define it as including more complex forms of coupling (Tollefsen et al. 2013). The research on

conversational analysis shows, however, that there is a broad range of embodied and ecological processes integrated into such events, and that different circumstances (e.g., environment structures) and different intentions elicit different dynamical balances among these processes. Accordingly, all of these semiotic resources, and the linguistic practices that go with them, help to clarify and complicate our social understanding of others.

4.2 How Things Shape Our Conversations

A very old television show called *What in the World,* which aired weekly on CBS in the United States in the 1950s, would start with some weird outer-space type music and a view of an obviously fake image of earth. Then it seemed you were flying through some clouds when a tremendously deep and dramatic voice would pierce through the music and the clouds, announcing, rather than asking: 'WHAT IN THE WORLD'. Once the clouds cleared you were looking as some very odd object sitting on a table, and you were saying to yourself, "What in the world is that?" This fantastically dramatic opening was the beginning of a weekly show on archeology that must have seemed relatively boring to anyone other than an archeologist. The odd object was something that somebody had unearthed at an archeological dig, in some far away place, and which was now housed in the University of Pennsylvania Museum. The show consisted of two or three thoroughly academic archeologists (including Froelich Rainey) who were given the task of identifying the object, saying where it was from, what time period it belonged to, and what it was used for. They were experts, of course, and they almost always figured it out, although sometimes they were stumped. The thing (whatever it happened to be that week) was clearly and explicitly the focal point of the conversation and every sentence was in some way linked to the meaning of this material object. This specific type of extreme thing-tied type of conversation happens here and there and every now and then, but it is not the rule.

I point to this example to make a distinction. In some cases, as in the television show, a thing is the center of the conversation – it is the object of the conversation and the conversation revolves around it in an explicit way. But, as I suggest, this is not the rule – it's not the way that things usually function in conversation. More generally things tend to operate in the background, serving a pragmatic role in assisting (or sometimes interfering in) the conversation. In the television studio production of *What in the World*, the microphones serve this sort of purpose. Depending on the type of microphone in use, it may facilitate moving around and gesturing; or it may keep the speaker in place, behind a podium; or it may require holding, in which case it will limit or change the dynamics of one's gesture.

Martin Heidegger (1962 §15) famously made this distinction: we are not primarily in the world of the *Vorhandenheit* – the present-at-hand – a way of regarding the world theoretically as a collection of things, or regarding a thing as something to be explained, as in the case of the archeological object on the table. We are rather

primarily in the world pragmatically, manipulating things grasped as something *Zuhanden* – ready-to-hand – and ready to use as part of some ongoing project. Our primary way of being-in-the-world is a form of material engagement with things that serve as instruments or pieces of equipment.

Lambros Malafouris (2013) calls our attention to the importance of material things and how our engagements with such things shape the way we think and act. He digs into archeological history to provide examples of how things (such as the Mycenaean Linear B tablets, or stone-age tools) can act as props or instruments for extending knowledge and for enabling action. His focus is on instrumental actions, however, rather than communicative actions or joint actions (see Gallagher and Ransom 2016). Although, in some regard, Malafouris shifts away from a communicative logic to a more general enactive one (2013, 90), he still provides a means to think of communicative practices as material practices where the medium forms part of the message, and the material specifics of the communicative practice have significance. The fact that a nationalist symbol is made of stone or iron rather than wood has significance; the process of signing one's name rather than simply stamping it or sending an electronic signature adds a certain affective force to the sign (91ff).

> Indeed, when it comes to material signs – and in direct opposition to the Saussurian contention concerning the arbitrariness of signification – to make the letters "black or white, raised or engraved, with pen or chisel" is in most cases the defining feature of the signification process. Moreover, not only the physical form of the letters but also the physical properties of the medium upon which they are inscribed are being dynamically implicated in the semiotic process. (Malafouris 2013, 93)

Beyond inscriptions, however, many other things enter into the semiotic process. Consider, for example, the case where we are both dressed in jeans and T-shirts and we meet at the local pub. What we might talk about seemingly doesn't depend on our mode of dress. If I were to say that my brother is getting married, you might nod and move on to a funny story about a wedding. Yet, if I were dressed in a tuxedo and I said the same thing, it may take on a more complex significance that would motivate a different remark from you – for example, 'So what are you doing here'? Likewise, the effect of my jeans and T-shirt in a court of law if I am the defense attorney would be quite different from the seeming non-effect of dressing the way lawyers typically dress for court, and would be different still from dressing in a tuxedo in such a context. "The distinctive properties of the material world bring about meaning in ways that language cannot, and vice versa" (Malafouris 2013, 95). Of course, the materiality of the message is not divorced from the cultural practices and locations in which it is instantiated – pub versus courtroom; telling stories versus litigating. The connection here between specific institutions and the institution of meaning through the media of material culture and practices is important (Gallagher 2013). Malafouris cites Martin Byers (1992, 415) on this point:

> In action-constitutive terms, all material culture items have action-constitutive force. Just as the judge's 'ermine' is a constituent element of her/his judicial speech action by endowing the wearer with institutional authority which simultaneously transforms the utterance of specific words into a legal declaration, so the (utilitarian) axe is also a constituent element

> of the user's material action, endowing its user with institutional or virtual property, e.g., that of being a peasant, thereby transforming the physical labour into the material action of discharging feudal dues, etc.

In the same way that the marker thrown on the hopscotch square is an act of sense making only because of the rules instituted for the game of hopscotch, and only in this way can some action be the violation of a rule and a determination in the subsequent encounter, the material structures of courtroom, pub, or feudal estate afford, and sometimes constitute the semiotic field and the framework for social interactions and shape the meaning of our conversations in ways which are not specifically in our awareness. Notions of external scaffolding as found in the concept of extended mind discussions (Clark 2008; Sterelny 2010, 2012), and material anchoring as found in discussions of distributed cognition (Hutchins 1995) can be taken in a close to literal way when the stuff of conversation and interaction is material stuff and institutional structure.

4.3 Wide Alignments

I indicated that the notion of intersubjective alignment is defined in different ways by different authors, from simple matching to more complex forms of coupling (Rothwell et al. 2017; Tollefsen et al. 2013). In this section I want to explore the range of elements that contribute to alignment, and to define alignment in wide, inclusive terms that include the broad range of embodied, ecological and material processes integrated into such events. I want to include intentions as well as circumstantial contextual features that define the dynamical balances among the elements of conversation. What kinds of principles can be drawn from the empirical work that has been done on alignment?

On one influential reading, intersubjective alignment involves 'a loosely interconnected set of cognitive processes' organized to facilitate social interactions (Tollefsen et al. 2013, 49). The interconnections are dynamic and adaptive to environmental constraints as well as to higher-order cognitive processes such as individual and shared or collective intentions, and at the same time they generate ongoing, shared cognitive states. Should we think of this as involving bottom-up processes, as Tollefsen, Dale and Paxton suggest? Or should we consider a dynamical mix of bottom-up and top-down processes, the latter sometimes leading to an inhibition of alignment since alignment is not always something good. This terminology suggests a set of vertically organized operations, some higher-order and some lower-order. I propose that it is more productive to think of these dynamical interactions on the model of what I'll call 'Goldstein's gestalt'. Kurt Goldstein, who made a well-known distinction between abstract and concrete cognitive processes, suggests that we should view the relation between such processes not on a hierarchical model, but on the model of a gestalt.

> Although the normal person's behaviour is prevailingly concrete, this concreteness can be considered normal only as long as it is embedded in and codetermined by the abstract attitude. For instance, in the normal person both attitudes are always present in a definite figure-ground relation. (Goldstein and Scheerer 1964, 8)

The dynamical relations themselves shape the *relata*, so that one can think of memory or imagination, not as separate processes from conversational or interactional practices, but as continuous with and intervening in speech, gesture, posture, movement and action. The kind of coupling described in terms of alignment clearly depends on basic motoric processes where we become attuned to the other's verbal and non-verbal expressions. This attunement sometimes involves a motoric/neuronal resonance where individual systems join in the formation of a larger interactional system. A good example of this is the formation of a 'joint body schema' in synchronic joint action. At the level of basic M-intentions, Soliman and Glenberg (2014), studying joint action, and body-schematic (motoric) coordination, have found evidence for a joint body schema understood as an extension of each individual's peripersonal space to include the other person. In their experiment, two participants coordinate for 5 min by moving a flexible wire back and forth to cut through candles, requiring close coordination to keep the wire tight. Similar to experimental results that show how tool use extends the body schema or expands peripersonal space, the coordinated action produces modulations that show up in both neuronal and behavioral measures.

This data could be interpreted in two different ways. Either, (1) what changes are simply processes in each individual – my body schema/peripersonal space expands; and your body schema does too. These are subpersonal changes that may generate an *individual* sense of joint agency – a feeling of being in sync with the other. Or, (2), on a more enactivist reading, the two bodies form a larger action system, so that the joint body schema belongs only to this larger system (two parts constituting a larger whole). This would be a form of physically attuned embodied alignment and may very well be limited to synchronous *concrete* interactions. Notably, not all joint actions involve this type of alignment, and this might motivate an objection that the phenomenon is limited to bottom-up processes. Importantly, however, Soliman and Glenberg (2014) show that these processes are permeated by more abstract (or so-called higher-order), and specifically cultural factors. They reject a dualistic account of culture which assumes that 'culture is a package of propositional constructs like values, beliefs, and world views' (2014, 207), and that gives primacy to such abstract propositional values or belief structures for defining culture, and relegates more embodied practices as secondary. In contrast, Soliman and Goldberg (2014, 209) argue 'culture is not a psychological structure that exists apart from more mundane learning and behavior'. As an enactivist might say, culture doesn't exist apart from everyday practices. Rather, 'the psychological underpinning of culture consists of an individual's sensorimotor tuning arising from and guiding social interactions' (Ibid). Not only are body schematic processes tuned by material practices, as in the case of tool use, they are also modulated by cultural differences. Citing the work of

Markus and Kitayama (1991) on individualistic *versus* interdependent self-conceptions in different cultures, they show that the joint body-schema effect was most pronounced when subjects were from socially interdependent versus individualist cultures.

> [C]ulture enters the scene not as a self-contained layer on top of behavior, but as the sum of sensorimotor knowledge brought about by a bodily agent interacting in a social and physical context. As such, culture diffuses the web of sensorimotor knowledge, and can only be arbitrarily circumscribed from other knowledge. (Soliman and Goldberg 2014, 209)

In such cases the alignment process is not a top-down phenomenon; rather the analysis shows the gestalt structure of these processes, involving not just brain plasticity, but a metaplasticity where changes in brain, body, and environment modulate each other in a holistic fashion (Malafouris 2013). Joint actions involve material engagement and rich contexts, and these can be explicated in terms of differences in cultural background or *habitus*. Such things in fact characterize many everyday actions and practices, as well as rituals within institutional contexts (Gallagher 2013; Gallagher and Ransom 2016).

4.4 Conclusion

Conversations and joint actions, even when to some degree planned and intentional, involve spontaneous yet organized dynamics that both transcend and depend upon the materiality of embodiment and environment. Some aspects remain unplanned simply because one does not know for sure what the other person is going to say, or what one's own response to that will be. Yet, with all the elements of spontaneity and the unplanned, communicative and joint action processes come into alignment and form a trajectory just because we are embodied and embedded in material contexts that extend into cultural practices, shape affordances, and generate institutional frameworks.

All of these elements are moving parts – bodies that posture and gesture, and align to each other, material things that scaffold and anchor, cultural practices that push and pull in a variety of directions, institutional rules and structures that keep things in line or that bend with spontaneous interactional forces – parts of a metaplastic, dynamical gestalt. These processes often lead to predictable rule-governed consequences; but every now and then we find ourselves saying, to ourselves, or to one another: 'what in the world!'

Acknowledgements Thanks to Lambros Malafouris and Deb Tollefsen for helpful discussions on these topics. Thanks also for the support provided by the Humboldt Foundation's Anneliese Maier Research Award.

References

Byers, M. A. (1992). The action-constitutive theory of monuments: A strong pragmatist version. *Journal for the Theory of Social Behavior, 22*(4), 403–446.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford: Oxford University Press.

Cole, J., Gallagher, S., & McNeill, D. (2002). Gesture following deafferentation: A phenomenologically informed experimental study. *Phenomenology and the Cognitive Sciences, 1*(1), 49–67.

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Does social interaction constitute social cognition? *Trends in Cognitive Sciences, 14*(10), 441–447.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science, 208*(4448), 1174–1176.

Duranti, A., & Goodwin, C. (Eds.). (1992). *Rethinking context: Language as an interactive phenomenon*. Cambridge: Cambridge University Press.

Gallagher, S. (1996). The moral significance of primitive self-consciousness. *Ethics, 107*(1), 129–140.

Gallagher, S. (2008). Intersubjectivity in perception. *Continental Philosophy Review, 41*(2), 163–178.

Gallagher, S. (2013). The socially extended mind. *Cognitive Systems Research, 25–26*, 4–12.

Gallagher, S., & Ransom, T. (2016). Artifacting minds: Material engagement theory and joint action. In C. Tewes (Ed.), *Embodiment in evolution and culture* (pp. 337–351). Berlin: de Gruyter.

Gallagher, S. (in press). *Action and interaction*. Oxford: Oxford University Press.

Goldstein, K., & Scheerer, M. (1964). *Abstract and concrete behavior: An experimental study with special tests*. Evanston: Northwestern University. Reprint of *Psychological Monographs* 53 (2), 1941.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics, 32*, 1489–1522.

Heidegger, M. (1962). *Being and time*. (J. Macquarrie, & E. Robinson, Trans.). New York: Harper & Row.

Hobson, P. (2002). *The cradle of thought*. London: Macmillan.

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

Malafouris, L. (2013). *How things shape the mind*. Cambridge, MA: MIT Press.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224.

Merleau-Ponty, M. (1964). *The primacy of perception* (J. Edie, Trans.). Evanston: Northwestern University Press.

Merleau-Ponty, M. (2012). *Phenomenology of perception*. (D. A. Landes, Trans.). London: Routledge.

Pacherie, E. (2006). Towards a dynamic theory of intentions. In S. Pockett, W. P. Banks, & S. Gallagher (Eds.), *Does consciousness cause behavior? An investigation of the nature of volition* (pp. 145–168). Cambridge, MA: MIT Press.

Rothwell, C. D., Shalin, V. L., &Romigh, G. D. (2017). Quantitative models of human-human conversational grounding processes. [Private Circulation via Research Gate].

Searle, J. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.

Soliman, T., & Glenberg, A. M. (2014). The embodiment of culture. In *The Routledge handbook of embodied cognition*. Taylor and Francis.

Sterelny, K. (2010). Minds: Extended or scaffolded. *Phenomenology and the Cognitive Sciences, 9*(4), 465–481.

Sterelny, K. (2012). *The evolved apprentice*. Cambridge, MA: MIT Press.

Tollefsen, D. P., Dale, R., & Paxton, A. (2013). Alignment, transactive memory, and collective cognitive systems. *Review of Philosophy and Psychology, 4*(1), 49–64.

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech: The beginning of interpersonal communication*. Cambridge: Cambridge University Press.

Trevarthen, C., & Hubley, P. (1978). Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year. In A. Lock (Ed.), *Action, gesture and symbol: The emergence of language* (pp. 183–229). London: Academic.

Trevarthen, C., Aitken, K., Papoudi, D., & Robarts, J. (1998). *Children with Autism: Diagnosis and intervention to meet their needs* (2nd ed.). London: Jessica Kingsley Publishers.

Chapter 5
Shared Intentionality and the Cooperative Evolutionary Hypothesis

Glenda Satne and Alessandro Salice

Abstract One important application of theories of collective intentionality concerns the evolution of social understanding and even of human thinking (Tomasello M, A natural history of human thinking, Harvard University Press, Cambridge, MA, 2014). A promising idea behind this approach is the Cooperative Evolutionary Hypothesis (CEH), namely, the idea that humans' capacity for social cooperation is at the heart of their ability to understand others' mental states and behavior, leading to an explanation of how humans came to share thoughts and language. However, some of the most popular defenses of CEH face important problems. In this paper, we take Tomasello's account (J Soc Ontol 2(1):117–123, 2016); A natural history of human thinking. Harvard University Press, Cambridge, MA, 2014; Origins of human communication. MIT Press, Cambridge, MA, 2008) as a leading example of the CEH which faces such insurmountable problems. In particular, we argue that Tomasello's analysis of cooperation and spontaneous help is problematic. We locate a source of such issues in the assumption that the right account of joint action and simple forms of shared intentionality is that which is offered by Bratman's theory of shared intentions. The second part of the article proposes and defends an alternative framework for understanding shared intentionality that can help substantiate CEH.

Keywords Cooperation · Collective intentionality · Evolution

G. Satne (✉)
Department of Philosophy, University of Wollongong, Wollongong, Australia
e-mail: gsatne@uow.edu.au

A. Salice
Department of Philosophy, University College, Cork, Cork, Ireland
e-mail: alessandro.salice@ucc.ie

A. Fiebich (ed.), *Minimal Cooperation and Shared Agency*, Studies in the Philosophy of Sociality 11, https://doi.org/10.1007/978-3-030-29783-1_5

5.1 Introduction

One important application of theories of collective intentionality concerns the explanation of the evolution of social understanding and for some even of human thinking (Tomasello 2014). A promising idea behind this approach is the *Cooperative Evolutionary Hypothesis* (CEH), namely, the idea that humans' capacity for social cooperation is at the heart of their ability to understand others' mental states and behavior, leading to an explanation of how humans came to share thoughts and language.

In this vein, Tomasello (2014, 1999a, b, Tomasello et al. 2005), has argued that a tendency to cooperate, coupled with an ability for socio-cultural learning, leads to the emergence of human specific sophisticated forms of cognition including the ability to entrain representations that are objectively valid. Thus, Tomasello claims that:

> the amazing suite of cognitive skills and products displayed by modern humans is the result of some sort of species-unique mode or modes of cultural transmission. The evidence [for this] […] is overwhelming. (Tomasello 1999a, 4)

Furthermore, following CEH, Tomasello argues that,

> [Thinking] is a solitary activity all right, but on an instrument made by others for that general purpose, after years of playing with and learning from other practitioners […]. Human thinking is individual improvisation enmeshed in a socio-cultural matrix. (Tomasello 2014:1)

In his account of CEH, Tomasello (2014) distinguishes two forms of social intentionality that correspond to a two-step evolutionary story. Both are forms of interaction and collaboration that are said to have paved the way for the emergence of human specific forms of cognition. First there is *joint intentionality*. This form of intentionality characterizes short-lived spatially and temporally located second-personal interactions between particular individuals. These are small-scale collaborations, triadic interactions in which the individuals involved share goals and jointly attend to situations and objects in the environment (Tomasello 2014, 48). According to Tomasello, this form of intentionality does not involve any grasp of objectivity in the sense of universal validity on the part of the interacting agents, rather individuals involved in these kinds of collaborations have a perspectival understanding of the world that encompasses a 'sense' or understanding of other perspectives towards the same objects. Second, there is *collective intentionality.* This kind of intentionality is at play when agents engage in large-scale collaboration. This form of engagement between agents goes beyond the here and now characteristic of joint intentionality. It involves the establishing—or the background presence of an already established—*cultural common ground* (e.g., institutions, conventions, and so on). This kind of intentionality, according to Tomasello, makes an "agent-neutral point of view" available to the agent (Tomasello 2014, 5/122). The point of view of a universal group that, in this framework, equates to the ability to grasp an objectively valid point of view.

Thus, in Tomasello's view the notion of joint intentionality pretends to describe the set of abilities that an organism is able to entertain without yet being capable of manipulating truth-functional representations of others and the world. It comprises mechanisms for social cooperation that enable individual learning and work in conjunction with mechanisms for the social inheritance of culturally evolved devices. When it comes to giving an account of joint intentionality Tomasello refers to the work of Bratman. Joint cooperative activities proper to joint intentionality are defined in terms of the satisfaction of Bratmanian conditions for joint intentional action (Tomasello 2014, 38ff). Accordingly, Tomasello states that three conditions for joint intentional action are sufficient for joint intentionality, i.e.:

If you and I are agents, and J is a goal, then:

(1) I must have the goal of doing J together with you,
(2) You must have the goal of doing J together with me,
and
(3) We must have "mutual knowledge, or common ground, that we both know each other's goals" (Tomasello 2014: 38)

Condition (3) involves meta-representations of other people's mental states, and hence higher order attitudes. It is to be noted that not every interaction of this kind will involve meta-representations, because there may be sufficient information in the situation available for the agents to know that they both want to J. In such cases, the agents know their mutual goals because they stand on 'common ground.' Nevertheless, Tomasello argues that a capacity for recursive mindreading needs to be presupposed on the part of the agents to make sense of the very possibility of them engaging in this kind of interaction. He argues that in conflicting situations, we draw back to recursive reasoning, showing that our capacity of knowing that we both want to J has "an underlying recursive structure" (Tomasello 2014, 38; also 2008). Thus, mechanisms for understanding the perspective of another agent towards the common goal and the situation need to be presupposed. Tomasello argues further that understanding the perspective of a partner in joint action requires simulating her abductive inferences ahead of time in order to anticipate how one might be understood by the other (Tomasello 2014: 94). Thus, agents capable of this sort of interaction have cognitive capacities that underpin recursive inferences, simulations, ongoing self-monitoring and meta-representations of self and other (Tomasello 2014:5, 9, 143). In this view, social cognition is part and parcel of joint intentionality involving both simulation theory (ST) and theory-theory (TT) of mind.

Tomasello's particular attempt to advance the CEH faces several problems (see Satne 2016). In this paper, we focus on one strand of such problems. Specifically, we argue that Tomasello et al.'s (Tomasello 2014; Warneken and Tomasello 2006, 2007) analysis of cooperation and spontaneous help is problematic. The root of the issue lies in assuming that the right account of joint action and simple forms of shared intentionality is given by Bratman's theory of shared intentions. We conclude by proposing an alternative framework for understanding shared intentionality that has the resources for overcoming these difficulties, fits well with empirical data, and can help substantiate CEH.

We proceed as follows. The first section outlines Tomasello's (Warneken and Tomasello 2006, 2007; Tomasello 2016) account of helping and cooperative behavior and its Bratmanian assumptions. The second section presents and discusses important problems of the view. The third puts forward of an alternative account, "a (minimal) collective view", that fits Warneken and Tomasello's empirical data while also dissolving the conundrums that affect their view. Finally, section four responds to the challenges and assesses the advantages and potential downsides of both views.

5.2 Tomasello's Account of Helping and Cooperative Behavior and Its Bratmanian Assumptions

Tomasello (2014, 2016) argues that children's early cognitive development provides important evidence for drawing hypotheses about the emergence and evolutionary trajectory of human cognitive capacities.[1] Consequently, studies focusing on early collaborative activities in children can be thought to provide important hints into the evolution of human cognitive capacities and potentially contribute to an account which would fit the CEH. But is Tomasello's account of cooperation a plausible way of substantiating the CEH? We think not. To see why we will focus on a series of experiments (and corresponding conclusions) of Tomasello's research on early helping and cooperative behavior in young children.

Warneken and Tomasello (2006, 2007) have studied joint action and helping behavior in young children. Such studies investigate the child's early understanding of shared intentions by focusing on two different kinds of behavior prominent in children between 14 and 24 months: (1) the child's tendency to help others achieve their goals, which is manifested in their engagement in further pursuing the incomplete actions of adults; (2) The child's ability to engage in shared cooperative activities, where the roles of each agent's actions are complementary and where all the interactors (children and adults) are pursuing the same goal together (Warneken and Tomasello 2006, 2007). In this view, only the latter is an example of shared intentionality and coincides in development with what we defined in the previous section as 'joint intentionality'.

According to Warneken and Tomasello, the first kind of cooperative behavior, i.e. helping behavior, is to be found in children starting at around 14 months of age. The experiments that test this kind of behavior are as follows: several tasks were performed in front of children where the experimenters encountered some kind of

[1] Tomasello endorses a more controversial thesis, the idea that 'ontogeny recapitulates phylogeny,' illustrated by his endorsement of the two-step intentionality theory for both ontogeny and phylogeny (see the final section of this paper for an illustrative quote). While we are not alone in thinking that this thesis is too strong (see Gould 1977 and Satne 2016), the claim that the developmental path of a cognitive capacity provides some evidence and constraints on its evolutionary trajectory is quite uncontroversial.

hindrance—his/her marker fell while drawing, a clothespin was dropped while hanging towels, a spoon accidentally fell into a box on top of which the experimenter's teacup was placed. In each case, the experimenter expressed annoyance when this happened ("Oops! I dropped my marker", etc.). This was contrasted with control conditions where the experimenter did not seem to care about what happened, or where she/he intentionally caused the hindrance (e.g., intentionally throwing the spoon into the box after stirring the tea). While in the former cases the children spontaneously engaged in the activity by helping the experimenter (by picking up the clothespin or marker, etc.), in the latter they did not. According to the authors, the outcomes of the experiments showed that young children have an altruistic tendency to act upon others' incomplete or impeded goals, and a correlated rudimentary capacity to understand someone else's intentional actions. Importantly, in their view, this is not yet a case of shared intentionality, because even if children are acting to achieve another person's goal, they are not coordinating their activities with the adult.

The second kind of behavior they tested is the one that, according to these researchers, can properly be called "joint intentional activity." As we will see below, such experiments are meant to satisfy Bratman's conditions for shared intentional activity.

In similar experiments, Warneken and Tomasello (2006) analyzed the behavior of children between 18 and 24 months. There, two agents (the child and the experimenter) must perform complementary roles in order to achieve a joint goal, e.g., pulling opposite ends of a tube to retrieve stickers that are hidden inside or placing a ball into a tube and catching it from the other side. It was shown that children as young as 18 months can successfully deal with these sorts of joint problem-solving tasks. When comparing both age groups, it was observed that this sort of activity is consistently and spontaneously carried out by 18- to 24-month-old children and becomes more skillfully and expertly performed over time. Moreover, in carrying out these experiments, the researchers found that children over 18 months of age consistently protest (verbally or non-verbally) if the adult disengages before the shared activity is completed and, significantly, that such protesting still arises even if the activity is successfully completed and the goal attained. In addition, they also observed a tendency in the children to repeat the action even when it was completed successfully. In their view, this showed that, in so doing the children were not pursuing an individual goal (e.g. get the ball) but rather that they were interested in the shared activity as such.

In summarizing their results, Warneken and Tomasello characterize spontaneous helping behavior as an example of individual intentionality: contributing to another person's goal manifests an understanding of the other person's goal directed behavior, but it is not yet a case of sharing a goal with another. This comes as no surprise since they explicitly state that shared intentional activity is to be understood in terms of Bratman's analysis of shared intentionality (see Warneken et al. 2006: 291; see also Tomasello 2014:38,40).

According to Bratman (1992) agents that engage in joint action have a number of individual and interlocked intentions that steer, guide, and monitor their joint activity. Given Jones and Jane and an action *J*—say, building a house—the conditions are that:

(1) The agents pursue *interdependent* goals—that is, Jones intends that Jane and Jones J together and Jane intends that Jones and Jane *J* together.
(2) The agents pursue the goal in an *intentional and non-coercive* way, that is, Jones intends that Jane and Jones *J* together in accordance with and partly because of Jane's intention as in (1) and Jane intends that Jones and Jane *J* together in accordance with and partly because of Jones's intention as in (1).
(3) The agents *intend that their strategies towards the goal mesh together*: Jones intends that Jane and Jones *J* together in accordance with and partly because of meshing sub-plans of (1) and Jane intends that Jones and Jane *J* together in accordance with and partly because of meshing sub-plans of (1).
(4) All the conditions are *common knowledge* for Jones and Jane.
(5) Among the agents there is public *mutual responsiveness* that tracks the goal.
(6) There is a *commitment to mutual help* and support among the agents.

In Bratman's framework, a goal J can be said to be shared *distributively*. Distributive goals are defined by Butterfill (2012) in the following way:

> An outcome O is a distributive goal of two or more agents' actions if, first, O is one to which each agent's actions are individually directed and, second, each agent's actions are related to the outcome in a way that it is possible for all of them together to succeed in bringing about O (Butterfill 2012: 849).

Since the conditions spelled out by Bratman's account of joint cooperative activity are not fulfilled in the first group of experiments in which there are no complementary roles for the action to be achieved, Warneken and Tomasello treat helping behavior in young children as a precursor to the kind of shared activity that is exhibited later, when children engage with adults in coordinating activities, what they take to be 'joint intentionality' proper. Accordingly, they characterize spontaneous helping behavior as an example of individual intentionality: contributing to another person's goal manifests an understanding of the other person's goal-directed behavior, but it is not yet a case of sharing a goal with another, as the conditions spelled out by Bratman are not fulfilled. The authors conclude that helping behavior in young children is a precursor to the kind of shared activity that is exhibited when they later engage with adults in coordinating activities, such as in the experiments above. Nevertheless, as we argue below, the specific features of the analysis they provide of shared intentional activity and helping behavior cast doubt on whether their conclusion really follows from the studies. Several features of children's behavior are not really explained by the Bratmanian model they use. Moreover, using such a model to explain children's engagement in shared activity leads to what could be taken as important shortcomings in their analysis.

5.3 Challenges for Tomasello's Account of Helping Behaviour and Cooperation in Young Children

5.3.1 Continuity and Overlap

While children seem to move naturally from helping others complete their actions (starting at 14 months) to engaging in joint action at 18 months, in Tomasello framework there is no motivation to see these two kinds of behavior as interconnected in any important way. Warneken and Tomasello themselves conclude that:

> Intraindividual comparisons of infants' performance on helping and cooperation [tasks] revealed no straightforward associations between the two suggesting that these activities differ in important ways. (Warneken & Tomasello 2007: 291, see also 291–292)

But this is because the account frames these two kinds of behavior as *a fortiori* independent. On the one hand, tests aimed at studying 14-month-old children helping adults complete their actions are conducted in scenarios where there aren't any individual motivations for the children to pursue the specific actions the adult is engaging with (this is something that the experiments explicitly control for), something that is a precondition for engaging in Bratmanian shared cooperative activities, as stated in condition (1) above—i.e., both individuals need to have the goal of J-ing together. Conversely, children's engagement in joint cooperative activities of a Bratmanian sort is completely independent of their tendency to help adults. This is because one can engage in Bratmanian cooperative activities without being motivated to help the other interactors. It is true that Bratman's account includes a commitment to mutual help and support as one of its conditions (condition (6) above). However, as Bratman does not really dwell on this condition extensively, there are different ways to understand it. One way to understand it relates to the kind of intentions appealed to by Bratman's theory: given that agents act upon individual intentions (and are thereby committed to individual, although interdependent, goals), an agent is under pressure to help other interactants to preserve the achievement of the goal (which is endangered by the problems encountered by the other). However, the helping behavior displayed by young children does not seem to be of this kind because children do not seem able to share intentions the way is described by Bratman's, and again this is explicitly controlled for in the case of the experiments analyzing helping behavior in 14–18 month-old children. One could think of this commitment to help and support independently of shared intentions – perhaps justified by general moral principles. Here, again, there are reasons to be skeptical about children in pre-school age having the resources to understand and act upon such a commitment.[2] Thus, even if it is true that as a consequence of engaging in the sort of cooperative activity that Bratman describes individuals *become* committed to offering support to each other—this is condition (6) above—this is so *insofar and*

[2] For an extensive review of empirical literature supporting the claim that young children do not understand such kind of commitments (see Michael et al. 2015).

only insofar as they are pursuing the distributively joint goal at issue. This is different from being naturally and generally motivated to help one another, i.e. a kind of helping behavior that is mainly independent of having the same goals as another.As a consequence, in Warneken and Tomasello's framework helping behavior and shared cooperative activities come out as concomitant types of behavior and not internally articulated developments. For this reason, helping behavior in children requires a completely different explanation on this model. Warneken and Tomasello (2006, 2007) attempt to offer such explanation by postulating the existence of a psychologically innate altruistic tendency to care for others and provide support. But even if humans are born with an innate tendency to help others, postulating such a natural altruistic tendency does not explain it. The problem is that within this framework, helping and cooperative behavior cannot explain one another: an innate tendency to help and support does not explain children's engagement in cooperative activities à la Bratman and engaging in Bratmanian cooperative activities does not explain or predict altruistic tendencies. For this reason, CEH becomes a mere postulate unable to account for the early development of shared intentionality in young children.

5.3.2 Explaining the Evidence

Warneken and Tomasello consider as important evidence of the child's ability to share goals in a Bratmanian sense the fact that she wants to repeat the action and protest if the adult disengages both before the action is completed or after it has been successfully completed. Crucially, according to the authors, this behavior suggests that children are not acting individually in such cases, but that they understand their partner's key contribution in what they are pursuing. Yet, granted that the fact that they want to repeat the cooperative sequence even when completed successfully could be considered as evidence of the special engagement and interest children have in these activities, the evidence is not very congenial with Bratman's conception of joint action and is not predicted by his theory. According to a Bratmanian understanding of shared cooperative activity, there would be no motivation to continue trying to reengage a partner when an action has been successfully completed. The key driving force of each individual's behavior is the goal they are pursuing—i.e., acting to achieve X (say to retrieve the ball from the tube) by means of the other's intention to do the same. But when the ball is retrieved, the goal is successfully reached and, arguably, there is no reason to continue to do it. Following a Bratmanian understanding of shared intentionality one should predict that children would protest only until their individual goal is achieved and not after successfully achieving it. In this framework, once the goal is attained there is no inherent motivation to protest or to attempt to reengage the partner.

5.3.3 Cognitive Demandingness

The third problem also relates to Bratman's conception of shared intentional action. The problem is that if we use, as Warneken and Tomasello do, Bratman's model for understanding shared intentionality in infants—what Tomasello (2014) calls 'joint intentionality,' which is the first possible form of shared intentionality—we need to attribute to children capable of participating in joint action an understanding of the concept of intention as applied to oneself and others and common knowledge of one's own intentions as well as those of the others. As we explained in the previous sections this involves commitment to the agent's being able to entertain higher-order nested propositional attitudes. Arguably, this is not a reasonable demand on the abilities required for young children to engage in joint action, not only because they engage in this sort of activity quite early, but crucially because engaging in some shared activities and joint actions may be essential for acquiring such a sophisticated understanding of the notion of intention and of other persons as agents (see Butterfill 2012, Satne forthcoming). For this reason, many have argued (Tollefsen 2005; Brownell et al. 2006; Michael et al. 2014; Zahavi and Satne 2015) that Bratman's conditions for shared intentional activity are too cognitively demanding to be attributed to young children. The claim coming from many, and despite the differences in their views, is that we need a simplified account of joint action if we want to make sense of young children's ability to engage in shared activities with others.

To illustrate this concern, consider in more detail condition (4) in Bratman's account, which corresponds to Tomasello's condition (3) above, the common knowledge condition. As explained above, Tomasello argues that common knowledge can be had in virtue of the fact that both the agents stand on common ground, but this is in turn grounded on an inferential recursive capacity of mindreading (see: Tomasello 2014, 38). Both in Tomasello's and Bratman's views, this means that each agent engaged in shared cooperative activity is required not only to believe that the partner agent has certain intentions, but also to believe that each agent believes that each agent has certain intentions. If construed along these lines, common knowledge relies on the agent's capacity to attribute beliefs to other individuals about one's beliefs (cf. Tollefsen 2005). Attributing beliefs to other individuals requires a theory of mind (ToM), and traditional research in social cognition suggests that children younger than four only have limited ToM-abilities.[3] It should be said, however, that this conclusion is currently a matter of heated debate. For example, other ToM tests (so called "indirect" tests, see: Low and Perner 2012) suggest that some ToM abilities surface much earlier. Also, Rubio-Fernández and Geurts (2013) suggest that children of 2.5 years of age have already acquired ToM-abilities.[4]

[3] This claim relies on children's alleged inability to pass the false belief test before the age of 4 (cf. Wellman et al. 2001); also (see Wimmer and Perner 1983).

[4] See Gallagher (2015) for critical discussion of the relevant tests, whether they evince ToM capacities, and, if so, of what kind.

Importantly, the problem here does not hang on the issue of when children develop some version of a ToM, for the recursivity of the beliefs required for common knowledge (their ranging over other beliefs, that is) would still represent an insurmountable challenge for young children (see Carpenter and Liebal 2012: 165). If this psychological evidence is taken seriously, then Bratman's theory is not able to explain joint action among preschool age children.[5] As Pacherie (2013: 1825) remarks, however, this conclusion is not particularly threatening for Bratman's account given that the explicit aim of this account is to identify a set of sufficient, but not necessary, conditions for shared intentionality. On the other hand, this is threatening for Tomasello and Warneken since they commit to Bratman's framework to explain children early abilities concerning shared intentionality.

In the next section, we sketch out an alternative account of 'minimal shared intentionality' which we argue can address the three difficulties presented above: (i) showing that there is an internal articulation between helping behavior and shared intentionality that can help substantiate a full-fledged CEH, (ii) explaining protest and reengagement targeted at the participation of others in joint activities, (iii) providing a less cognitively demanding explanation of the ability to cooperate with others than Tomasello's Bratmanian model. Taking all of these results together we argue that the approach to minimal shared intentionality here advocated is a plausible candidate for substantiating CEH, a task that is left open to be further pursued in the future.

5.4 Evolving Collective Intentionality

We argue in this section that the challenges faced by Tomasello's account of cooperation might be overcome by an altogether different conception of shared intentionality, one in which collaboration and help flow from the structure of the shared activity without the need to postulate any further factor, such as an innate altruistic tendency to support one another.

Instead of distinguishing two forms of shared intentionality—one joint and one collective—as Tomasello et al. do, we propose to understand shared intentionality as collective from the outset, having different stages of development (see Satne 2016). Collective or we-intentionality is the capacity to take part in a group or a '*we*'. According to this view, individuals display collective intentionality when they

[5] These arguments leave open the possibility for other psychological mechanisms to explain joint action in young children. Pacherie (2013), for example, proposes that joint action in young children is explained by their ability to group identify. However, the success of Pacherie's account depends on how the psychological notion of group identification is understood. On her view, group identification presupposes the ability to adopt the group's perspective. Because of this assumption, Salice and Miyazono (2019), argue that Pacherie's theory may also turn out to be too cognitively demanding for the purposes of explaining the first and very early forms of social interaction among children.

are part of a group that sets itself to do something. Our hypothesis is that the ability to adapt and interact with others by pursuing shared goals is prior to joint intentionality which involves representing mental states and other's perspectives, and that collaboration is based on a sense of engagement and group membership that starts to develop quite early in infancy (see Hobson 2002; Reddy 2015, see Over and Carpenter 2009; Buttelmann et al. 2013).[6]

This view takes as its point of departure the apparent fact that children are immersed in social practices from the beginning of their lives, e.g. interacting with family members, care-givers, and peers, including siblings and other social group or family members. It has been claimed that children shift towards a "socio-centric mode of reasoning" by the age of 3 (cf. Dunham and Emory 2014), when they begin to encode group membership and to display in-group favoritism. But recent research has traced the emergence of sociality back even earlier. Although it is a matter of debate whether these findings illustrate genuine understanding of group membership (cf. Buttelmann et al. 2013: 427), Kinzler and colleagues (2011) have found that preschoolers, when learning new information, trust native-accented speakers more than speakers with a different accent. Similarly, Buttelmann and colleagues (2013) show that, from the age of 14 months, children imitate those speaking in their native language (in-group members) more faithfully than those speaking a foreign language (out-group members). 18-month-old infants helping behavior is the target of Over and Carpenter (2009) experiments. In a series of experiments they show that the mere hint of affiliation dramatically increases prosocial behavior in infants that age, In these experiments, 18-month-old infants helped a person in need more often, and more spontaneously, when primed with photographs evoking affiliation (images of human-like puppets interacting) compared to when primed with photographs evoking individuality (images of human-like puppets not interacting or alone). These experiments suggest that there is a strong correlation between group membership and helping from very early age.

Following the experimental research, we propose that the form of intentionality that precedes and grounds what Tomasello calls 'collective intentionality' is a more primitive, minimal, but already collective form of intentionality. This is a kind of social intentionality that can be thought to underlie and support the development of the full-fledged kind that involves conventions, language, institutions, and objectively valid contents. It is a primitive kind of group behavior that is not preceded by 'joint intentionality' in Tomasello's sense, but rather is integrated with more basic forms of joint action. A form of 'minimal' collective intentionality that designates special forms of group acting and feeling that are displayed in jointly acting. Such early forms of joint action do not require representations of others' mental states from the outset, but spring from common background norms and habits that in turn provide a platform for the emergence of representational capacities. We surmise that this minimal variety of collective intentionality is associated with a minimal and

[6] As noted above, Pacherie provides a collective view of early joint action. For our reservations about the prospects of such an account for substantiating CEH, see fn. 4 above.

unsophisticated understanding of others' mentality and agency. Furthermore, the minimal form of collective intentionality at issue is basic also in the sense that it does not depend on taking part of already stable groups and might well take place on a one-to-one basis, e.g. when children interact in dyads with care-givers. Thus, minimal collective intentionality characterizes interactions that are particularly pervasive in the 1st years of life, namely, second-personal relations between particular individuals that interact with one another in intentional activities through which they pursue collective goals.[7]

Thus, the proposed account provides a way of substantiating CEH by committing to the idea of an evolving capacity for collective intentionality. This view proposes that social interaction begins without the exercise of perspective-taking and that with time and exposure to multiple sources of social learning and training a more developed form of collective intentionality—including the capacity to act and think within more complex groups—develops progressively, including those forms that involve the capacity to explicitly represent foreign perspectives. Essential to this progression is the emergence/learning of language, which is an external cultural reservoir of rules and norms, and other sorts of environmental scaffolds, in the context of which objective content would have its proper place.

Psychologically this means that the development of the capacity for collective intentionality breaks into several stages. In the case of infants, the ability to coordinate and attune to others' directed actions and reactions might suffice for making sense of a minimal capacity for collective intentionality, belonging to very simple activities in early infancy. For example, 2–4-month-old babies have been shown to accommodate and bodily anticipate the usual style of their care-givers pickup postures (see Reddy 2015; Reddy et al. 2013) and 6-month-old babies to bodily comply with intentional directives (including linguistic ones, Reddy et al. 2013, Reddy 2015, see also Rochat 2015). According to the minimal collective view, what is characteristic of these forms of interaction is that each participant's role in the activity is determined and structured by the shared activity in which they—as a group/we/dyad—are taking part, without the children planning ahead or intellectually predicting and representing the mental states of their partners (see Satne forthcoming).

For older children, the ability to identify with a group might be the basis for sharing more sophisticated goals with others (Rakoczy et al. 2009, see Salice 2015, Salice and Miyazono 2019): the ability to adopt the group perspective enables, e.g., the formation of long-standing goals whose achievement requires complex planning, coordination through time, and presupposes the exercise of normative pressure on the individuals to provide the expected contribution (Gilbert 2014). In addition, the progression from one to the other form of collective engagement involves the scaffolding of shared intentional activities by shared norms and the exercise of guided and individual practical reasoning, thus making it possible for more sophisticated reason-guided collective behavior to emerge (Brownell 2011).

[7] For a thorough argument for why one ought to consider second-personal interactions of this sort a form of collective intentionality, (see: Satne forthcoming).

While there is much more to be said about the details of our proposed alternative view on collective intentionality (for more see Satne 2016, Satne forthcoming, Satne and Salice 2015; Hutto and Satne 2015), our aim in this paper is to show how an account of minimal collective intentionality committed to the framework just presented can address the three difficulties presented above to Tomasello's two-step account of CEH. With that aim in mind, in the following we present a sketch of the psychological mechanisms sufficient for minimal collective intentionality. Finally, we show that the collective view here advocated can provide answers to the three challenges presented to Tomasello's way of substantiating CEH.

5.4.1 Minimal Collective Intentionality in Ontogeny

According to the view on minimal collective intentionality here advocated, young children have dispositions to engage in a kind of group behavior that does not require that they represent the other's perspectives on their actions.

Taking together a number of different empirical studies and psychological theories of development, in the following we describe three psychological dimensions of action and cognition that are sufficient for making sense of minimal collective intentionality in the sense defined. Those are (i) social conformism; (ii) basic forms of spontaneous interpersonal coordination; and (3) unsophisticated forms of social cognition. All three elements are key for making sense of shared action, and social learning, leading to the emergence and development of more complex collective and social behavior. In what follows we briefly describe each.

5.4.2 Social Conformism

"Social Conformism" describes a set of mechanisms for shaping children's dispositions to match the behavior of other members of the community and continue monitoring them for compliance (Haugeland 1998).[8] This cluster of mechanisms have been studied by several researchers under different headings: *viz.* mechanisms for 'mindshaping' (Zawidzki 2013), 'mutual adaptation' (Kiverstein and Roepstorff 2016), 'folk-psychological normative regulation' (McGeer 2007, 2015), 'in-group bias imitation' (Buttelmann et al. 2013), 'natural pedagogy' (Csibra and Gergely 2009), normative enforcement (Schmidt et al. 2012; Rakoczy et al. 2008) and imitation/emulation (Tomasello 1999a, b; Call et al. 2005; Tomasello and Rakoczy 2003;

[8] The core idea can be found in Haugeland (1990), who describes it in the following way: "when community members behave normally, how they behave is in general directly accountable to what's normal in their community; their dispositions have been inculcated and shaped according to those norms, and their behaviour continues to be monitored for compliance" (Haugeland 1990: 406).

Tennie et al. 2009, 2012).[9] These mechanisms of social conformity get the practice of learning and teaching off the ground and do not require individuals to purposefully comply with rules from the get-go. Instead of representing rules and others' assessments of their actions, individuals are sensitive to others' assessments through more minimal cognitive pathways such as emotional tuning to others' approval and disapproval (see Satne 2014).

5.4.3 Coordination

Coordination refers to the individuals' capacity to follow others and conform to group behavior (Pacherie 2013, 2015). Several studies have shown an innate tendency in humans to coordinate their own bodily movements and, progressively, more sophisticated activities with others (see Knoblich et al. 2011). Studies from interactive dynamics also call attention to these mechanisms of coordination or spontaneous coping (see De Jaegher and Di Paolo 2007 for a review). This capacity is crucial for understanding how joint coordinated activities are possible from a very young age and to understand their development into more complex coordinated activities later in life when combined with other, more sophisticated skills, including those involved in complex planning.

5.4.4 Basic Forms of Social Cognition

Early forms of social cognition play a key role in accounting for the ability to engage in minimal collective intentionality, for example, in explaining how children are able to identify others' approval or disapproval by tuning to someone else's reactions. Instead of having recourse to TT and ST or a combination of the two, as Tomasello does, this proposal aligns itself with the phenomenological tradition and in particular with the idea that at least some mental states, like basic emotions, are directly observable (see Merleau-Ponty 1965; Zahavi 2014; Gallagher 2001, 2015). Importantly, this tradition has challenged the idea of representation as the building block of cognition. Both enactivist accounts of cognition (see e.g. Noë 2006; Gallagher 2017; Varela et al. 1991; Hutto and Myin 2014) and phenomenological ones (Merleau-Ponty 1965; Gallagher 2001, 2015), reject the idea that what one is doing when cognizing the world and other subjects within it, is building up representations of objects and states of affairs. Rather, such accounts understand (basic) cognition in extensively, relational, interactive, and non-representational terms.

[9] There is a debate concerning the demandingness of the cognitive machinery required to make sense of this set of social comformism capacities (see: Tomasello 1999a, b; Rakoczy et al. 2008; Csibra and Gergely 2009; and for a critical discussion, Satne 2014). Here for reasons of space we advocate the minimal view without further discussing this issue.

According to this framework, cognition is not based on internal representations but conceived in terms of different sensorimotor responses to affordances in the environment, that is, in regard to our possible actions or the possible actions of others in a given situation (Gibson 1979).[10] Some views—e.g., the radical enactive approach (Hutto and Myin 2014, 2017)—hold that some forms of cognition are representational but deny that this is the case with root and, especially, early forms of cognition. In particular, they claim that representational cognition is built upon and depends on the mastery of special linguistic practices of claim-making (Hutto and Satne 2015; Hutto and Myin 2014, 2017). While we won't provide further arguments in favor of these accounts of cognition in this paper, we surmise that an account of the early forms of basic social cognition can appeal to recognitional dispositional capacities instead of representational ones to account for the developmental and evolutionary trajectories of human cognitive abilities.[11] In the same vein, some studies in developmental psychology have provided evidence that face-to-face encounters involve direct perceptual understanding of others' emotions and intentionally directed actions (Reddy et al. 2013; Rochat 2003, 2015) and it has been argued that these do not need to implicate the explicit or implicit use of representations, inferences, or simulations (Gallagher 2001; Zahavi 2008; Hutto 2017; Reddy et al. 2013; Zahavi 2014; Satne 2014). When it comes to substantiating the CEH, the perceptual account, briefly sketched out here, has some interesting advantages vis-á-vis ST and TT models of social cognition. It does not presuppose that either simulations of theoretical inferences are necessary for primitive forms of social cognition to emerge, allowing for a cognitively undemanding account of basic social cognition, and thus provides an important tool for the minimal collective view to account for very early forms of social interaction in babies as young as 2 months of age (Reddy 2008; Rochat and Striano 1999).

5.5 Tackling the Problems of Tomasello's Account of Helping and Shared Intentionality in Young Children

5.5.1 Continuous Development

According to the collective view outlined above, helping behavior in children is intrinsically articulated to their gradually developing understanding of themselves as part of a group or a 'we'. In the view here advocated, children would engage in

[10] See Reed (1991) for a non-representational account of social affordances and their developmental and evolutionary trajectories.

[11] For a view on minimal cognitive capacities based on Millikan's concept of push-me pull-you representations (PPRs), (see: Salice and Miyazono 2019). This account can be thought to be compatible with radical enactive views as long as one agrees with Millikan that the organismic responses at issue are intenṙional—i.e. extensional—and not intenṡional (for detailed discussion, see: Hutto and Myin forthcoming and Hutto and Satne 2015).

helping behavior insofar as they are contributing to a shared goal, a goal they have insofar as they belong to a group (maybe a dyad) to which they see the adult as also belonging. Thus, this sort of helping behavior is thought to be intrinsically related to a root ability to share goals with others, i.e. the capacity for minimal collective intentionality.

As was said above, Warneken and Tomasello (2007) claim that: "'infants' performance on helping and cooperation tasks revealed *no straightforward associations* between the two suggesting that these activities differ in important ways" (29, our emphasis). Contrariwise, the proposed collective view, developed according to a less cognitively demanding model of shared intentionality, claims that shared intentionality is based on a sense of involvement and membership that starts to develop quite early in infancy (see Hobson 2002; Reddy 2015). Our view may thus contribute to putting at the center of the empirical research on child social behavior the allegedly existent links between helping and cooperation—e.g. help could be measured with respect to in-group and out-group scenarios (see Over and Carpenter 2009, for an example). If help is proven to be modulated by group membership from very early on in development, then the emergence of social cooperation and group life in phylogeny—concerning whose existence there is solid scientific agreement—might contribute to an explanation of its emergence as well as its central role in modern human lives.

5.5.2 Explaining Protest and Reengaging Behavior

Allowing for collective action in early infancy allows us to make sense of children's complaining and reengaging behavior: the interest in redoing the actions targeted in Warneken and Tomasello's experiments would not be tied to the achievement of one's individual goal. It would be intrinsically tied to a sense of contributing to a collective goal—and to a developing interest in belonging and being recognized as a member of a group. In the collective view, these are the motivations of protest and reengagement behavior in children found in Warneken and Tomasello's studies. But repetition is key in this account of minimal collective intentionality, for group membership is built up through manifold and repetitive interactions with others.

5.5.3 Cognitive Demandingness

The collective view puts the possibility on the table that the ability to adjust to the behavior of groups or another individual is what makes it possible for children to start engaging in more and more complex shared activities with others. It claims that this sort of engagement might be possible without representing the intentional states of others and argues that coordination and attunement might suffice for very simple tasks in early infancy. For older children, the ability to identify with a group and

follow shared norms may offer a new platform for them to share collective goals with others. Both the former and the latter abilities may be thought to be cognitively less demanding than Bratman's conditions on joint action, being based on emotional sharing and imitation (Satne 2014; Buttelmann et al. 2013; Over and Carpenter 2009), rather than perspectival representations, meta-representations and complex instrumental reasoning.

In light of the foregoing observations, one might conclude that this account of minimal collective intentionality seems to offer a promising alternative to Tomasello's view concerning shared intentionality in that it allows one to make sense of the three problems that Tomasello and Warneken face. That is, it predicts and accounts for continuity between cooperation and helping, it explains and predicts the data on child helping behavior (their protesting behavior and their tendency to want to repeat the joint action), and it is a less cognitively demanding account. Thus, our alternative account seems to be a good candidate for substantiating CEH.

5.5.4 Adjudicating Between the Views

Tomasello himself has acknowledged that this proposal "has some merit" (Tomasello 2016:122), in his words:

> [..]Satne has a somewhat different proposal. Her problems with the intermediate step of joint intentionality lead her to basically try to avoid it altogether. She proposes that we should just think about children becoming enculturated into the wider group from the beginning, without any intermediary stage in which they form special kinds of relationships and engage in special kinds of activities jointly with other individuals. There is no problem in principle with this approach, and I am sure it applies to some animal species […] And I am open to the idea that children are adapting to the group in parallel to their forming special kinds of individual relationships with others. (Ibid).

But he complains that

> It is just an empirical fact that young children are first skillful with joint intentionality before they are skillful with collective intentionality, and it seems obvious that the kinds of relationships and interactions that children have experienced with other individuals should have some effect on the way they relate to and interact with the group […] I simply do not see how we can leave out of account the kinds of second-personal relationships that are so important to human social interaction, for example, joint commitments that are made to other individuals.[….]. And I would say that empirically it is the case that infants and young children interact in quite sophisticated ways when interacting with a single other individual, and when they are in groups they mostly simply act in parallel with others or they engage with another individual. (ibid).

Might these remarks pose a problem for the proposed strategy on how to substantiate the CEH?

As it is apparent, the point that Tomasello is making in the quoted passage is about ontogeny. As we said before, many in the field of evolutionary studies doubt that ontogeny does in fact recapitulate phylogeny (For an overview, see Gould

1977). But if that exact mirroring relation is placed in doubt, then it becomes possible to claim that, when it comes to phylogeny, the one-to-one explanation is quite implausible. This is especially so because the scarce archaeological evidence we have regarding shared activities in phylogeny—vestiges of ritual, symbolic, and artistic-craft practices—does not point to such one-to-one relations but rather to the collective hypothesis. Thus, in the case of phylogeny the primitive collective view is much more plausible.

What about ontogeny? Tomasello claims that children do not interact collectively before they interact in one-on-one second personal relationships with specific others. But this is not a direct objection to the collective view, for, as said before, minimal collective intentionality does not depend on the existence of already stable groups, rather it describes and encompasses one-on-one interactions. Those are characterized as second-personal relations between particular individuals that interact with one another in intentional activities through which they pursue *collective* goals. The discussion with Tomasello then boils down to the issue of how one interprets second-personal relations. On his view, they are Bratmanian joint cooperative activities. For the collective view, they are kinds of we-intentional relations. We have already pointed to some reasons for preferring the collective view. Yet, the matter now seems to turn on empirical evidence concerning ontogeny. Could the minimal collective view of ontogeny be defended on empirical grounds?

While evidence regarding older children remains controversial, experiments with very young babies—2–8 months—provide a clue as to why the model offered here can be thought to be better suited than Tomasello's alternative when it comes to addressing the ontogeny of social interactions.

Reddy (2003) argues that babies as young as 2 months old, before any evidence of shared attention, respond to others in emotion-rich interactions and that these early interactions provide them with a *know-how* understanding of the other person being intentionally directed towards them. Reddy provides evidence of babies at 2 months anticipating intentional actions directed at them (for example, they accommodate their bodies anticipatorily when they are about to be picked up by adults, Reddy 2015) and of compliance with intentional directives at 6 months (including linguistic directives: Reddy et al. 2013; Reddy 2015, see also Rochat 2015). This provides evidence for the idea that (inter)action—involving intentional responses and complementary-structured reactions to others—is developmentally prior to understanding and distinguishing two distinct individual perspectives and sharing distributive goals, as Tomasello's proposal suggests. These latter abilities seem to depend on social cognitive abilities, including basic forms of ToM, that are thought to develop later. There are also studies involving 8-month-old babies that support the collective view. Ramenzoni and Liszkowski (2016) performed a number of experiments in which 8-month-old babies that do not yet communicate via the pointing gesture use instrumental actions with the apparent expectation that the other will adopt and predict their goals. The authors tested object-reaching actions of 8-month-old babies in two contrasting scenarios: when an adult was present and when no one was present, selectively changing the position of the desired objects with respect to the infants' action boundaries. They found that the infants in the solo

scenario only reached for the object when it was within their action boundaries, while they selectively increased their reaching for far-out objects outside their action boundaries when a parent or a less familiar adult sat alongside. The authors assess their results in the following way:

> This proposal shares much with the 'shared intentionality' hypothesis (Tomasello et al. 2005), although we do not wish to make the claim that infants in their first year of life have a differentiated understanding of shared plans and goals. Rather, this latter understanding may be a later developmental achievement that builds on further social interactional experiences (Butterfill 2012). Earlier in development, *infants may simply assume that other individuals participate in their everyday perceptual and action experiences*, such that […] they *come to expect a helping hand* when their own falls short of reaching a goal. (Ramenzoni and Liszkowski 2016: 7, our emphasis)

The minimal collective view gives a very congenial explanation of the behavior of the infants in the experiment, according to it infants expect the adult to contribute to their actions because they take the adult to be involved in a collective action with them, not because they have an intellectual grasp of the notion of intention as applied to themselves and the adult. Rather, the baby has practical understanding of the activity she is engaged in as a joint activity in which the adult also has a role to play—namely, in this case, that of providing her with the object she is reaching for (see Satne forthcoming).

The important point here is that the collective intentionality explanatory framework does not presuppose the kinds of cognitively demanding capacities that Tomasello attributes to young children to explain second-personal interactions, and thus it should be preferable as a hypothesis concerning how infants, including young babies from 2 to 8 months, understand others when interacting with them in one-on-one scenarios.

5.6 Concluding Remarks

The aim of this paper was to provide a plausible alternative to Tomasello's way of pursuing the Cooperative Evolutionary Hypothesis. While the proposal was outlined in broad strokes, its main attraction is that it might amount, if substantially developed further, to an account of shared intentionality that dispenses with the condition that shared intentionality requires representing the other's perspectives from the get-go. In this way, it sets the stage to tell a story about the emergence and development of such representational capacities. Introducing a minimal form of collective intentionality to account for early development may then provide one with the right framework and tools to pursue the CEH or so we have argued.

Acknowledgements Previous versions of this paper were presented at Aarhus University, Denmark (2016); Maynooth University, Ireland (2016); at the University of Wollongong, Australia (2016); at Monash University, Australia (2017); and at the AAP 2017 Annual Conference at the University of Adelaide. We are grateful for the valuable feedback from audiences at such venues.

References

Bratman, M. (1992). Shared cooperative activity. *The Philosophical Review, 101*, 327–341.

Brownell, C. A. (2011). Early developments in joint action. *Review of Philosophy and Psychology, 2*, 193–211.

Brownell, C., Ramani, G., & Zarwas, S. (2006). Becoming a social partner with peers: Cooperation and social understanding in one- and two-year olds. *Child Development, 77*(4), 803–821.

Buttelmann, D., Daum, M., Zmyj, N., & Carpenter, M. (2013). Selective imitation of in-group over out-group members in 14-month-old infants. *Child Development, 84–2*, 422–428.

Butterfill, S. (2012). Interacting mindreaders. *Philosophical Studies, 165*(3), 841–863.

Call, J., Carpenter, M., & Tomasello, M. (2005). Copying results and copying actions in the process of social learning: Chimpanzees (Pan troglodytes) and human children (Homo sapiens). *Animal Cognition, 8*(3), 151–163.

Carpenter, M., & Liebal, K. (2012). Joint attention, communication, and knowing together in infancy. In A. Seemann (Ed.), *Joint attention. New developments in psychology, philosophy of mind and social neuroscience* (pp. 159–182). Cambridge, MA: The MIT Press.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense making. *Phenomenology and the Cognitive Sciences, 6*(4), 485–507.

Dunham, Y., & Emory, J. (2014). Of affect and ambiguity: The emergence of preference for arbitrary Ingroups. *Journal of Social Issues, 70*, 81–98.

Gallagher, S. (2001). The practice of mind: Theory, simulation or primary interaction? *Journal of Consciousness Studies, 8*(5–7), 83–108.

Gallagher, S. (2015). The problem with 3 year-olds. *Journal of Consciousness Studies, 22*(1–2., 2015), 160–182.

Gallagher, S. (2017). *Enactivist interventions*. Oxford: Oxford UP.

Gibson, J. (1979). *The ecological approach to visual perception*. New York: Psychology Press.

Gilbert, M. (2014). *Joint commitment: How we make the social world*. New York: OUP.

Gould, S. (1977). *Ontogeny and phylogeny*. Cambridge, MA: Harvard UP.

Haugeland, J. (1990). The intentionality all-stars. *Philosophical Perspectives, 4*, 383–427.

Haugeland, J. (1998). *Having thought: Essays in the metaphysics of mind* (pp. 305–361). Cambridge, MA: Harvard University Press.

Hobson, P. (2002). *The cradle of thought*. Oxford: Palgrave Macmillan.

Hutto, D. (2017). Basic social cognition without mindreading: Minding minds without attributing contents. *Synthese, 194*(3), 827–846.

Hutto, D., & Myin, E. (2014). *Radicalizing enactivism*. Cambridge, MA: MIT Press.

Hutto, D., & Myin, E. (2017). *Evolving enactivism*. Cambridge, MA: MIT Press.

Hutto, & Myin. (forthcoming). Much Ado about nothing? Why going non-semantic is not merely semantics", *Philosophical Explorations*.

Hutto, D., & Satne, G. (2015). The natural origins of content. *Philosophia, 43*(3), 521–536.

Kiverstein, J. & Roepstorff, A. (2016). The proof of the pudding: Coordination and common ground.

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.

Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: State of the art. *British Journal of Developmental Psychology, 30*, 1–13.

McGeer, V. (2007). The regulative dimension of folk psychology. In: Daniel, D. Hutto, & Matthew Ratcliffe (Eds.), *Folk psychology re-assessed* (pp. 137—156). Kluwer/Springer Press.

McGeer, V. (2015). Mind-making practices: The social infrastructure of self-knowing agency and responsibility. *Philosophical Explorations, 18*(2), 259–281.

Merleau-Ponty, M. (1965). *Phenomenology of perception*. London: Routledge.

Michael, J., Christensen, W., & Overgaard, S. (2014). Mindreading as social expertise. *Synthese, 191*(5), 1–24.
Michael, J., Sebanz, N., & Knoblich, G. (2015). The sense of commitment: A minimal approach. *Frontiers in Psychology, 6*, 1968.
Noë, A. (2006). Experience without the head. In T. Gendler & J. Hawthorne (Eds.), *Perceptual experience* (pp. 411–433). Oxford: Oxford University Press.
Over, H., & Carpenter, M. (2009). Eighteen-month-old infants show increased helping following priming with affiliation. *Psychological Science, 20*(10), 1189–1193.
Pacherie, E. (2013). Intentional joint agency: Shared intention lite. *Synthese, 190*(10), 1817–1839.
Pacherie, E. (2015). Modest sociality: Continuities and discontinuities. *Journal of Social Ontology, 1*(1), 17–26.
Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology, 44*(3), 875–881.
Rakoczy, H., Warneken, F., & Tomasello, M. (2009). Young children's selective learning of rule games from reliable and unreliable models. *Cognitive Development, 24*, 61–69.
Ramenzoni, V., & Liszkowski, U. (2016). The social reach: 8-month-olds reach for unobtainable objects in the presence of another person. *Psychological Science, 27*(9), 1278–1285.
Reddy, V. (2003). On being the object of attention: Implications for self–other consciousness. *Trends in Cognitive Sciences, 7*(9), 397–402.
Reddy, V. (2008). *How infants know minds*. Cambridge, MA: Harvard UP.
Reddy, V. (2015). Joining intentions in infancy. *Journal of Consciousness Studies, 22*(1-2), 24–44.
Reddy, V., Liebal, K., Hicks, K., Jonnalagadda, S., & Chintalapuri, B. (2013). The emergent practice of infant compliance: An exploration in two cultures. *Developmental Psychology, 49*(9), 1754–1762.
Reed, E. (1991). Cognition as the cooperative appropriation of affordances. *Ecological Psychology, 3*(2), 135–158.
Rochat, P. (2003). Five levels of self-awareness as they unfold early in life. *Consciousness and Cognition, 12*(4), 717–731.
Rochat, P. (2015). Self-conscious roots of human normativity. *Phenomenology and the Cognitive Sciences, 14*(4), 741–753.
Rochat, P., & Striano, T. (1999). Emerging self-exploration by 2-month-old infants. *Developmental Science, 2*, 206–218.
Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science, 24-1*, 27–33.
Salice, A. (2015). There are no primitive we-intentions. *Review of Philosophy and Psychology, 6*(4), 695–715.
Salice, A., & Miyazono, K. (2019). Being one of us. Group Identification, Joint Actions, and Collective Intentionality, Philosophical Psychology. https://doi.org/10.1080/09515089.2019.1682132
Satne, G. (2014). Interaction and self-correction. *Frontiers in Psychology, 5*, 798–791. 798-11.
Satne, G. (2016). Joint and various (collective) forms of intentionality. *Journal of Social Ontology, 2*(1), 105–116.
Satne, G., & Salice, A. (2015). Helping behaviour and joint action in young children. *Phenomenology and Mind, 9*, 98–106.
Satne, G. (forthcoming). Understanding others by doing things together. An enactive account. *Synthese*.
Schmidt, M., Rakoczy, H., & Tomasello, M. (2012). Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition, 124*(3), 325–333.
Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: On the evolution of cumulative culture. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364*(1528), 2405–2415.
Tennie, C., Call, J., & Tomasello, M. (2012). Untrained chimpanzees (Pan troglodytes schweinfurthii) fail to imitate novel actions. *PLoS One, 7*(8), e41548.

Tollefsen, D. (2005). Let's pretend! Children and joint action. *Philosophy of the Social Sciences, 35*(1), 75–97.
Tomasello, M. (1999a). *The cultural origins of human cognition*. Cambridge, MA: Harvard UP.
Tomasello, M. (1999b). Emulation learning and cultural learning. *Behavioral and Brain Sciences, 21*(5), 703–704.
Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.
Tomasello, M. (2014). *A natural history of human thinking*. Cambridge, MA: Harvard University Press.
Tomasello. (2016). Response to Commentators. *Journal of Social Ontology, 2*(1), 117–123.
Tomasello, M., & Rakoczy, H. (2003). What makes human cognition unique? From individual to shared to collective intentionality. *Mind and Language, 18*(2), 121–147.
Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *The Behavioral and Brain Sciences*. 2005 Oct*, 28*(5), 675–691.
Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.
Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science, 311*, 1301–1303.
Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy, 11*(3), 271–294.
Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Development, 77-3*, 640–663.
Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false-belief. *Child Development, 72*, 655–684.
Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128.
Zahavi, D. (2014). *Self and other*. Oxford: Oxford UP.
Zahavi, D. (2008). Simulation, projection and empathy. *Consciousness and Cognition, 17*(2), 514–522.
Zahavi, D., & Satne, G. (2015). Varieties of shared intentionality: Tomasello and classical phenomenology. In J. Bell, A. Cutrofello, & P. Livingston (Eds.), *Beyond the analitic-continental divide: Pluralist philosophy in the twenty-first century* (pp. 305–325). London: Routledge.
Zawidzki, T. (2013). *Mindshaping. A new framework for understanding human social cognition*. Cambridge, MA: MIT Press.

Chapter 6
Minimal Cooperation and Group Roles

Katherine Ritchie

Abstract Within philosophy, cooperation has primarily been analyzed in the context of theories of collective intentionality. These discussions have largely focused on interactions between pairs or small groups of agents who know one another personally. Cooperative game theory has also been used to argue for a form of cooperation in large unorganized groups. Here I consider a form of minimal cooperation that can arise among members of potentially large organized groups (e.g., corporate teams, committees, and governmental bodies). I argue that members of organized groups can be minimally cooperative in virtue of playing roles in an organizational structure and having a common goal. The minimal form of cooperation considered here is not grounded in collective intentions involving symmetric mental states, special collective intentional modes, or joint commitments. More generally, I show how considering minimal cooperation in the context of organized groups provides an opportunity to reevaluate the extent to which the social world and social phenomena depend on internalist mental factors (e.g., intentions, beliefs) and externalist non-mental factors (e.g., laws, job descriptions). The view of minimal cooperation among members of organized groups developed here provides support for an externalist rather than internalist theory of at least one social phenomenon.

Keywords Cooperation · Social groups · Roles · Externalism · Collective intentionality

Within philosophy, cooperation has primarily been analyzed in the context of theories of collective intentionality. On these views, cooperation is taken to involve complex often symmetric mental attitudes. These theories center their discussions on small groups usually involving pairs of agents who interact with mutual knowledge of (some of) the others' intentions, beliefs, ends, and commitments (Bratman et al.). An alternative account of cooperation relies on game theory. On these

K. Ritchie (✉)
Department of Philosophy, CUNY Graduate Center and City College of New York, New York, NY, USA

A. Fiebich (ed.), *Minimal Cooperation and Shared Agency*, Studies in the Philosophy of Sociality 11, https://doi.org/10.1007/978-3-030-29783-1_6

accounts, the focus is on cooperation in large unorganized groups, like participants in political demonstrations or flash mobs. Paternotte, a proponent of using cooperative game theory to analyze minimal cooperation, takes collective intentionality accounts to center on "egregious cooperation" and argues that an analysis could be "weakened without making cooperation disappear" (2014, 48). Here I focus on a third sort of group—organized groups—and develop a novel account of minimal cooperation that relies on neither collective intentionality nor cooperative game theory.

Organized groups—like corporate teams, committees, and legislative bodies—need not be composed of pairs or small groups of agents with mutual knowledge of collective beliefs, intentions, goals, and so on. They are not mere coalitions of agents, as in cooperative game theory. Instead, they involve individuals and organizational structures with roles for members to play. I argue that members of organized groups can be minimally cooperative in virtue of playing roles in an organizational structure and having a common goal. The view departs from the analyses of cooperation that begin with collective intentionality and those that begin with game theory. The minimal form of cooperation is not grounded in complex mental states/representations or in beneficial payoffs. We will see that roles individuals play in organizational structures are central to account for the form of minimal cooperation developed here.

Considering cooperation between members in organized groups will also prove illuminating in assessing the extent to which theories of social interaction, and perhaps social ontology more generally, should be construed in internalist or externalist terms. Call views of social phenomena and entities centered on mental representation, mental states, or mental contents **internalist views of the social**.[1] Call views that rely on features that are external to mental states or mental contents—for instance documents, laws, or material resources—**externalist views of the social**.[2] When analyzing cooperative interaction between two agents working in close proximity, it is plausible that both agents have representations of the other agent and of (some of) her beliefs, desires, and intentions. That is, an account that is largely internalist is plausible. However, once we move to a scale at which there are multiple agents interacting in organizations that involve defined roles and action-types, the situation is different. Agents need not personally know all the people with whom they cooperate or have complex representations of others' mental states, or so I will argue. It is useful to have well defined roles for senators and for members of a corporate team as they require one to do less mental work to understand what others believe, intend, and so on. It is costly to represent others' mental states. It is far

[1] Even if one is a mental externalist, one could be an internalist about the social. On such a view one holds that social entities, facts, collective intentionality, etc. depend on mental states or mental contents that are themselves dependent on external features of the natural or social environment.

[2] The distinction is not exclusive as views might include both mental and non-mental features. A view might be more or less internalist/externalist. See Torrengo (2017) for a recent defense of externalism about institutions. I discuss his view further in Sect. 6.4.

easier to represent roles which are simplified in ways that agents—in their full psychologically complicated glory—are not. Playing a role is central to the more externalist account of minimal cooperation among members of organized groups developed here.

The article is structured as follows. I begin by considering the nature of organized groups (Sect. 6.1). In particular, I consider the ways organized groups are structured to involve roles that are defined in ways that are interdependent. Then I consider three prominent views of collective intentionality and how they falter when extended to large organized groups (Sect. 6.2). Since members of organized groups can be minimally cooperative and their cooperative behavior is not well captured by extending a theory of collective intentionality, we need to look elsewhere to understand this form of minimal cooperation. Next, I argue for a view of minimal cooperation that relies on roles group organizational structures impose and a common goal (Sect. 6.3). Then I consider internalist and externalist views of the social and whether we should expect a unified account of cooperative activity (Sect. 6.4). Most theorists hold largely internalist theories of social entities and phenomena according to which mental states and representations ground or constitute the social world and social phenomena. I show how the view of minimal cooperation among individuals in organized social groups supports externalism about at least some social phenomenon. Finally I draw concluding remarks (Sect. 6.5).

6.1 Organized Groups and Group Roles

Before an analysis of cooperation by members of organized groups can be undertaken, we need a better understanding of the nature of organized groups. In Ritchie (2013, 2015, forthcoming) I argued that groups like corporate teams, committees, clubs, the U.S. Senate, and British Parliament are entities with organizational structures. Organized groups are not merely structures. That is, they are not identical to structures. Rather they are entities that have members that are structured or organized in a particular way; they are structured wholes.

The structures of organized groups consist of roles and relations between them. Roles are defined in terms of relations to other roles, tasks that role-players are allowed or required to carry out, and in some cases specific features a role-player must have. Relations between roles might be hierarchical or non-hierarchical. Relations that involve deference and power are hierarchical. For instance, a role might allow a role-player to give orders to individuals playing other roles. Relations of seconding a motion or reporting on a project involve relations between group members that are non-hierarchical. Relations between roles also capture the ways playing a role depends on other roles being played. In discussing organizations Miller (2014) states that "there is a degree of interdependence among … roles" and that "the performance of the constitutive tasks of one role cannot be undertaken, or

cannot be undertaken except with great difficulty, unless the tasks constitutive of some other role" are undertaken. Roles are largely defined in terms of relations; they are interdependent.

Roles are also defined in terms of powers, norms, and responsibilities. These might involve relations to other role players. For instance, some roles involve the power to give directives to other role holders. Norms about how to properly interact might also be part of what defines a role. Roles might also involve powers, norms, or responsibilities pertaining to particular tasks. For instance, some role might involve being responsible for the group's media presence, while another gives power and responsibility to enable a role player to control fundraising efforts.

Finally group organizations might place particular restrictions on who can play a role. For instance, one might be required to be born in a particular country, be at least 35, or meet other external requirements like being appointed or elected by non-group members.[3] A role may allow for multiple role-players or only allow for a single role-player. These could also be specified by requirements on roles.

Consider a simple example of a committee with roles for president, vice president, secretary, treasurer, and member without a leadership role. The role of the president might involve calling meetings to order, delegating tasks to members of the committee that do not have leadership roles, initiating votes, and so on. The president role might be specified so that only someone with a particular nationality can play the role. The member role plausibly allows for more than one role-player, while other roles might allow for only a single role-player.

The Senate, a corporate team, and other organized groups are not just organizational structures. The structure of a group captures its functional organization, but an organized group itself is an entity that *is* structured or that *has* a structure. While a precise metaphysical account of organized groups is not needed for our purposes here, being clear that organized groups are not identical to structures is important for several reasons. First, the same organizational structure might be had by more than one group. For instance, there might be many corporate teams or many committees with roles defined in the same way. If groups were identical to structures, there could not be two groups with the same structure. Groups should not be identified merely in virtue of having the same organizational structure. Members of a group matter for group identity conditions.

Second, organizational structures themselves might be abstract entities. Many philosophers take relations to be abstracta. If organizational structures are complexes of relations they would also be abstract entities. Yet, groups seem to be causally relevant and have spatial and temporal locations; they do not seem to be abstracta.

[3] The conception of roles relied on here bears similarities to the way Koslicki defines places in structures. She states that structures are "entities which make available positions or places for other objects to occupy, provided that these occupants satisfy the type restrictions imposed by the structure on the positions in question" and which impose on the objects "a particular configuration or arrangement" (2008, 235–6). Here I focus on more than just configuration or arrangement, but the focus on relations and type restrictions fits with part of the picture I am offering.

Groups are, therefore, not just organizational structures. They are entities that have structures or that are structured.

An organized group comes to be when some people play the roles required by a group organizational structure. When some person, *a*, stops playing a role *r* and another person, *b*, starts playing *r* the group's membership changes. While *a* was a member of the group, she no longer is. More formally, we can define membership in an organized group as follows:

> *Organized Social Group Membership*: Some things, X, are the members of a group with an organizational structure S at time t and world w if, and only if, together X play the roles in S (i.e., X are related or normatively bound in the ways required by S).[4]

The definition of membership allows for organized groups to change members and captures how membership relates to a group's organizational structure.[5]

6.2 Collective Intentionality and Cooperation

Analyses of cooperation have been addressed largely in the context of collective intentionality. One might think the explanation as to why discussions of cooperation have focused on collective intentions is obvious—cooperation involves collectively intending to φ with others. So, the thought goes, an account of collective intention will also be an account of cooperation. I will ultimately challenge the idea that all cooperation requires collective intentionality. To see why, we first need to consider how collective intentionality is understood.

Definitions of cooperation usually involve "a set of individual intentions that have certain origins and enjoy certain relations, all of which is common knowledge (that is, public or transparent) among agents" (Paternotte 2014, 47). While sharing a common core, views vary according to whether they involve primitive irreducible *we*-intentions, the ways they take individual intentions and plans to be involved, and whether plural subjects are relied upon. They also vary in being causal or normative explanations.

Bratman, Gilbert, and Searle offer three prominent accounts of collective intentionality and cooperation.[6] Bratman (1992, 1999, 2014) argues for a reductive account of what he calls 'shared cooperative activity'. He analyzes "we intend to J" in terms of individual attitudes as follows:

---

[4] This definition is as in Ritchie (forthcoming), but departs from those in Ritchie (2013, 2015). Note that the predicate 'together X play the roles in S' is a collective predicate. That is, it applies to some things X without applying to each of the individuals (contrast with 'is tall' in 'the players are tall', compare to 'surround the building').

[5] Organized groups might also persist through changes in their organizational structure. Here I will not develop a theory of *how much* structural change a group can sustain.

[6] For an account relying on a so-called *we*-mode (see Tuomela 2006). Miller (2001) gives an account of collective ends or goals that does not rely on collective intentions.

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

The account reduces collective intentionality to individual intentions with subplans that are consistent and can be jointly satisfied (i.e., that "mesh") and common knowledge of one another's intentions.

Gilbert (1989, 2006) argues for an account based on joint commitments and plural subjects. She takes a plural subject to be formed via a joint commitment to do something as a body. Joint commitments involve obligations to the others with whom one formed the commitment to carry out the activity to which they committed. They also give one the right to rebuke other parties in the joint commitment if they fail to act in a way that is appropriate to their goal obtaining. Gilbert holds that forming a joint commitment requires all of the parties bound by it be involved in its creation (2006: 135, 138, 168). She states "[t]he parties to a joint commitment … are those who comprise both its creator and its subject" (135). Going further she states that it is a "condition that all of the parties must be involved in the creation of a joint commitment" (ibid.).[7] Further, prior to forming a joint commitment all involved must express "readiness for joint commitment" which must be common knowledge among them (2006: 138). In her final analysis Gilbert states that "[t]wo or more people are acting together (doing something together) if and only if: (1) they are jointly committed to espousing as a body the appropriate goal; (2) they are fulfilling the behavioral conditions associated with the achievement of that goal; (3) their satisfaction of these conditions is motivated in each case by the existence of the joint commitment" (2006: 146). Gilbert's account is non-reductive, as committing to do something as a body does not reduce to individual commitments or beliefs.

Finally, Searle (1990, 1995, 2010) argues for a view of collective intentionality that relies on primitive irreducible *we*-intentions, but which does not require a plural/group subject or meshing sub-plans. He argues that when engaged in cooperative collective action participants each have intentions of the form 'we intend that we φ'. According to Searle this involves a special collective or *we*-mode of intending that is not reducible to mere individual intentions in an *I*-mode. Searle's account also involves common knowledge. He states that "[c]ooperation *implies* the existence of common knowledge or common belief," but common knowledge "together with individual intentions to achieve a common goal is not by itself sufficient for cooperation" (2010, 49). Rather, he holds that "[c]ooperation requires the collective intention to

[7] There are, however, points at which Gilbert appears to want to walk back from this strict requirement. For instance, she states that there are cases "where one or more people sign on to a joint commitment originally created by others" (2006: 134). Given the role creation is given in her account of plural subjects and joint commitments, the theory would need to be modified to allow for joint commitments to bind those who are not co-creators.

cooperate," which is an intention that *we* cooperate or that 'we intend that we φ' (2010, 58).[8]

It is worth noting that the core cases on which Bratman, Gilbert, and Searle focus involve small groups of individuals. Gilbert states that "[i]n the example of acting together on which I focus there is no hierarchy, the people involved certainly know of one another and are in the midst of a relatively substantial personal interaction" (2006: 99).[9] Similarly, after mentioning basketball teams and orchestras, Bratman states that "to keep things simple" he will focus only on "shared cooperative activities that involve only a pair of participating agents and are not the activities of complex institutions with structures of authority" (1992: 327). He suggests that we can "gain some insight" into shared intentionality by starting small and that "[p]erhaps our theory of small-scale shared agency can, with due adjustment and further additions, be extended to … larger social organizations" (2014: 8). Starting with simple cases can be a good methodological strategy. Simple cases are easier to analyze; they have fewer "moving parts", if you will. If one is concerned with understanding how cooperative or collective behavior comes to be, simple cases rather than highly complex institutional cases also seem like a more fruitful place to begin inquiry. However, even if starting with simple cases is justified, it would be hasty to conclude that the same analysis of simple cases can always be scaled up to larger organized groups.

Groups like teams, committees, and legislative bodies are very different from two people going for a walk together or three people cooperating while preparing a meal. They involve organizations that have defined often hierarchical roles. Organized groups can vary in members across times and worlds and may be extremely large. They seem different from Gilbert's plural subjects which are formed whenever some individuals form a joint commitment to do something as a body. Moreover, individuals who are currently members of a group might not have been involved in forming the group, its organizational structure, or its goals. Recall that Gilbert requires that all parties bound by a joint commitment must be involved in creating it. In passing several times she states that individuals might join joint commitments, but her theory as espoused in (2006) would need to be modified to allow for this. In considering Gilbert from here, I will be considering her account as stated. On that account, members of many organized groups do not form plural subjects. Yet, she requires plural subjects in her account of acting together.

[8] Searle also discusses what he calls 'collective recognition', a notion that is weaker than cooperation. He says that "collective recognition need not be a form of cooperation and thus does not require a collective intention to cooperate" (2010, 58). Here I am concerned with considering whether cooperation should require collective intentionality in the way theorists have often argued. So, I focus on Searle's view of cooperation rather than collective recognition.

[9] Although, Gilbert (2006) takes it to be possible for there to be large plural subjects which are impersonal (i.e., in which not all members know one another personally), anonymous (i.e., when members do not know that particular individuals exist at all), and hierarchical.

Members of some groups—like a large legislative body or a corporate team—might not know every other members. They also might not know which individuals are playing which roles or whether there are multiple individuals playing some roles. If a member, *a*, does not know that *b* is a member of the group, it is implausible to suppose that *a* has representations of *b*'s intentions, beliefs, or plans. Knowing that there are particular roles being played is not enough to enable one to have representations of mental states. After all, roles don't have mental states! Further, members of an organized group might not know what many roles involve. They might not even know what roles the group organization includes.

Without knowing who is playing a role the accounts given by Bratman and Searle appear not to apply. On Bratman's account individuals are required to have knowledge of others' intentions in order to engage in cooperative shared activity. Having a *we*-intention, as required in Searle's account of cooperation, might not be possible without knowing who is included in the group or plurality picked out by *we*. Even the condition of common knowledge or belief is challenged by the epistemic opacity about individual identities that can occur in large organizations. For instance, if p being common knowledge among a group G requires that each member of G have representations of others knowing that p (as well as iterations of higher-order attitudes), then knowing who the members of G are is required for common knowledge.[10]

Members of organized groups certainly seem to be able to cooperate. Members of teams, committees, and legislative bodies do complete goals, often in ways that are more efficient and effective than an individual or unorganized group of individuals could. Any account of cooperative activity that is meant to be general, must be able to account for such cases. The accounts just considered cannot, at least in their current forms, correctly accommodate *all* cases of cooperation. Moreover, it does not appear that they can be easily modified to handle the cases I focus on here. This is particularly so for Bratman and Searle's accounts. The methodological assumption of starting with small group one-off interactions and scaling up to large group long-term interaction should be abandoned.

In the next section I argue for a form of minimal cooperation that relies on roles in a group organizational structure. Collective intentionality as analyzed by Bratman, Gilbert, and Searle is not necessary for minimal cooperation among members of organized groups. Before turning to what I take is required for minimal cooperation in organized groups, what I am not arguing for needs to be clearly specified.

My aim is not to argue that the accounts of collective intentionality and cooperation just considered fail. One of the accounts, or another in the same spirit, might be correct for small groups working together on a specific joint action. For instance, a

[10] Not all definitions of common knowledge require that agents know who all of the members of the set of group are. Gilbert (2006) argues that 'population common knowledge' can hold even in cases in which group members do not know one another, but do know that there is a relevant population. So Gilbert's account does not succumb to this version of the problem. However, given that she does spell out joint commitments and plural subjects in a way that requires creation by all those committed, her account still falters. See also Paternotte (2014) for discussion of common knowledge in conditions of anonymity and further citations.

case of two people organizing a closet together might require symmetric attitudes involving representations of others' mental states. Moreover, Bratman and Gilbert explicitly state that these are their primary target cases.[11] So, my argument is not against their accounts given their particular target phenomena.

I am also not arguing that members of organized groups *never* cooperate in ways that should be analyzed in terms of joint commitments, representations of others' intentions, or *we*-intentions. Rather, I argue that there are cases of minimal cooperation among members of organized groups that do not involve collective intentions as spelled out in any of the accounts just considered. The accounts considered above fail to give necessary conditions for minimal cooperation. A theory relying on collective intentionality might be correct for cases involving extensive or what Paternotte calls "egregious" cooperation, but minimal cooperation among individuals playing roles in group organizations does not place such stringent mental requirements on agents.

Finally, I am not claiming that the creation of organized groups or organized group structures is independent of collective intentionality. It might be the case that an account of collective intentionality like those just considered is often part of the explanation for how organized groups or a group structures *come to be*. For instance, some individuals might form a *we*-intention to create a certain group or a group with particular defined roles. Tomasello suggests that cooperative interaction in humans requires that "participants coordinate their roles—their plans and sub-plans of action, including helping the other in her role as needed—which are interdependent" (2009: 61). These minimal roles that are formed in one-off joint actions are plausibly the precursor to the reified institutionalized roles that humans create in organized group structures. I am not arguing for a view of how organized groups or group structures come about. As far as I am concerned here, it is an open question whether a group coming to be requires collective intentions. That is a question in social ontology, rather than a question about the conditions required for cooperation. Instead I am arguing that even if complicated symmetric attitudes, the formation of a joint commitment, or special *we*-intentions are needed to bring about an organizational structure or a group, these are not necessary for members of organized groups to minimally cooperate.

To put the point another way, consider diachronic and synchronic accounts of cooperation in organized groups. It might be true that a complex representational account with symmetric attitudes is required to create an organized group with a particular structure composed of roles and relations. A diachronic account may rely on a view of cooperation like one discussed above. Yet, once an organized group with various defined roles exists, minimal cooperation need not involve anything as mentally complex as that posited by Bratman, Gilbert, or Searle. A synchronic account of minimal cooperation among members of an organized group, might be quite different. The organizational structure allows for group members to minimally

[11] Although Gilbert (2006, especially Ch. 8) argues that the account can be extended to large social groups like societies. I return to her account below.

cooperate by playing particular roles.[12] Once a group is "built" the ladder of collective intentionality that was used in constructing it can be "knocked away".[13] Once an organizational structure is in place, complex mental work involving mutual knowledge, representations of others' mental states, and formation of joint commitments are not necessary for cooperation.

6.3 (Minimal) Cooperation + Roles

In Sect. 6.1, I argued that organized groups are structured wholes. They have structures with roles that are defined in ways that depend on relations to one another. The interdependence of roles relates to the powers, obligations, and rights role-players have. It also relates to the way actions undertaken by one role-player might be part of larger actions that involve other role-players. Organizational structures with defined roles allow for a division of labor and also for a division of knowledge. No one member of a team might understand the entire range of activities and roles involved. Organizational structures can allow many individuals together to carry out complex actions that no one member could fully understand or complete.

Let's consider an example. Suppose a consulting firm is tasked with determining whether a merger between Company A and Company B would benefit Company A. In order to determine what to recommend, thereby meeting their goal, the firm puts together a consulting team with various roles. Suppose further that the team has many members who are located across multiple offices and that many team members do not know of one another. Roles involve responsibilities and obligations that normatively bind role-players in various ways. The consulting team includes roles that require role-players to research similar past mergers and pass findings on to members who will include them in a report. Other members have roles that require analyzing a merger's impact on stockholder and customer perception. They too report their findings to members tasked with writing a final report. And so on. Through many members playing their assigned roles—that is, carrying out tasks

[12] Tollefsen suggests something similar about group agents. She states that "the performance of joint actions on the basis of group ends, shared intentions, joint commitments, or we-intentions might very well be the way in which corporate agents form and sustain their agency over time… group ends, joint commitments, shared intentional activity, and we-intentions might all be part of what is happening internally within corporate groups, and this produces a pattern of group behavior that exhibits unified agency" (2015: 47). Here I am focused on members of a group cooperating, rather than on whether a group is an agent, nevertheless there are similarities in my argument and Tollefsen's remarks.

[13] The argument made here bears similarities with Bar-On's (1995) reconstrual of Gricean intention-based semantics. She argues that problems for the Gricean can be avoided if speaker intentions aren't needed to fix meaning *now* (i.e., intentions are not needed for a synchronic account of nonnatural meaning that has been conventionalized), even if meanings *were* fixed by intentions in the past (i.e., speaker intentions are part of a diachronic account of nonnatural meaning). She also uses the metaphor of "kicking away the Gricean ladder".

and interacting in ways team roles require—the team concludes that Company A should merge with Company B as it will benefit Company A to do so.

In the example, some individuals might have been working closely with others in ways that fit with accounts like Bratman's, Gilbert's, and Searle's. Yet, other members never directly interact and have no knowledge of one another. The team had the goal of determining whether a merger should be recommended. Every member, let's suppose, knows that she is part of a team that has said goal. Each intends to act in a way defined by her role to meet the goal. The team met its goal in a way that involved collaboration and, it would be natural to say, cooperation. Yet, the case does not involve cooperation on any of three accounts considered in the last section. Let's see why.

Bratman requires that each of the individuals participating in shared cooperative activity know that each other individual with whom they are cooperating has an intention of the form 'I intend that we J'. Suppose that Kai and Simona are two members of the consulting team. Neither knows the other personally and neither knows that the other is playing a particular role in the team. Kai does not know that Simona has an intention of the form 'I intend that we determine whether a merger between Company A and Company B should be recommended'. Kai does not even know that Simona exists, so she knows nothing about her mental states! If Bratman's account had to be met for some individuals to be engaged in cooperative action, the members of the team would not count as cooperating even in a minimal sense. The account fails to capture that this is a case of cooperation.

Gilbert requires that two or more people cooperatively carrying out an action together are motivated by the existence of a joint commitment (this is condition (3) in her analysis quoted above). She also holds the joint commitments must be created by all of the parties bound by the commitment.[14] In the case of the consulting team, both conditions could fail. For instance, Simona might not be motivated by the joint commitment to meet a certain goal, but rather to a commitment to play the role that she has been assigned. She might want to play that assigned role because she cares about the firm, others' perceptions of her, keeping her job, or for any number of other reasons. She still seems to be at least minimally cooperating, even if she is not *motivated* by a joint commitment. This alone is sufficient to show that Gilbert's account would not classify the team members as cooperating. Further Simona might not have contributed to forming the commitment to determine whether Company A should pursue a merger with Company B. In fact, *no member* of the team needs to have been involved in setting this as the team's end. The commitment to that end might have been set by those higher up in the larger hierarchical organization of the consulting firm. Given that the team members are cooperating, meeting Gilbert's account cannot be necessary for minimal cooperation.

[14] As noted above, Gilbert appears to want to allow for cases in which to join a joint commitment that they did not create. However, her view as espoused in Gilbert (2006) carries an explicit requirement that joint commitments require co-creation by all those they bind. If her account is going to handle cases like those involving organized groups considered here, it needs to be modified so that not all parties of a joint commitment are required to be creators.

Searle requires *we*-intentions and common knowledge for cooperation. The case under consideration can be augmented to explicitly involve members of the team who lack *we*-intentions but who are, at least in a minimal sense, cooperating with others. Suppose that Layla is another member of the team. Layla has no thoughts in an irreducible *we*-mode, but Layla does intend to play her role as well as possible. She understands that playing her role involves others playing their roles. For instance, she knows that part of playing her role is to pass information on to others on the team and that others cannot successfully play their roles without her playing her role. Yet, she never has a *we*-intention. She thinks about what she is doing and about the goal she's taking part in, but lacks *we*-intentions. By amending the case in this way, it does not seem that *we*-intentions have been smuggled in. And if they have been an argument to that effect is needed.[15] Moreover, it seems that Layla is cooperating with others on the team to provide information for a report and on determining whether to advise Company A to pursue a merger. This provides evidence that Searle's account also fails to establish necessary conditions for cooperation.

The corporate team case involves roles that *require* individuals playing them to cooperate, at least in a minimal sense. Playing a role places normative demands on how role-players are to interact and what they are supposed to do. If roles are played, role-players will interact in ways that aim at a shared end in what appears to be a case of cooperative activity. This is so even if members of the team do not know of others and do not have representations of others' mental states. The accounts of collective intentionality fail to categorize the behavior of the members of the team as cooperative; a different account of minimal cooperation is needed.

An account of the cooperation in the cases just considered needs to take seriously the relevance of roles in an organizational structure. I propose that there is a minimal form of cooperation in which some individuals cooperate in φ-ing just in case they play roles in an organized group structure that work towards a common goal. Roles in a group structure function together to allow the larger group to achieve ends or goals. They are defined in ways that are interdependent, that are normatively binding, and that require interaction to achieve goals. Minimal cooperation is achieved when playing roles is combined with a common goal.

The minimal sort of cooperation I propose here is not as robust as cooperative activity as theorized by Bratman, Gilbert, Searle, and others. It need not involve being motivated by a joint commitment, having representations of others' intentions, or having special *we*-intentions. But, I take it that it fits with a common sense understanding of cooperation. In the case studies, individuals are working together by playing roles to achieve a goal. That certainly is, at least, a minimal form of cooperation in the way we ordinarily understand it.

[15] More generally one might challenge the notion of we-intentions as a special primitive irreducible mode of intending. Tollefsen says that we-intentions "seem pretty mysterious" and that "[n]othing in our experience and in our everyday practice of ascribing intentions to others…seems to confirm their existence" (2015: 33).

The way roles are institutionalized or reified in group organizational structures allows for more minimal mental requirements on agents engaged in cooperative activity. Members of certain sorts of groups—those with organizational structures—can cooperate in a way that requires less internal mental representation and less in terms of motivational requirements. Individuals must play their roles that aim at a common goal. Playing a role requires standing in the right relations to other group members, completing particular tasks, and having certain rights and obligations. The complex mental demands placed on individuals by accounts that start with pairs of individuals coordinating action are not necessary when individuals play roles in groups with organizational structures. While individuals in groups will have intentions and might have representations of others' mental states, these are not required for minimal cooperation given the existence of defined roles in an organized group structure.

One might worry that the account I have offered is *too* minimal. Let's consider two cases that might be used to press the worry. First, consider a spy network. The individuals in the network might not know anyone else in the network. Moreover, they might not understand what their roles are, what end they are helping to work towards, or how they are contributing to that end. Nevertheless, the combined efforts of those in the network might fulfill the end. Are they cooperating? Intuitions about such cases are mixed. The account I have given could be adapted to accord with both stances on the case by expanding on what is required for a common goal. If one takes the spies *not* to be minimally cooperating, the account could be adapted to require some mental requirements on having a common or shared goal. For instance, one might argue that minimal cooperation in φ-ing requires playing roles in an organized group structure that they work towards a goal that all know. In the spy network case, not all know the goal, so the case does not involve minimal cooperation. In contrast, if one takes the spies in the network to be minimally cooperating, one could take having a common goal to require less. It might require that roles are functionally integrated to achieve an end. Or, one might require just that someone with authority over the group know the goal for it to be a common goal.

Second, consider the parts of a car engine. Each part plays a role so that together they function to convert chemical energy into mechanical energy in order to move the car. They do not, however, seem to be cooperating even in a minimal sense. Just functioning in concord is seemingly not enough for cooperation. While the account of minimal cooperation I have given emphasizes roles and function, it does not take the parts of a car engine to be cooperating. To see why, note that I argued that there is a form of minimal cooperation that involves playing roles in *an organized group structure*. Roles in group organizational structures involve norms. Role-players are obligated to one another and have responsibilities to complete tasks, give directives, and report to others. The organizational structure of a car engine is defined in terms of ways parts operate, but not in terms of norms. The members of the corporate team are correctly classified as minimally cooperating while the parts of the car engine are not taken to be minimally cooperating. While more could be said to develop a theory of minimal cooperation among members of organized groups in greater detail, I will leave the account as sketched here. In the next section I briefly consider how the account fits into the larger project of theorizing social phenomena.

6.4 Internalism, Externalism, and Continuity in Social Theorizing

Accounts of social phenomena, social entities, and social facts have largely centered on mental representation of some form or other. Moreover, as we saw there has also been a common assumption that theorizing about the social can start small and then be scaled up. Torrengo argues that many hold a **continuity thesis** he spells out as follows: "[s]ocial phenomena are determined by collective intentions and their shared contents in small groups as much as (and in the same ways as) in complex social situations" (2017: 70).[16] The preceding discussion could serve as a challenge to certain views that center on mental representations, mental states, or mental content. Whether it challenges the continuity thesis depends on how small scale cases should be understood. To clarify the discussion, I return to the terminology introduced earlier.

As I use the terminology internalist views of the social are views that take social phenomena or entities to be grounded, constituted, constructed … by mental representation, mental states, or mental contents. On my usage externalist views of the social are those that take social phenomena or entities to be grounded, constituted, constructed… by features that are external to mental states and mental contents, like documents, laws, or material resources.

As we saw above prominent accounts of collective intentionality and cooperation rely on representations of others' intentions, special ways of intending, and common knowledge. Such views are, at least to a large extent, internalist. Views that are more internalist than externalist are also common in social ontology. Collective attitudes, intentions, and representations of a kind itself are often taken to constitute or ground social entities and facts.[17]

Other views in social ontology make mental states, contents, and representations less central.[18] For instance, Thomasson argues that some social kinds might not depend on representations of themselves. She states social entities like "racism, economic recessions, class systems, and gender-biased power structures are

[16] Note that Torrengo himself argues that the continuity thesis is false.

[17] For instance, Searle (2010) argues for a social ontology that relies on collective acceptance of constitutive rules. Hacking (1999) argues for a view on which mental representations of "interactive kinds" lead individuals to "modify their behavior" when they become aware of being classified in a particular way (e.g., as a woman, as a refugee, as a Black person) they may "modify their behavior accordingly" (1999: 32). Mallon recently defined social construction projects as accounts that explain "by appeal to our practices of representing" (2016: 1). He argues for an account of social kinds or roles "that are structured by the representations of human categories and, over time, by the causal effects of such representations" (2016: 210).

[18] Khalidi argues along the same lines holding that the existence of some social kinds fail to require that attitudes "be directed towards the kind itself" (2015: 104). See Ruben (1985) for arguments that certain social phenomena can exist without being noticed (e.g., exploitation). See Guala (2010) for arguments that we might lack knowledge of social kinds, even if they depend on collective intentions.

typically not intentionally created (either directly or indirectly) by accepting constitutive rules about entities of that kind" (2009: 549). They are, instead, "byproducts of more basic social and institutional facts" (ibid.). A general view of social ontology should not be so internalist so as to *necessitate* that representations of social entities are operative in the construction of the social world.

The argument I gave above goes further. It is not just that minimal cooperation does not require a representational state about *being cooperative.* It is the stronger general claim that collective intentions, recognition, acceptance, and joint commitments are not necessary for minimal cooperation at least when individuals are members of an organized group. The view is more externalist than a view that only goes so far as to reject that a social kind K relies on representations of K.

The argument I gave is closer to views that give accounts of social entities in terms of non-mental features. For instance Epstein (2015) argues that facts about social entities can depend on facts that do not involve mental states, intentions, or even people. Torrengo (2017) recently argued for an externalist view of social entities. On his view institutional facts and entities are not always grounded in collective belief, acceptance, or intentions, rather institutions and institutional statutes, duties, and rights "are characterized by [and grounded in] the tendency to defer to elements that are external to the content of collective intentions—such as laws, declarations, and contracts" (2017: 67). I agree that not all social entities and not all social facts are characterized by or grounded in ways that a largely internalist view of the social requires. Moreover, the argument I have given shows that externalist views of the social can be extended beyond social entities to at least one social phenomenon.

Whether the preceding argument supports a rejection of the continuity thesis depends on what is required for cooperation in pairs or small groups. Here I have argued that members of organized groups can cooperate without shared mental states, *we*-intentions, or joint commitments. I have not, however, argued for a view of what is required for cooperation among individuals who are not members of organized groups. So, the status of the continuity thesis is left open and two possibilities arise. If an account like Bratman's, Gilbert's, or Searle's is correct for cooperation at a small scale, then the preceding argument does require a rejection of the continuity thesis. A more internalist account relying on collective intentions would explain cooperation in pairs or small groups. A more externalist account that relies on playing roles in an organizational structure would explain other cases of (minimal) cooperation. This would rejection of the continuity thesis and a methodology that seeks a continuous and single theory for cooperation in all cases.

Alternatively, one might take the preceding discussion to provide the beginning of a framework for a theory of cooperation that upholds the continuity thesis. On this view, cooperation would always involve playing roles in some structure or other—whether it is the structure of a legislative body, that of a family unit, or that involved when two people attempt to move a piano together. Such a view would be very different from the sorts of views that have dominated the literature on cooperation and collective intentionality. It would uphold the continuity thesis at least to the extent that roles would be relied on in all cases of cooperation. It is plausible,

however, that any viable version of a role-based view of cooperation would also posit some discontinuity in the mental requirements for cooperation in small-scale one-off interaction and for large-scale long term cooperation in organized groups.

6.5 Concluding Remarks

I have argued that there is a form of minimal cooperation among individuals who are members of organized social groups that does not depend on shared mental states, special *we*-mode intentions, or joint commitments. Roles in group organizations can be defined so that playing a role and having a common goal involves cooperating with others. The account I have given is not a general account of cooperation. In its present form it applies only to cases in which individuals play roles in an organized group structure (i.e., to members of organized groups). Yet, at least in these cases, externalism about the social holds for phenomena like (minimal) cooperation.

Social phenomena and social ontology have largely been understood in internalist terms. The roles that representations and mental states play in the creation of the social world are interesting and need to be better understood. Yet, they should not be overemphasized. As organizations become complex and institutions are reified, representations and mental states matter less. Even if internalist mechanisms are needed to build some foundational aspects of the social world, the internalist ladder can be "knocked away" once complex roles and institutions have been created. As we extend our focus beyond pairs of individuals or small groups to complex institutionalized groups and contexts, the social—even social phenomena like cooperation—can and should be theorized in ways that are more external.

Acknowledgements This paper has significantly improved through conversations and comments on previous versions. In particular I thank Derek Anderson, Kent Bach, Barrett Emerick, Jeremy Evans, Nick Leonard, Rebecca Mason, Matt Moss, Matthew Rachar, Jules Salome, Tuomo Tiisala, the members of an audience at a colloquium at the University of San Francisco, and an anonymous commentator.

References

Bar-On, D. (1995). Meaning reconstructed: Grice and the naturalizing of semantics. *Pacific Philosophical Quarterly, 76*(2), 83–116.

Bratman, M. E. (1992). Shared cooperative activity. *Philosophical Review, 101*(2), 327–341.

Bratman, M. E. (1999). *Faces of intention: Selected essays on intention and agency*. Cambridge: Cambridge University Press.

Bratman, M. E. (2014). *Shared agency: A planning theory of acting together*. New York: Oxford University Press.

Epstein, B. (2015). *The ant trap: Rebuilding the foundations of the social sciences*. Oxford: Oxford University Press.

Epstein, B. (forthcoming). "What are social groups? Their metaphysics and how to classify them." *Synthese*: 1–34.
Gilbert, M. (1989). *On social facts*. London: Routledge.
Gilbert, M. (2006). *A theory of political obligation: Membership, commitment, and the bonds of society*. Oxford: Oxford University Press.
Guala, F. (2010). Infallibilism and human kinds. *Philosophy of the Social Sciences, 40*, 244–264.
Hacking, I. (1999). *The social construction of what?* Cambridge, MA: Harvard University Press.
Khalidi, M. A. (2015). Three kinds of social kinds. *Philosophy and Phenomenological Research, 90*(1), 96–112.
Koslicki, K. (2008). *The structure of objects*. Oxford: Oxford University Press.
Mallon, R. (2016). *The construction of human kinds*. Oxford: Oxford University Press.
Miller, S. (2001). *Social action: A teleological account*. Cambridge, UK: Cambridge University Press.
Miller, S. (2014). Social institutions. In Zalta, EN (Ed.).*The Stanford encyclopedia of philosophy*. Winter 2014. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2014/entries/social-institutions/
Paternotte, C. (2014). Minimal cooperation. *Philosophy of the Social Sciences, 44*(1), 45–73.
Ritchie, K. (2013). What are groups? *Philosophical Studies, 166*(2), 257–272.
Ritchie, K. (2015). The metaphysics of social groups. *Philosophy Compass, 10*(5), 310–321.
Ritchie, K. (forthcoming). Social structures and the ontology of social groups. *Philosophy and Phenomenological Research.*
Ruben, D.-H. (1985). *The metaphysics of the social world*. London: Routledge & Kegan Paul.
Searle, J. (1990). Collective intentions and actions. In P. R. C. J. Morgan & M. Pollack (Eds.), *Intentions in communication* (pp. 401–415). Cambridge, MA: MIT Press.
Searle, J. (1995). *The construction of social reality*. New York: Free Press.
Searle, J. (2010). *Making the social world: The structure of human civilization*. Oxford: Oxford University Press.
Thomasson, A. (2009). Social Entities. In R. Le Poidevin, P. Simons, A. McGonigal, & R. Cameron (Eds.), *Routledge companion to metaphysics* (pp. 545–554). London: Routledge.
Tollefsen, D. P. (2015). *Groups as agents*. Malden: Polity.
Tomasello, M. (2009). *Why we cooperate*. Cambridge, MA: The MIT Press.
Torrengo, G. (2017). Institutional externalism. *Philosophy of the Social Sciences, 47*(1), 67–85.
Tuomela, R. (2006). Joint intention, we-mode and I-mode. *Midwest Studies in Philosophy, 30*(1), 35–58.

Chapter 7
Towards a Blueprint for a Social Animal

Stephen Butterfill and Elisabeth Pacherie

Abstract In this chapter, we attempt to answer the question, By what steps could members of a group capable of acting together with a purpose, coordinating flexibly, communicating cooperatively and deceiving competitors be constructed from creatures with minimal social skills and cognitive abilities? The method we use is creature construction: the idea is to adopt the perspective of a designer tasked with specifying a sequence of creatures, where each is independently viable and has the capacities of its predecessors together with some new capacity which enables it to overcome limits its predecessors faced. In creature construction, the aim is not to characterise actual species, nor to describe actual evolutionary or developmental processes. Instead the aims are to understand how various forms (or prototypes) of joint action are related to, and diverge from, each other; and to identify limits on what can be achieved with a given set of cognitive and social skills.

We start with Alphonso and his kin, whose social cognition is limited to tracking the goals of others' actions. We show that despite little cognitive sophistication, the salience and triangulation heuristics enables them to initiate simple joint actions requiring coordination. One group of their descendants, Beki and her kin, develop abilities to produce pointing gestures and object-directed vocalisations, that enable them to enlist others not yet as partners but as social tools, thus extending the range of situations in which they can rely on the salience and triangulation heuristics. Another group of Alphonso's descendants, Bemi's kin, learn the art of strategic deception, acquiring increasingly elaborate tactics for manipulating others' action possibilities. This advantages them in competition. Finally, the Kimi, who are mixed descendants of both the Beki and the Bemi, inherit the former's communicative abilities and the latter's abilities for tactical deception. Progressively integrating the two allows them to develop new capacities of selective deception.

We argue that although our creatures do not yet have all the cognitive capacities classical accounts imply are needed for joint action, they have proxies for some of these capacities. These proxies allow them to coordinate in a limited but useful

S. Butterfill (✉)
Department of Philosophy, Social Sciences Building, University of Warwick, Coventry, UK
e-mail: s.butterfill@warwick.ac.uk

E. Pacherie
Institut Jean Nicod, ENS, EHESS, CNRS, Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France

A. Fiebich (ed.), *Minimal Cooperation and Shared Agency*, Studies in the Philosophy of Sociality 11, https://doi.org/10.1007/978-3-030-29783-1_7

range of ordinary circumstances. Further, relying on such proxies provide ways of avoiding both omni-doxasticity and omni-intentionality when acting together.

Keywords Shared intentionality · Creature construction · Goal tracking · Communication · Tactical deception

7.1 Introduction

Human social and societal life is built on thoughts, intentions, motivations and feelings that bind us and our actions together. One essential aspect of this is joint action. Despite its foundational role in all social and cultural life, joint action has only recently become a topic of inquiry in philosophy and the cognitive sciences (for example, see Bratman 2014; Gilbert 2013; Tomasello 2009; Rakoczy 2017). A number of classical accounts of joint action make substantive demands on cognitive abilities (as has been argued by Butterfill 2012; Pacherie 2011; Pacherie 2013; Pacherie and Dokic 2006; Tollefsen 2005; Obhi and Cross 2016). By contrast, we pursue a minimalist agenda with the aim of isolating some undemanding forms of joint action (or prototypes for joint action) and their more modest cognitive requirements.

Our method is creature construction, which has a distinguished history in philosophy (Bratman 2000, 2014; Grice 1974; Strawson 1959). The idea is to adopt the perspective of a designer tasked with specifying a sequence of creatures, where each is independently viable and has the capacities of its predecessors together with some new capacity which enables it to overcome limits its predecessors faced. As Grice described it,

> The method [...] is to construct (in imagination, of course) [...] a sequence of types of creature, to serve as [...] models for actual creatures. [...] The general idea is to develop sequentially the psychological theory for different brands [...] (which of course is unlikely ever to be more than partial). (Grice 1974, 37)

Whereas Grice's project was 'to compare what one thus generates with the psychological concepts we apply to suitably related actual creature', our aims here are more modest. We will not propose that any of our constructed creatures resembles actual individuals very closely. Our claim is merely that the creatures we construct exhibit possible cognitive profiles. These fictional creatures are, however, loosely inspired by discoveries about developmental, comparative and cognitive psychology, as our footnotes reveal.

We will use the method of creature construction to attempt to answer the question, By what steps could members of a group capable of acting together with a

purpose, coordinating flexibly, communicating cooperatively and deceiving competitors be constructed from creatures with minimal social skills and cognitive abilities? The aims are to understand how various forms (or prototypes) of joint action are related to, and diverge from, each other; and to identify limits on what can be achieved with a given set of cognitive and social skills.

We take for granted in what follows that the creatures we are constructing have some ability to track cause and effect. This is, after all, a blueprint for a social animal.

7.2 Alphonso's Kin

Consider first Alphonso and his kin, whose social cognition is limited to tracking the goals of others' actions. This enables them to distinguish between, say, grasping and pushing actions. Their goal tracking is pure in the sense that it does not depend on representing intentions or other mental states (see Gergely et al. 1995; Csibra et al. 2003 for one account of how this is possible). What joint actions are they capable of?

Suppose Alphonso is walking in the mountains when he comes across one of his kin pushing against a small boulder that is blocking their paths. There is no way around the boulder: the only option is to move it. But it is manifestly too heavy for any individual to move. Alphonso joins in pushing the boulder. At first they are both pushing as hard as they can but it doesn't move much. Gradually they fall into a rhythm of pushing and releasing simultaneously, rocking it harder and harder until eventually it gives way altogether.

Moving the boulder counts as a joint action in at least a minimal sense. Minimally, a joint action is an event grounded[1] by two or more agents' actions (compare Ludwig 2007). This is a very broad notion of joint action. Whenever two or more agents' actions have a common effect and there is an event comprising the actions and their common effect, the actions will ground this event. Because Alphonso's and his kin's rocking actions have a common effect, namely moving the boulder, this is sufficient for their actions to ground a joint action.[2]

That Alphonso and his kin perform actions directed to a single goal is not entirely accidental because the environment has provided a single obstacle which they must overcome. But their both simply pushing is not enough: to move the boulder, coordinated action is required. How could their actions be coordinated? For Alphonso and his kin, coordination is nonintentional. In this case, the actions of Alphonso and

[1] Events D1, ... Dn *ground* E just if: D1, ... Dn and E occur; D1, ... Dn are each part of E; and every event that is a part of E but does not overlap D1, ... Dn is caused by some or all of D1, ... Dn. (This is a generalisation of the notion specified by Pietroski 1998.)

[2] Joint action of this kind is found in nonhuman primates (e.g. Suchak et al. 2016; Visco-Comandini et al. 2015) and human infants (e.g. Rakoczy 2017).

his kin could be coordinated thanks to a combination of two things: the joint affordance a large, barely movable rock presents them[3]; and entrainment, the process of synchronizing two or more rhythmic behaviours with respect to phase.[4]

Because their actions are coordinated in this way, there is a sense in which the joint action performed by Alphonso and his kin is purposive. Their actions are directed to moving the boulder. Importantly, because they are coordinated in such a way as to make the boulder's moving more likely, this is not just a matter of each agent's action being directed to moving the boulder. Rather their actions are collectively directed to this goal: it is thereby a joint goal *in the following minimal sense*.

A *goal* is an outcome towards which an action is directed (so not a mental state). And the actions which comprise a joint action can be *collectively directed to an outcome* in this sense: there is an outcome to which these actions are directed and this is not, or not only, a matter of each action being individually directed to that outcome.[5] A *joint goal* of a joint action is an outcome to which the actions comprising it are collectively directed. This sense of joint goal is broad enough to allow that the jointness may not be at the level of intentions or representations of the subject; it can instead be found at the level of the coordination mechanisms. So for Alphonso and his kin's actions to have a joint goal does not yet imply that they are aware of this, nor that the joint goal is an any sense a purpose they share in acting.

The boulder is a relatively simple case because the environment provides a single most salient goal to pursue. There is also just one most salient means to pursue this goal, and the demands on coordination are limited to the need for actions to be synchronized. But what happens when there is more than one goal to pursue, none more salient than the other? In such situations there will often be uncertainty about which goal any individual will pursue. On different trees there are two large bunches of jackfruits to harvest and efficient harvesting requires several people to participate, so that it would be futile for Alphonso and his kin to pursue different goals. How could Alphonso and his kin overcome such uncertainty, even without being aware that there is uncertainty about goals? If any of them makes a move to harvest jackfruits from one of the trees, she will be aware that some additional contribution to this action is needed (although she need not appreciate what the contribution is; compare Vesper et al. 2010). How is the additional contribution to be secured?

---

[3] A *joint affordance* is an affordance for the agents of a joint action collectively. That is, it is an affordance for these agents and this is not, or not only, a matter of its being an affordance for any of the individual agents. For evidence that joint affordances exist, (see M. Richardson et al. 2007; Davis et al. 2010; Doerrfeld et al. 2012).

[4] Entrainment is found in many species' behaviours (e.g. Backwell et al. 1998). In humans, entrainment can occur automatically—that is, independently of the subject's tasks and motivations (e.g. Varlet et al. 2015)—and without awareness (e.g. Richardson et al. 2005). For a review of emergent forms of coordination in joint action, (see Knoblich et al. 2011).

[5] Given how we have defined joint action (in terms of two or more agent's actions grounding a common effect), it follows that not all joint actions have joint goals. This is because for an outcome to be a joint goal, it is not sufficient that it be a common effect of two agent's actions. Rather, the actions must be collectively directed to this outcome in the above sense.

Suppose Alphonso goes ahead unilaterally and attempts to start harvesting jackfruits as if on the assumption that the additional contribution will be forthcoming.[6] When he acts as if on this assumption, others can detect the goal of his actions. Providing they do so, his acting on this assumption makes the goal to which his actions are directed more salient or more attractive to others. This in turn makes it more likely that the assumption will turn out to be true, which can mean that acting on the assumption is reasonable.

Here, then, is a simple strategy that has the effect, not always but often enough, of enabling coordination when there is no single most salient goal to pursue (for example when several bunches of jackfruits could be harvested):

1. Pick a goal to pursue.
2. If it is not too costly to end up acting alone, go ahead and act as if on the assumption that any necessary additional contribution will be forthcoming.
3. If this turns out not to be the case, change your objectives.

A similar strategy is used by Alphonso and his kin when there are complementary roles and it is uncertain who is to do what. For example, to harvest the jackfruits, one agent needs to climb the tree and break the stems while the other should stand below and catch the falling bunches. Alphonso proceeds by selecting and performing a role, as if on the assumption that the other will perform the complementary role. Where this turns out not to be the case, he will eventually change role or objectives.

However, there are situations in which it is too dangerous or otherwise too costly to act as if on the assumption that others' contributions will be forthcoming. To illustrate, suppose Alphonso and his kin have dispersed in the woods to forage for mushrooms when a large pig comes by. This is an opportunity for them to hunt the pig, but their actions will need to be tightly synchronized as tackling the animal alone would be too dangerous. If only they had common knowledge, they would be able to safely rely on the assumption that each will join pursuit of the pig. But Alfonso and his kin do not attribute, and are not aware of, knowledge states. So they do not have common knowledge. To avoid both missing an opportunity and risking disaster, they need a proxy[7] for common knowledge, one that requires no awareness of knowledge states.

---

[6] We write 'as if on the assumption' because Alphonso need not actually assume that the additional contribution will be forthcoming. Rather, his actions may be produced in such a way that they often rely for their success on others' contributions without Alphonso himself having any view on the matter. For comparison, an agent's actions can be produced in such a way that they rely on unsupported objects falling without the agent herself having any view on this. Note that this is possible even when there are significant exceptions (not all unsupported objects fall when underwater); compare Perry (1993, 202).

[7] To say that A is a *proxy* for B is to say that, at least within a useful if limited range of circumstances, A can fulfil some of the functions of which B would fulfil were B present in those circumstances. Where there are two proxies for B, one may be a *better proxy* than the other insofar as it can serve more of the functions associated with B, or insofar as it can serve functions associated with B in a broader range of circumstances.

One crude proxy for common knowledge involves stimuli such as loud noises or strong smells. Suppose the pig is making a loud and distinctive noise; or, more generally, that Alphonso and his kin are in a position to perform actions directed to goals which specify an object or event associated with a loud noise or strong smell. Suppose further that the noise or smell is salient enough that there could be little doubt that everyone nearby had picked it up, so that its existence would ensure common knowledge among people who were capable of having common knowledge.[8] Where such salient stimuli identify the pig, the risk to Alphonso of relying on the assumption that his kin's goal will be the goal of hunting the pig is reduced. So Alphonso and his kin can use salience as a proxy for common knowledge. This is the *salience heuristic*: where they would not ordinarily rely on the assumption that additional necessary contributions will be forthcoming, Alphonso and his kin will rely on this assumption where an object or event associated with the target of potential actions is both sufficiently salient and sufficiently more salient than any other object or event.[9]

Relying exclusively on salience in this way would stringently limit the range of situations in which Alphonso and his kin can perform joint actions. There will be many cases in which an event is not very salient but others appear likely to be affected by it. For example, they may exhibit characteristic responses to it, such as a startled response or a certain twitching of a nose; or they may have a line of sight to it.[10] Suppose Alphonso, observing the pig, can observe that he and the pig are each linked to his kin, where this link is a matter of the thing (himself or the pig) causally influencing his kin, or else it is a matter of his kin having a line of sight to the thing (to himself or to the pig). This triangular situation ensures that if Alphonso acts, there is less of a risk that additional contributions will not be forthcoming. This is the *triangulation heuristic*.[11]

Although neither is infallible, the salience and triangulation heuristics pick out situations in which, often enough, there could well be common knowledge among more sophisticated individuals. So Alphonso and his kin can use these heuristics as proxies for common knowledge.[12]

---

[8] At least some nonhumans are sensitive to what others hear (Santos et al. 2006), at least within limits (Bräuer et al. 2008).

[9] Note that it is salience itself, not a belief about, or representation of, salience, which drives the salience heuristic. This heuristic adapts Lewis' suggestion that common knowledge can be arrived at through public events (1969). He suggests an event E is the basis for common knowledge that P if (i) E is public and (ii) E indicates that P. For instance, a pig's grunting is a public event and indicates the presence of a pig and is therefore a basis for common knowledge.

[10] Line of sight calculations, in at least limited form (but perhaps along with more sophisticated forms of perspective taking), appear widespread in nonhumans. See, for example, (Kaminski et al. 2006; Bräuer et al. 2004; Bugnyar et al. 2004; Okamoto-Barth et al. 2007).

[11] The triangulation heuristic is to joint attention as the salience heuristic is to common knowledge (compare Rakoczy 2017).

[12] Our approach is similar in spirit to that of Carpenter (2009) who discusses 'common knowledge, in the sense of what is known or has been experienced together' (p. 383). Note, however, that she is committed to characterising common knowledge rather than merely a proxy for it.

These heuristics demonstrate that Alphonso and his kin can achieve a range of joint actions. And since the salience and triangulation heuristics work with larger groups as well as dyads, they can even achieve joint actions involving many individuals. But there are also limits on when they can act together, and on when they can avoid acting together. To illustrate, consider two situations. First, Alphonso and one of his kin hear a pig approaching and hide in order to ambush it. But the pig they heard turns out to be not one but two pigs. This defeats the salience heuristic, as neither pig is sufficiently more salient than the other. And they have no other way to determine which pig to attack. Second, sometimes members of Alphonso's kin come across a resource, such as a berry patch, which they would ideally exploit alone. If others are alerted, the original discoverer will get little. They have no way to prevent more dominant individuals from pilfering berries. While Alphonso and his kin are unequipped to overcome either of these problems, his descendants are more fortunate.

7.3 Beki's Kin

At this point in our construction, Alphonso's descendants divide into two groups, Beki's and Bemi's. Beki and her kin are frequently confronted with situations like the 'two pig' situation described above: situations in which there are multiple, equally salient possible goals achieving which would require joint action. In these situations they desire to act, and are both excited by a potential target of action and frustrated by their inabilities to act. But Beki and her kin develop abilities to produce pointing gestures and object-directed vocalizations. Initially, the gesture or vocalization is not a thoughtful attempt to communicate but merely an expression drawn out of them by a situation. Among the causes of their gestures and vocalizations is a combination of excitement at a potential target of action and frustration at not being able to act.

None of Beki's kin understand these gestures and vocalizations except perhaps as expressions of frustration. Despite this, on some occasions the gestures and vocalizations do function to draw others' attention to objects. They thereby have the effect, unintentionally, of extending the range of situations in which Beki and her kin can rely on the salience heuristic. For example, in the 'two pig' situation, Beki vocalizes at one of the pigs thereby making this pig more salient than the other to her kin. At this point they can rely on the salience heuristic (a proxy for common knowledge introduced above), and so capture the pig.

The gestures and vocalizations also extend the range of situations in which Beki and her kin can rely on the triangulation heuristic. To illustrate, consider the 'hidden pig' problem: there is a pig nearby which only Beki is linked to, and which is too dangerous to be tackled alone. Frustration at missing an opportunity to catch the pig triggers a vocalization directed to it. The vocalization draws Beki's kin's attention to the pig, thereby linking them to it as well. In this way, the conditions necessary for relying on the triangulation heuristic are met, and so Beki's and her kin capture the pig.

At this first stage, Beki's kin's communicative actions have limited effects because they are unintentional responses to exciting objects in frustrating situations (among other things). But over time they observe and become familiar with the causal effects of their gestures and vocalizations on their kin and on other animals around them. This allows them to produce gestures and vocalizations with the goal of bringing about their familiar effects. What was merely an expression of excitement and frustration (among other things) has become an instrumental action resembling a communicative act. This greatly expands the range of actions they can perform. Initially their goals were limited to acting on the physical environment. Now a new set of goals is open to them, namely influencing the behaviour of their kin. They can, for example, call to make another come towards them, or vocalize towards an object in order to direct another's actions to that object. This amounts to their coming to recognize each other as social tools (Warneken et al. 2012).

Over time, Beki's kin come to make and distinguish a range of gestures. Some of these center around danger. What was initially an involuntary vocal response to danger gradually becomes an action they produce with the goal of bringing about its familiar effects, so that it functions as a warning.[13]

Merely involuntary vocalizing and gesturing enabled Beki's kin to solve the 'two pig' problem sporadically, when their actions happened to make the pig more salient or linked it to their kin. But having voluntary control over their gestures and vocalizations enables them to gesture and vocalize strategically, suppressing gestures and vocalizations when these could hinder success, and using them to enlist others as social tools when they promote success. This gives them a systematic way of solving the 'two pig' problem. And, more generally, it means that when uncertainty and the costs of acting alone would otherwise prevent action, communicative actions can put them in a position to act.

7.4 Bemi's Kin

Bemi's kin occupy a region in which resources are scarce. This confronts them with challenges quite different from those facing Beki's kin, who inhabit a region with abundant resources. Competition for food becomes intense, and those who come across ripe berries must either be able to defend their find or consume it before others notice. This is especially challenging for weaker individuals: to stray too far from the group is dangerous, but to forage too close to others means retaining little of the food found.

The weaker of Bemi's kin acquire abilities for tactical deception (Byrne and Whiten 1985). When finding berries, the weaker individuals will occasionally

[13] Vervet monkeys have a range of danger calls for eagles, for pythons and, for leopards. Infant vervet monkeys are not very discriminating: they will produce the alarm call whenever they see a big bird in the sky; only later do they become able to distinguish eagles from nonthreatening big birds (Seyfarth et al. 1980).

refrain from exploiting the food source while others are around. Although such restraint will increase their chances of getting the berries for themselves, their actions need not be performed with any intention to gain an advantage. It may be, for example, that rising anxiety associated with anticipation of conflict over food immediately suppresses any inclination to feed. But at some point, the weaker among Bemi's kin learn that they can avoid having food stolen when foraging in proximity to others by delaying consumption of food for as long as possible while they are present. Now the acts (or omissions) of tactical deception are performed with an intention to avoid theft. But they are not yet performed with any insight into others' mental states.

Over time the value of freezing up or refraining from exploiting food is reduced as competitors come to associate these behaviours with the presence of food. This in turn leads to an escalation of tactical deception. The weaker among Bemi's kin begin to act as if the food was absent and walk on as others pass by.[14] And what at first was a tendency not to eat when others are around gradually becomes a tendency not to eat when others are linked to you or to the food. (As stipulated above, being linked is a matter either of having a line of sight to, or else of being causally influenced by, the food). They also become discriminating in when tactical deception is used, relying on it against stronger but not weaker competitors.

Eventually some of Bemi's kin realise that being linked to food is a precondition not only for stealing it but also for performing any action concerning it; and that what goes for food goes for any kind of object. So they appreciate that, for example, if someone is not linked to a sleeping snake she will not be able to avoid it. In this situation, one seizes the other's head and forcibly links her to the snake.

Bemi's kin take an important further step when their experience of manipulating whether competitors and partners are linked to things gradually clues them into the realisation that you need not be linked to something right now in order to act on it; in many cases, it is sufficient to have been linked to it at some point in the recent past. This enhances their abilities to prevent more dominant individuals from stealing their berries. By concealing a cache of berries from a dominant competitor regardless of whether they are currently hungry enough to steal, members of Bemi's kin can avoid theft when the competitor later becomes hungry again.

The story of Bemi's kin is one of gradually elaborating tactics for manipulating others' action possibilities. Merely involuntary freezing in the presence of competitors enabled Bemi's kin to protect some of their food discoveries from theft. Involuntary freezing was sometimes ineffective and occasionally even led to missing

[14] Compare De Waal (2016)'s description of what happened when experimenters hid grapefruits on the island where the colony of chimpanzees spend the day: 'After releasing the apes onto the island, a number of them passed over the site where we had hidden the fruits under the sand. Only a few small yellow patches were visible. Dandy, a young adult male, hardly slowed down when he ran over the place. Later in the afternoon, however, when all the apes were dozing off in the sun, he made a beeline for the spot. Without hesitation, he dug up the fruits and devoured them at his leisure, which he would never have been able to do had he stopped right when he saw them. He would have lost them to dominant group mates' (De Waal 2016 Chap. 2).

opportunities to eat, but it did provide Bemi's kin with the experiences necessary to acquire more refined tactical deception. Learning that freezing is associated with having food stolen less often, they began intentionally to delay gathering or consuming food when others were currently linked to them. This further enriched their experiences of how others' actions are associated with facts about what they are, or have recently been, linked to, allowing them to manipulate those links. This in turn created new opportunities to learn about when others' actions succeed and when they fail.

7.5 The Kimi

Climate change forces Beki's and Bemi's groups to migrate, and they end up in close proximity. In their new environment, food comes mostly from large animals. Tight action coordination involving multiple complementary roles is therefore needed for catching prey. But food is also scarce enough that a find needs to be protected from pilfering by others, which will often require strategic deception.

Success in hunting large animals requires tight coordination among a fairly large number of individuals, a division of roles, and the ability to anticipate another's complementary action in order to coordinate your own with it. Through repeating successful behaviours, the joint actions of Beki's and Bemi's kin come to follow conventional patterns in the way they unfold and in who does what. These conventional patterns resemble action scripts and function as precursors of planning. Here Beki's kin have an advantage: as they identify patterns in their past successful behaviours, they can use their communicative abilities to assign roles. But in communicating, Beki's kin often alert Bemi's kin, and so end up losing much of the food.

For their part, Bemi's kin rely on the triangulation and salience heuristics to coordinate in capturing an animal, and so rarely succeed unless they are together when they encounter an animal. The infrequency of their successes means that their joint actions involve less conventional patterns than those of Beki's kin, which further widens the gap between their hunting and that of Beki's kin. But when Bemi's kin do capture an animal, their tactical deception means that they rarely suffer from pilfering.

Both Beki's and Bemi's kin are thus struggling to survive, although for different reasons. As their societies disintegrate, members of the two groups occasionally reproduce and raise children who become skilled in both communication and tactical deception. These children belong to neither Beki's nor Bemi's group but are outcasts. As outcasts they are often thrust together, but without thereby forming a group in their own right. They are however sometimes forced to act together, as surviving alone is impossible. And in acting together they have an advantage over Beki's and Bemi's kin: they are simultaneously communicators, who enable coordination, and tactical deceivers, who can avoid pilfering. Repeated successes in acting together results in them forming stable groups. These are the Kimi.

Despite their advantages, the Kimi are initially vulnerable because their abilities to communicate are not fully integrated with their abilities for tactical deception. When a potential target appears, the Kimi will gesture to link other group members to the animal, so enabling cooperative action. In doing this they attract competitors, especially Bemi's kin, who, being much less successful at obtaining food themselves, rapidly learn to follow the Kimi and steal from them.

Eventually some of the Kimi realise that their gestures and vocalizations are drawing competitors to them at just the wrong moment. Just as their ancestors began to intentionally delay foraging, so they come to suppress gestures and vocalizations when competitors are around. Although this initially makes things better for them, Bemi's kin become so dependent on them that the new strategy only means the Kimi are rarely apart from their competitors. It is only when there is some pressing danger that Bemi's kin will not pursue the Kimi.

Through following the Kimi so closely, Bemi's kin gradually come to associate the Kimi danger calls with danger, running away whenever these calls are made. On detecting the association between danger calls and their competitors' flight, Kimi groups have the opportunity to put danger calls to a new use. Some now begin to use danger calls to scare Bemi's kin away. Initially these fake danger calls cause both Bemi's kin and the Kimi to respond as if there was danger. At this stage, they are only useful where a Kimi has found a food source and does not need to cooperate with her kin. But as more and more Kimi come to use or encounter fake danger calls, perhaps observing apparently anomalous combinations of danger calls followed by feeding behaviour, they gradually come to produce and respond to the danger calls in a more nuanced way. Meanwhile Bemi's kin, who are not so close, lack opportunities to observe the anomalous combinations and so fail to learn to differentiate genuine from tactically deceptive danger calls. As the new practice of tactical deception takes hold among the Kimi, they come to accompany danger calls with nonvocal communication about food, thereby ensuring cooperation from their group co-members nearby even while giving a danger call. The Kimi have now combined Bemi's tactical deception with Beki's communicative abilities. They can selectively deceive.

7.6 Conclusion

There in fact are no kin of Alphonso, Beki, Bemi or Kimi among us, we assume. But fiction is a philosophical tool that can help us to understand by what steps members of a group capable of acting together with a purpose, coordinating flexibly, communicating cooperatively and deceiving competitors could be constructed from creatures with minimal social skills and cognitive abilities. We started with Alphonso and his kin, whose social cognition is limited to tracking the goals of others' actions. Despite little cognitive sophistication, the salience and triangulation heuristics enables them to initiate simple joint actions requiring coordination such as gathering and hunting. But they are dependent on the environment to provide favourable

circumstances for coordination, and vulnerable to pilfering by dominant individuals. One group of their descendants, Beki's kin, are equipped to coordinate in less serendipitous circumstances. Beki and her kin develop abilities to produce pointing gestures and object-directed vocalizations. They can suppress gestures and vocalizations when these could hinder success, and use them to enlist others not yet as partners but as social tools, thus extending the range of situations in which they can rely on the salience and triangulation heuristics. Meanwhile, another group of Alphonso's descendants, Bemi's kin, learn the art of strategic deception, acquiring increasingly elaborate tactics for manipulating others' action possibilities. This advantages them in competition. Finally, the Kimi, who are mixed descendants of both the Beki and the Bemi, inherit the former's communicative abilities and the latter's abilities for tactical deception. Progressively integrating the two allows them to develop a new capacity for selective deception.

Although the creatures we have been constructing do not yet have all the cognitive capacities classical accounts imply are needed for joint action, they have proxies for some of these capacities. Alphonso's kin already have proxies for common knowledge and joint attention. These proxies allow them to coordinate in a limited but useful range of ordinary circumstances. Whether or not the things they do together are strictly speaking cases of joint action, they will at least appear very much like joint actions.

Further, having these proxies continues to be useful for the descendants of our creatures. It frees them from relying on explicit beliefs and sophisticated forms of reasoning in many ordinary situations. As Perry says, even though the descendants may have greater cognitive sophistication, their designers will want to avoid 'omnidoxasticity'. Instead, '[a] more efficient way for Mother Nature to proceed is to fit our psychology to the constant factors in our environment, and give us a capacity of belief for dealing with the rest' (Perry 1993, 202). Of course Perry's focus is an individual acting alone: for much of the time, at least, it would be unfortunate to have to rely on beliefs about gravitational forces in reaching for a glass of water, say. The proxies Alphonso, Beki, Bemi and Kimi rely on provide ways of avoiding both omnidoxasticity and omni-intentionality when acting together.

As our creatures become more complex, their social environments become more complex. There is more variability and less constancy, which makes it more likely that the limits of what can be done with the proxies will matter. Take the triangulation heuristic, for example. According to this heuristic, if you observe that another is linked to you and to the target of a potential joint action, where the joint action is mutually desirable, it is safe to perform the joint action on the assumption that the other will participate (see [sec:alphonso]). With their communicative capacities, Beki's kin are positioned to overcome many limits of this heuristic (as illustrated with the 'two pig' problem in [sec:beki]). However, their descendants, the Kimi, end up using their communicative capacities for deceptive ends. This may create a need to distinguish deceptive from sincere communication and to keep track of how frequently others deceive. Such a need might be met by the emergence of a crude

theory of mind.[15] No less importantly, as our creatures and their descendants get better at navigating social complexity, they are increasingly likely to run up against another kind of limit. For example, action scripts and flexible role assignments allow them to exploit frequent and predictable events. But the better they get at coordinating around these events, the more need they may have to rapidly change their approach. When some of the Kimi chance on a better way of tracking their prey, there is no way for the group to exploit this fortuitous discovery. Introducing the better way of tracking prey cannot be done directly through conventions and action scripts, which can change only gradually. Instead this would require some form of planning ability. As these examples suggest, our creatures' proxies for common knowledge, joint attention and the rest may therefore be drivers of development in this sense: needs for greater cognitive sophistication arise from hitting their limits.

References

Backwell, P., Jennions, M., Passmore, N., & Christy, J. (1998). Synchronized courtship in fiddler crabs. *Nature, 391*(6662), 31–32. https://doi.org/10.1038/34076.

Bratman, M. E. (2000). Valuing and the will. *Noûs, 34*(supplement 14), 249–265. https://doi.org/10.1111/0029-4624.34.s14.13.

Bratman, M. E. (2014). *Shared agency: A planning theory of acting together*. Oxford: Oxford University Press. http://0-dx.doi.org.pugwash.lib.warwick.ac.uk/10.1093/acprof:oso/9780199897933.001.0001.

Bräuer, J., Call, J., & Tomasello, M. (2004). Visual perspective taking in dogs (Canis Familiaris) in the presence of barriers. *Applied Animal Behaviour Science, 88*(3), 299–317. https://doi.org/10.1016/j.applanim.2004.03.004.

Bräuer, J., Call, J., & Tomasello, M. (2008). Chimpanzees do not take into account what others can hear in a competitive situation. *Animal Cognition, 11*(1), 175–178.

Bugnyar, T., Stöwe, M., & Heinrich, B. (2004). Ravens, Corvus Corax, follow gaze direction of humans around obstacles. *Proceedings of the Royal Society of London B: Biological Sciences, 271*(1546), 1331–1336. https://doi.org/10.1098/rspb.2004.2738.

Butterfill, S. (2012). Joint action and development. *The Philosophical Quarterly, 62*(246), 23–47.

Byrne, R. W., & Whiten, A. (1985). Tactical deception of familiar individuals in baboons (Papio Ursinus). *Animal Behaviour, 33*(2), 669–673. https://doi.org/10.1016/S0003-3472(85)80093-2.

Carpenter, M. (2009). Just how joint is joint action in infancy? *Topics in Cognitive Science, 1*(2), 380–392. https://doi.org/10.1111/j.1756-8765.2009.01026.x.

Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science, 27*(1), 111–133.

[15] The emergence of theory of mind might involve the proxies in a further way. It may be that Beki's kin come to recognise the conditions specified by the salience and triangulation heuristics as sufficient conditions for goal-directed action to occur. So it is no longer simply that the heuristics describe what they do: they themselves identify the heuristics as things which should guide their actions. This involves them taking the first steps towards becoming mindreaders. In recognising conditions specified by the heuristics as sufficient for action to occur, they begin to appreciate how which actions someone performs can be influenced by which objects she has a line of sight to and which objects are causally influencing her.

Davis, T. J., Riley, M. A., Shockley, K., & Cummins-Sebree, S. (2010). Perceiving affordances for joint actions. *Perception, 39*(12), 1624–1644. https://doi.org/10.1068/p6712.

De Waal, F. (2016). *Are we smart enough to know how smart animals are?* New York: WW Norton & Company.

Doerrfeld, A., Sebanz, N., & Shiffrar, M. (2012). Expecting to lift a box together makes the load look lighter. *Psychological Research, 76*(4), 467–475. https://doi.org/10.1007/s00426-011-0398-4.

Gergely, G., Nadasky, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*, 165–193.

Gilbert, M. P. (2013). *Joint commitment: How we make the social world*. Oxford: Oxford University Press. http://0-dx.doi.org.pugwash.lib.warwick.ac.uk/10.1093/acprof:oso/9780199970148.001.0001.

Grice, P. (1974). Method in philosophical psychology (from the banal to the bizarre). *Proceedings and Addresses of the American Philosophical Association, 68*, 23–53.

Kaminski, J., Call, J., & Tomasello, M. (2006). Goats' behaviour in a competitive food paradigm: Evidence for perspective taking? *Behaviour, 143*(November), 1341–1356. https://doi.org/10.1163/156853906778987542.

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: Theory and data. In B. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 51, pp. 59–101). San Diego: Academic.

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

Ludwig, K. (2007). Collective intentional behavior from the standpoint of semantics. *Nous, 41*(3), 355–393. https://doi.org/10.1111/j.1468-0068.2007.00652.x.

Obhi, S. S., & Cross, E. S. (2016). *Shared representations: Sensorimotor foundations of social life*. Cambridge, MA: Cambridge University Press.

Okamoto-Barth, S., Call, J., & Tomasello, M. (2007). Great apes' understanding of other individuals' line of sight. *Psychological Science, 18*(5), 462–468. https://doi.org/10.1111/j.1467-9280.2007.01922.x.

Pacherie, E. (2011). Framing joint action. *Review of Philosophy and Psychology, 2*(2), 173–192. https://doi.org/10.1007/s13164-011-0052-5.

Pacherie, E. (2013). Intentional joint agency: Shared intention lite. *Synthese, 190*(10), 1817–1839. https://doi.org/10.1007/s11229-013-0263-7.

Pacherie, E., & Dokic, J. (2006). From Mirror neurons to joint actions. *Cognitive Systems Research, 7*(2–3), 101–112.

Perry, J. (1993). *The problem of the essential indexical: And other essays*. Oxford: Oxford University Press.

Pietroski, P. M. (1998). Actions, adjuncts, and agency. *Mind, New series, 107*(425), 73–111. http://www.jstor.org/stable/2659808.

Rakoczy, H. (2017). Development of collective intentionality. In K. Ludwig & M. Jankovic (Eds.), *The Routledge handbook of collective intentionality* (pp. 407–419). London: Routledge.

Richardson, M. J., Marsh, K. L., & Schmidt, R. C. (2005). Effects of visual and verbal interaction on unintentional interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance, 31*(1), 62–79. https://doi.org/10.1037/0096-1523.31.1.62.

Richardson, M. J., Marsh, K. L., & Baron, R. M. (2007). Judging and actualizing intrapersonal and interpersonal affordances. *Journal of Experimental Psychology: Human Perception and Performance, 33*(4), 845–859.

Santos, L. R., Nissen, A. G., & Ferrugia, J. A. (2006). Rhesus monkeys, Macaca Mulatta, know what others can and cannot hear. *Animal Behaviour, 71*(5), 1175–1181. https://doi.org/10.1016/j.anbehav.2005.10.007.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Animal Behaviour, 28*(4), 1070–1094. https://doi.org/10.1016/S0003-3472(80)80097-2.

Strawson, P. F. (1959). *Individuals*. London: Meuthen.

Suchak, M., Eppley, T. M., Campbell, M. W., Feldman, R. A., Quarles, L. F., & de Waal, F. B. M. (2016). How chimpanzees cooperate in a competitive world. *Proceedings of the National Academy of Sciences, 113*(36), 10215–10220. https://doi.org/10.1073/pnas.1611826113.

Tollefsen, D. (2005). Let's pretend: Children and joint action. *Philosophy of the Social Sciences, 35*(75), 74–97.

Tomasello, M. (2009). *Why we cooperate*. Cambridge, MA: MIT Press.

Varlet, M., Bucci, C., Richardson, M. J., & Schmidt, R. C. (2015). Informational constraints on spontaneous Visuomotor entrainment. *Human Movement Science, 41*(June), 265–281. https://doi.org/10.1016/j.humov.2015.03.011.

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks, 23*(8–9), 998–1003. https://doi.org/10.1016/j.neunet.2010.06.002.

Visco-Comandini, F., Ferrari-Toniolo, S., Satta, E., Papazachariadis, O., Gupta, R., Nalbant, L. E., & Battaglia-Mayer, A. (2015). Do non-human Primates cooperate? Evidences of motor coordination during a joint action task in macaque monkeys. *Cortex, Special issue: Neuro-cognitive mechanisms of social interaction, 70*(September), 115–127. https://doi.org/10.1016/j.cortex.2015.02.006.

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. https://doi.org/10.1111/j.1467-7687.2011.01107.x.

Chapter 8
Modest Sociality, Minimal Cooperation and Natural Intersubjectivity

Michael Wilby

Abstract What is the relation between small-scale collaborative plans and the execution of those plans within interactive contexts? I argue here that joint attention has a key role in explaining how shared plans and shared intentions are executed in interactive contexts. Within singular action, attention plays the functional role of enabling intentional action to be guided by a prior intention. Within interactive joint action, it is joint attention, I argue, that plays a similar functional role of enabling the agents to act in a collaborative way such that their actions are rationally guided by a prior shared intention. This understanding of joint attention – as having a key functional role of enabling the rational guidance of joint intentional action by a prior shared intention – allows for an alternative understanding of the kind of minimal cooperation that infants can engage in. On this understanding, infants' capacity to engage in joint actions is already an incipient capacity to engage in rational and intentional joint actions, albeit a capacity that is necessarily scaffolded by an adult rational co-partner.

Keywords Modest sociality · Minimal cooperation · Michael Bratman · Natural pedagogy · Shared intentions · Collective intentionality · Joint action · Joint attention · Natural intersubjectivity.

Joint action and shared collaborative activity can take place in a multitude of ways. There are long-term institutionalised collaborations – such as the construction of the International Space Station – occurring within a hierarchical structure, containing defined roles, interchangeable agents, and taking place over vast stretches of space and time. At a more modest level, there are non-institutionalised, not necessarily hierarchically-based, small-scale collaborative plans, such as going on a family trip, or hosting a surprise birthday party. At an even more fine-grained level, there are

M. Wilby (✉)
School of Humanities and Social Sciences, Anglia Ruskin University, Cambridge, UK
e-mail: michael.wilby@anglia.ac.uk

A. Fiebich (ed.), *Minimal Cooperation and Shared Agency*, Studies in the Philosophy of Sociality 11, https://doi.org/10.1007/978-3-030-29783-1_8

on-the-hoof, face-to-face interactions, such as taking a walk together, or moving an obstacle out of the way of the road.

In a way that runs partly parallel to this, the literature on Collective Intentionality and Shared Agency has tended to focus on three families of collaborative activities:

(a) Massively Shared Agency
(b) Modest Sociality
(c) Minimal Cooperation

The first has cross-over with topics in *political philosophy*, the second with topics in *philosophy of mind,* and the third with topics in the *philosophy of cognitive science.*[1]

It is not obvious, or perhaps even plausible, to suppose that there might be one overriding framework to capture all of these and related phenomena. Yet there remain questions of how they relate to each other. On the face of it, in at least some central cases, massively shared agency would not be possible without forms of modest sociality, and modest sociality would not be possible without forms of minimal cooperation. So the different areas must interface with each other in some places and at some points.

My concern here will be with the relation between (b) and (c). In particular, rather than discussing Minimal Cooperation as an alternative type of cooperation to Modest Sociality, I want to discuss the form that Minimal Cooperation might take when viewed as a way of executing shared (prior) intentions that arise as elements of plans within Modest Sociality (Bratman 1999, 2014). Approaching Minimal Cooperation in this manner, will, I think, allow for a different emphasis on what is required of the type of cooperation that humans might engage in at the face-to-face level. This kind of project differs in emphasis from the projects that have tended to motivate accounts of Minimal Cooperation.[2] The explicit motivation for such minimal accounts has tended to point to the discrepancy between childrens' capacity for shared agency, on the one hand, and the cognitively-heavy demands that accounts of Modest Sociality put on the protagonists, on the other. That is: infants can engage in collaborative activities, but they don't seem to have the sophisticated metarepresentational capacities that theories of Modest Sociality typically require agents to have for shared agency (Tollefsen 2005; Pacherie 2011; Butterfill 2012; Fantasia et al. 2014). The apparent discrepancy has led some to argue either that (a) Minimal Cooperation is to be applied to a distinct *explanandum* to that to which Modest Sociality applies – i.e., that collaborative infants are engaging in something other than genuine shared agency; or that (b) Minimal Cooperation is an alternative

[1] See, for instance, the collection of articles in Chant et al. (2014) which cover a range of ground. The phrase 'massively shared agency' comes from Shapiro (2014).

[2] Although (see Pacherie 2013 and Tollefsen 2014), for approaches similar to the one I shall be adopting. The question of how prior intentions can 'interface' with motor representations in the individual case is a topic that is beginning to receive quite a lot of attention, see in particular Butterfill and Sinigaglia (2014), Mylopolous and Pacherie (2017), Brozzo (2017), Brownstein (2014), and Shepherd (2017).

explanans for the same phenomena that Modest Sociality is tackling – i.e., that infants do engage in genuine shared agency, but it should be modelled in an alternative way to the way in which it is modelled in the adult case (see Butterfil and Sebanz 2011 for a summary of these alternatives). There might well be room for both of these views – collaborative activities come in a range of flavours after all – and neither need be seen as competitors to accounts of Modest Sociality, which, on Bratman's view at least, is multiply realisable.

My aim here is slightly different. I want to question what kind of account of Minimal Cooperation would be required for agents who *can* engage in the more sophisticated form of Modest Sociality, in a way that makes rational contact with the plans, conceptions and schemes that such agents are involved in. I shall argue that by taking such an approach we will find that we can understand forms of minimal cooperation that are at once suitable for those agents who see those minimal collaborations as elements within larger plans (i.e. adults), and for those agents who are not yet capable of seeing their collaborations in such terms (i.e. infants and young children). This is a useful endeavour because: (a) much – although not all – of the evidence that relates to infant social capacities is in their interactions with adults; and (b) it gives a glimpse into the question of how infants might be initiated into more sophisticated normative commitments and reason-giving that comes with engagement in larger plans and activities.

This chapter will be split into three parts: Sect. 8.1 shall outline an understanding of shared intentions within Modest Sociality, based on Michael Bratman's work. Section 8.2 shall outline the notion of (what I shall call) *Natural Intersubjectivity,* which regards joint attention as having a pivotal role in coordinating and explaining small-scale interactions – of the kind that Minimal Cooperation is often used to explain – as manifestations of shared intentions. Section 8.3 shall argue that this notion of Natural Intersubjectivity is able to say something about the development of young children's sociality, enculturalisation and mindreading skills.

8.1 Modest Sociality

The term *Modest Sociality* refers to shared agency – i.e. shared intentions and their attendant cooperative actions – within small adult groups, in the absence of asymmetric, hierarchical authority relations that typically constitute social relations within institutional structures. As Bratman puts it, the focus of Modest Sociality is:

> primarily with duets and quartets rather than symphony orchestras with conductors, with small teams of builders rather than large and hierarchically structured companies, with small and informal neighbourhood groups rather than county governments, with small group discussion rather than deliberations in the US Senate, and with friendship and love rather than legally constituted marriage (Bratman 2014: 7).

Bratman distils his account of Modest Sociality[3] into the following elements, which he states are "somewhat compressed sufficient conditions for our shared intention to J" (Bratman 2014: 103)[4]:

(a) **Intention condition**: We each have intentions that we J; and we each intend that we J by way of each of our intentions that we J … and by way of relevant mutual responsiveness in sub-plan and action, and so by way of sub-plans that mesh.
(b) **Belief condition**: We each believe that if the intentions of each in favour of our J-ing persist, we will J by way of those intentions and relevant mutual responsiveness in sub-plan and action; and we each believe that there is interdependence in persistence of those intentions of each in favour of our J-ing.
(c) **Interdependence condition**: There is interdependence in persistence of the intentions of each in favour of our J-ing.
(d) **Common knowledge condition**: It is common knowledge that a–d.
(e) **Mutual responsiveness condition**: Our shared intention to J leads to our J-ing by way of public mutual responsiveness in sub-intention and action that tracks the end intended by each of the joint activity by way of intentions of each in favour of that joint activity.

(a)–(d) relate to shared intentions, typically held prior to the joint action which they motivate, while (e) relates to the execution of those shared intentions within joint action. The core element to focus on here is the *Intention Condition*. The Belief Condition, Interdependence Condition and Common Knowledge Condition serve as ways of stabilising collaborations, to ensure smooth running, trust, mutual deliberation and bargaining. I shall turn to the *Mutual Responsiveness Condition* in Sect. 8.1.2, since this is where the account makes most obvious contact with Minimal Cooperation.

Consider an example. Suppose that we intend to dance the tango together. What does such an intention involve? The account parses this intention as follows: we each have an intention that we dance the tango, and we each intend that we dance the tango by way of each of those very intentions, as well as by way of relevant meshing sub-plans and actions. What makes the shared intention to dance the tango together different from us each having our own private intention to dance with the other is that we each intend that the dance come about by way of each of our respective intentions (and our respective and meshing sub-plans and actions by which we will realise this action). If all goes to plan, we will end with a joint action that comes about by way of each of our respective intentions – and meshing sub-plans – to perform that joint action.

Bratman's approach is, as he calls it, *Constructivist*. Constructivism can be seen as a general methodological approach to understanding complex phenomena that constructs those complex phenomena out of simpler building blocks (Grice 1974). In the case of social agency, the building blocks are features of individual agency that combine to generate social phenomena. This is achieved, on Bratman's account,

[3] There are a number of rival accounts of Modest Sociality (most notably Gilbert 2009; Searle 1990; Tuomela 2007), but I shall focus here on Bratman's account at least partly because of its starting assumption of constructivism.

[4] The non-distilled accounts can be found in Bratman (1992), and then extended in *Shared Agency* (esp Chap. 4).

by putting the weight of the analysis on the *content* of the intentions. The intentions are standard, individual intentions. Within social contexts, they take on the same functional and normative load, as they would do within individual contexts. They are distinctively social only insofar as (a) two agents are involved; and (b) the content of the agents' intentions make reference to a joint activity and to the other agent's intentions with regards to that joint activity.

In order to do its Constructivist job there is already much that is contained within the *form* of an intention on Bratman's account. I shall discuss this now in a bit more detail, for, in doing so, we should be able to get a better grip on the challenges involved in working out how Bratmanian Modest Sociality can join-up with a more Minimal Coordination.

8.1.1 The Planning Theory of Intention

The central building block of Bratmanian Modest Sociality is the *Planning Theory of Intention* (Bratman 1987, 2007) that applies in the first instance to individuals. It resists the traditional idea that intentions can be understood reductively in terms of clusters of beliefs and desires. Rather, intentions are distinctive states of mind that are "typically elements within larger plans" (Bratman 1987: 28). Their functional role is to help coordinate our activities (both personal and social) in a way that compensates for our limited cognitive resources of memory, time and processing power. I shall outline some of the key feature of this theory.

Firstly, *plans are hierarchical and partial.* If one has an intention to give a talk in a far-away country, then that intention will be a core element of the plan that it generates. The plan will be *hierarchical*: the intention generates and embeds further sub-intentions (buying a ticket, booking accommodation, writing the talk), all of which can be decomposed into further embedded sub-sub-intentions. The plan is also *partia*l: if I am to give a talk at the conference, then at some point I will need to find my way from my accommodation to the lecture hall – but that part of the plan can be left to later; I don't need to fill in all the elements in advance.

Secondly, *plans are stable in ends and more flexible in means.* By intending to give a talk in a far-away country, I thereby settle on a course of action and start thinking of a partial plan of how to fulfil it. The intention is stable in that it provides a settled objective, around which one draws up one's plans. Once it is in place, then there is a presumption of inertia that one will carry out the intention. This stability of ends allows for a flexibility of means – the lower down the hierarchy one goes the more flexibility one will allow in one's sub-intentions, so as to fulfil one's more robust core intention (e.g. one might intend to take the 10.00 am flight, but if that flight is not available, then, all else being equal, one will book a later flight, rather than just abandoning one's intention to give a talk).

Thirdly, *plans are subject to norms of consistency*. Plans, and the elements of which they are composed, must be coherent, both internally and epistemically. For instance, one shouldn't intend that one give a talk at the same time that one is

intending to visit a sick relative. Nor should one intend to give a public talk at a place where one believes no-one will be. When one buys a plane ticket one should ensure that it goes to the place where one intends to give the talk, and one should make sure it goes from an airport which is most appropriate for one to reach. All of this will require an internal means-end coherence between intentions and sub-intentions; consistency across intentions and across beliefs, as well as consistency in the agglomeration of plans, intentions and beliefs.

Fourthly, *intentions are motivational.* Intentions, in addition to the structural role mentioned above, provide agents with a conative, pro-attitude towards acting in a way that fulfils their content.

As can be seen from this, the emphasis on intentions as typically being elements of plans is an emphasis on *future-directed intentions*. This opens up the question of how intentions – as elements of partial plans – are connected to *intentional actions*. In order to approach this question, I want to make use of a three-way distinction, adapted from Elisabeth Pacherie (2006, 2011).

We can distinguish between: (i) distal intentions; (ii) proximal intentions; (iii) intentions in action.

The first are future-directed intentions, with a stability of content, subject to norms of consistency, and that generate, and are elements of, small-scale partial plans. We can note three aspects of these kinds of intentions: (a) They are states of intention, rather than intentional actions; (b) they tend to specify *types of outcomes* – e.g. that I present a talk in Geneva – rather than *token actions*; (c) because they specify general *outcomes* – states of affairs – rather than specific actions, then they can naturally be described with a *'that clause'*.[5]

The second are intentions that are directed towards particular token actions or token outcomes at a particular time – e.g. that I start speaking *now*. They will tend to be sub-intentions (or sub-sub-intentions, etc.) and so will tend to be more flexible and context-driven. We can note three aspects of these kinds of intentions: (a) They are states of intention, rather than intentional actions; (b) they tend to specify particular, token actions, rather than generalised outcomes; (c) when they specify particular actions, then they are most naturally described using the grammatical construction of 'intention to'.

The third are the execution of an intention. We can note three aspects of these kinds of intentions: (a) they are intentional actions, rather than states of intention; (b) because they are actions, then they involve token actions rather than types of actions or generalised states of affairs; (c) because they are actions, then they are most naturally described using the grammatical adverb 'intentionally'.

[5] The grammatical object of an '*intention that*' takes a propositional form, rather than an agential or infinitival form (see Ferrero 2013). Consequently, an 'intention that' needn't (but could) be self-reflexive: One could intend, for instance, that a parcel be delivered (if one were a manager of a postal company, for example), without there being any suggestion that it is oneself who delivers the parcel. By contrast, an *intention to* deliver a parcel – where the grammatical object of such an intention is an action, with a hidden but implied reflexive pronoun (PRO) – implies that the action is necessarily carried out by the subject of the sentence.

This is a very rough sketch of a set of important distinctions within action theory, and there is certainly room for both dispute and development with how I have set them out. Before turning to the 'Interface Problem' of how Future-Directed Intentions relate to Intentional Actions, I want to briefly say something about how the Planning Theory relates to shared intentions.

8.1.2 Shared Intentions and the Planning Theory

The Planning Theory tells us that there is more to having a shared intention that we *J* than merely two agents having volitional mental states with interlocking contents. It tells us, rather, that the agents have a *shared plan*.

In particular, it is worth making a distinction – which can be easily blurred – between a prior intention and a plan. Intentions are *elements* of plans. The plan itself is composed of a number of intentions which relate to each other in a hierarchical and coherent manner. If we think of a plan as something like a hierarchically structured 'to-do list', then there is a sense in which agents with shared intentions (which are in fact two intentions that interlock) genuinely *share* a plan (as a singular thing). Without a sense that there is one plan in place, then the constraints of coherence and consistency will get lost. The agents in question might possess different *copies* of this plan (different vehicles of the plan exist in the different agents heads), and the copies will likely contain different details which are relevant to playing 'their part', but, if there is to be the interconnectedness that is required – of coherence, consistency and agglomeration – then there is a sense in which there is a singular plan that the agents are both consulting. (It might be that the Common Knowledge Condition in Bratman's condensed outline effectively plays the role of making this plan public, but I shan't discuss that here).

8.1.3 Shared Intentions and Mutual Responsiveness

We have established the following set of distinctions between three faces of intentions (Table 8.1):

If the above is correct, then, in the execution of a prior intention (which might be dispersed and discontinuous over time), we move from the general to the specific, from intended outcomes to executed actions, and from an intention *that* to an

Table 8.1 Three faces of intention

Future-directed	Present-directed	Intentional action
Intention that *p*	Intention to φ	φ-ing intentionally
General	Specific	Specific
Directed at outcomes	Directed at actions	Actions

intention *to*. In Bratman's account, this is mentioned only insofar as this condition obtains:

(e) **Mutual Responsiveness Condition**: Our shared intention to J leads to our J-ing by way of public mutual responsiveness in sub-intention and action that tracks the end intended by each of the joint activity by way of intentions of each in favour of that joint activity.

Although this doesn't tell us how the aspects are interfaced, it is useful to consider, since it contains within it all three faces of (shared) intention: the shared future-directed intention; the shared present-directed intention; and the shared intentional activity (Tollefsen 2014). That is to say, 'our shared intention to J' can be understood as a shared future-directed intention. Our 'sub-intentions' can be understood as being – depending on context – either future-directed sub-intentions, or present-directed intentions, and 'our J-ing' can be understood as a shared intentional activity.

So, we can parse the Mutual Responsiveness Condition as follows: our each intending that p (and our each intending that p by way of both those intentions and by way of meshing present-directed intentions and meshing actions, etc.), leads to our jointly intentionally φ-ing, by way of the 'public mutual responsiveness' of each of our respective future-directed sub-intentions that q and our present-directed intentions to φ that track the future-directed intentions that p.

For instance, suppose that we intend that we dance the tango for the next dance, and that this is common knowledge, in that it is something we have openly agreed to do. How is this shared intention executed? According to this account, we each intend the following: <that we dance the tango by way of each of our intentions to do so, as well as by our meshing future-directed sub-plans, our meshing present-directed sub-intentions, and our meshing actions>. This is the future-directed shared intention. The claim then is that this future-directed intention is executed because it 'leads to' our intentionally dancing *this very dance we are dancing together now*, and it led to this by way of a 'public mutual responsiveness' in both our future-directed sub-plans (that I hold out my hand, and that you stand up), and our present-directed intentions to dance this very dance we are dancing (the lead dancer intentionally moving as so, and the follower moving as thus), where these present-directed intentions (as well as the future-directed sub-intentions) track the initial future-directed intention (the initial agreement) that we dance the tango together.

The account claims that future-directed intentions 'lead to' intentional actions *by way of* present-directed intentions that *track* the future-directed intentions. So, the relation between a future-directed intention and an intentional action is not purely causal. One way to understand how this tracking exhibits itself in the linguistic expression of the intentional action is as an instance of what GEM Anscombe (1957) calls 'the intention with which a thing is done'. So: an intentional action to φ is done with the intention that p.[6] That linguistically exhibits, but does not yet explain, the relation that holds between a future-directed intention and an intentional action.

[6] Where there is a future-directed sub-intention that q involved, one might just as truly describe it as an intention to φ with the intention that q. For instance, if we have a future-directed intention

To conclude this section: I have argued that a joint intentional action can, in some cases, be understood in terms of being an executed element of part of a shared prior plan. When it is understood in this way, we can describe the action as <jointly φ-ing with the intention that p>, where p is the content of the governing intention of the plan (and/or sub-plan). For instance, we might be dancing this very dance (a token instance of the tango) with the sub-intention that we dance the tango, and the fuller prior intention that our guests have an enjoyable evening. Here we have a prior shared intention (that our guests have an enjoyable evening) that leads to the joint intentional action, which is carried out by way of an intention to dance this very dance in a way that tracks both the sub-intention and the fuller prior intention. In a hierarchical way, both the sub-intention and, higher up the hierarchy, the prior intention, continue to shape and direct the execution of the action.[7]

8.2 Natural Intersubjectivity

How are prior shared intentions able to shape and direct the execution of a close-encounter joint action such that the action normatively tracks that prior shared intention? The claim that I shall pursue here is that it is via *joint attention* (or, at least, that joint attention has a pivotal role to play). Joint attention, I argue, functions to coordinate multiple agent's actions in a way that allows for the rational execution of a prior shared intention.[8] This coordination allows for the agents' prior shared intention to guide the joint intentional action in a way that is under the agents' reliable conscious control. By jointly attending to the object, event or activity, the agents are able to monitor, control and respond to the ongoing activity, including the roles that each agent is playing as well as the salient features of the situation – such as the objects of their attention – in a way that normatively tracks, because rationally guided by, the shared prior intention.

This conception of a key functional role of attention as a matter of *conscious rational guidance of action* is related to various programmes within the Philosophy

that we hold a party, that might involve sub-plans that we decorate the room we hire in a colourful way. This in turn might lead to the jointly intentionally hanging *this* decoration *here*. This action is done intentionally. And we can describe it as being doing with the intention that we hold a party. And/or we can describe it as done with the intention that we decorate the room colourfully.

[7] Recent discussion of the relation between prior intentions and intentions-in-action has centred around a two-tier framework, in which a more primitive minimalist system runs partly independently of a slower, more purposeful, cognitive system that is conscious and propositional. This discussion has involved, inter alia, an account of how primitive agents (animals or infants) might be able to engage in joint actions via shared goals in the absence of the former (Butterfill 2012, 2013); and how the more complex system-2 might interface with the more primitive system-1 (Butterfill and Sinigaglia 2014; Sinigaglia and Butterfill 2015). See Christensen and Michael (2016) for a critical discussion of this proposal.

[8] Fiebich and Gallagher (2013) argue for the related idea that Joint Attention has the functional role of "[reducing] the risk of shared intention failing" (2013: 586).

of Attention (Campbell 2002; Wu 2011; Smithies 2011; Watzl 2017). I shall begin with a discussion of the relation between attention and action in the individual case, and then turn to a discussion of how this plays out in the joint case.

8.2.1 Attention and Rational Guidance

Attention is not sufficient for attentive action. Merely attending to an object or a scene is not going to necessitate acting on it (although it might raise urges or suggest specific actions). If attention is to have the coordinating and guiding role that I am claiming of it, then there needs to be something additional to perceptual attention involved; there must also be a conative aspect that combines with the cognitive aspect of attention. The additional conative aspect, I suggest, comes in through the distal, governing prior intention – when one is acting attentively, then one's actions are under the normative *guidance* of one's higher-order intentions/plans by means of one's attention, where such intentions have a motivational force that carries through into the executed actions. By extension – in a manner that I shall develop in Sect. 8.2.3 – joint intentional action comes to be under the normative guidance of prior planned shared intentions by means of joint attention.

There are three aspects of the concept of guidance which are of relevance here.

Firstly, guidance is something achieved by a person rather than by a sub-process, or "independent causal mechanism", such as a motor representation (Frankfurt 1978). On the assumption that higher-order plans have "agential authority" that amounts to a form of 'self-governance' (Bratman 2007), then it is these higher-order plans that are doing the guiding.

Secondly, guidance is not the same as merely *causally effecting an outcome*. Guidance involves a personal level awareness of what one is doing and needs to do; how one is to do it; and of the circumstances that one is in, that would allow one to intervene, correct and control one's actions as and when necessary in order to fulfil one's plans and intentions. Guidance will involve exercising one's conscious control at some moments, submitting to the unreflective, skilled 'flow' (Csíkszentmihályi 1975; 1990; Schear 2013) of action at others, and consciously intervening in or correcting those actions at still others. When one is unreflectively engaged in the flow of a skilled activity – as a well-practised pianist might be – then the activity remains under the guidance of one's plans and intentions, such as the intention to play a particular piece at a certain tempo (Velleman 2008). The skilled pianist will be ready to intervene and correct slips, and be ready to take more active control at certain points of certain sequences of fine-grained motor control that might at other times be performed automatically (see Brozzo 2017).

Attention is a capacity by which one can make rationally informed decisions about how or when to control, intervene, correct, or 'go with the flow' in pursuit of a goal or plan: it is by means of attention that one is able to determine what skills to put into action and when, and to which objects or locations one should direct one's

skills towards. If one was not attending to the scene that one was in – the activity itself as well as the surroundings – then one would have immense difficulty exercising one's capacities successfully. One's attention can be directed in a 'top-down' manner, as when one is guiding one's actions in a controlled and deliberative way; or it can grabbed in a 'bottom-up' manner by unexpected or unanticipated events in the environment, allowing one to intervene or correct when one is acting 'automatically'.[9] Such unexpected events might require one to change or abort one's more flexible sub-intentions in pursuit of a more robust distal intention, or even to unexpectedly achieve one's intentions and plans by capitalising on a lucky accident. The notion of 'guidance' allows for unanticipated interruptions – events that grab or disrupt one's attention – to nevertheless have a place within a rational and planned intentional schema.

Thirdly, guidance sees an intentional action as an actualisation of a prior intention. When one's intentional actions are carried out under the rational guidance of one's prior intention, then we can say that the action is carried out with the intention that *p*, where *p* is the content of the prior intention.[10] When one is acting intentionally with the intention that *p*, then there is a sense in which the planned intention is operative *in* the action, rather than being a mere causal precursor to the action. The planned intention, in having 'agential 'self-governing' authority', has guidance control of a range of one's sub-systems and bodily movements. Intentional actions are thereby "exercises of conceptual capacities", to use McDowell's (2009) phrase.[11]

8.2.2 Attention and Demonstrative Reference

I have suggested that attention plays a role in exercising the rational guidance of intentional action by a prior intention. As we shall discuss in the next section, in the case of joint attention, we can say that joint attention takes the role of *rationally coordinating* joint intentional actions such that they are exercises of, because rationally guided by, a shared prior intention. The question I want to discuss first, in this section, is how attention *simpliciter* realises the broad functional role I have sketched.

A key functional role of attention, then, is to enable intentional actions such that they are rationally guided by a higher-order intention and plan. How does it do this? How is this role *realised*? Attention can be seen to be a form of demonstrative

[9] Or one can make use of a combination of these and guide one's attention in a 'top-down' manner in order to find something that will grab one's attention and 'pop-out' from the scene.

[10] And the action can be intentional under more than one description (Davidson 1963), depending on the plan/sub-plan that we choose to highlight.

[11] Luthra (2016) calls this kind of view *Rationalism about agential control*: whereby "our ability to determine, as agents, how we act consists solely in guidance of our actions through the exercise of those sorts of rational capacities—capacities for intention, practical judgment, practical reasoning, and the like" (Luthra 2016: 2272)

reference (Campbell 2002; Dickie 2015) – a way of specifying particular objects, processes, features, locations or events within one's perceptual field, that makes them accessible to thought; 'this thing over here' and 'that thing over there'. In doing this, attention can act as a kind of interface between world, thought and action (Campbell 2002, Ch 7). As mentioned earlier, prior intentions will, by their nature, be more general than the specific actions that actualise those intentions. Attention allows the general to be linked to the particular by identifying particular objects, processes, features, locations or events that can act as determinants of the general concepts that feature within elements of plans and sub-plans (see Evans 1982, and the notion of a fundamental ground of thought, for an outline of how this connection might be made).

Suppose, for instance, that one intends that one have baked beans for dinner tonight. One needs, in that case, to be able to identify a can of baked beans when one sees one. Further, the sub-plans, as and when they come about and get filled in, will involve further general concepts that need to be identified and realised by particulars in one's environment (pans, spoons, plates, etc.), as well as particular actions (reaching, grabbing, etc.) and spatial relations (to the right, up the top, to the left, etc.). This requires the agent to be able to have demonstrative thoughts about these objects, relations and actions, in a way that will link up with the plans. Those demonstrative thoughts are grounded in attention – it is by means of attention that one can identify *that object* as a can of baked beans, and *that object* as a pan, and then, when it comes to action, calculate the distance that needs to be traversed between them (*that* distance), and then perhaps, by means of indexical predicates, identify the needed movement and actions: I should do *that a*ction (Heal 1997; Butterfill and Sinigaglia 2014).[12]

An intention that specifies, for instance, <that I cook baked beans tonight>, gets actualised partly by means of a demonstrative concept grounded in attention that specifies <that *those* are baked beans>. One can then insert that specification into the generalised plan, so as to intend the more specific plan <that I cook *those* beans>, eventuating, via numerous further, finer-grained sub-plans, involving pots and pans, in an intention to cook the beans, and the very act of intentionally cooking the beans. On a hierarchical planning theory, such intentions will be filled out with particulars – demonstratives, indexical predicates, etc. – all the way down, as sub-plans are filled in. At each stage the plans and sub-plans are guiding the actions by means of the attentive reference to places, locations, features, etc., that fulfil the plans. The demonstratives have the role of allowing the agent to insert objects, actions, properties and locations into a propositional level, planning schema.

To summarise this and the preceding section: I have argued that attention plays the role of enabling the rational, guided execution of intentional actions by prior intentions and prior plans. It plays this role partly by means of demonstratively identifying the specific features of the environment – objects, locations, processes,

[12] Although much of the philosophical and psychological literature on attention has been concerned with object attention, I am assuming here that one can likewise attend to processes and actions and events.

actions – that can act as particular instantiations of the elements that figure in prior plans in only a general way. One then acts on those objects, locations, processes and actions in accordance with what would be required for the fulfilment of the plans, in a way that is guided by those plans, and not merely caused by them. On this picture, motor representations play a role as the sub-personal mechanisms by which this guidance is carried out, but not as intentions or even intention substitutes (cf, Butterfill and Sinigaglia 2014).

8.2.3 Joint Attention and Coordination

The above sketches a role for attention within action, whereby attention is the means by which plans and sub-plans can guide intentional actions. I now want to argue that *joint attention* plays a corresponding role in interactive cases of joint agency, where part of that corresponding role is to act as a means of *rationally coordinating* joint action in accordance with, because guided by, a prior shared intention.

It is important here to fix the kind of situations I am talking about. There are of course ways of executing shared intentions which don't require interaction. If we are searching for a lost dog, for instance, and you go north and I go south, then there is no interaction, even though there is a shared intention that the dog be found by one of us, and various sub-intentions to be executed in accordance with that plan. My concern, rather, is with interactional contexts with coordinated actions. These will be *cooperatively loaded* actions – actions where coordination (and not merely 'meshing') is essential to their fulfilment.

'Joint attention', as I shall be using the term, refers to a situation where two or more agents are both attending to an event, object, process, feature or location, and *they are aware of this whole awareness,*[13] which is to say that their attention, and its common content, is 'out in the open' and 'mutually manifest' to each.[14] Attention (and by extension joint attention) here is to be understood not as a full-blown propositional attitude, with a 'that'-clause and an independently assessable truth-value, but rather as an unsaturated demonstrative element that can figure in a propositional attitude, such as a demonstrative singular thought, that can, in turn, act as the realisation of an element of a plan. For instance, attention to an object – a can of beans, say – can give specific content to a general intention <that one cook beans for dinner tonight>, by specifying the particular can of beans that can fulfil that intention.

Now, the claim at hand is that joint attention has the role of coordinating an interactive joint action in accordance with, because rational guided by, a shared prior

[13] This way of phrasing it comes from Peacocke (2005).

[14] This understanding of joint attention is a short-hand for a number of alternative ways of filling out the idea that might take a truncated, iterative, relational or fixed-point form (see Barwise 1988). I shall leave it unstated what form it should take, since any attempt to do so would require a full-length paper of its own, although my sympathies are with the relational approach (Campbell 2005; Eilan 2007; Seemann 2012; Wilby 2010).

intention. Suppose, to take John Searle's (1990) well known example, we intend that we make a Hollandaise Sauce together. We have a plan. We know how to carry it out. We know the ingredients needed and the recipe to follow. And we have an open agreement that we will make it this afternoon. When the time comes we will need to identify the objects, processes, angles, locations, etc., that will be necessary to carrying out this prior shared intention. We will need to coordinate on the details – we will need to work out who is pouring, who is stirring, where the bowl is, the timing, etc. What we attend to will help determine what we use and how we use it in a way that tracks the shared prior intention. That we are both attentive to the elements that comprise the situation we are in is clearly central to linking our activities and our plans. It is because we are both attending to the unfolding activities – by means of attending to the objects, locations, relations, processes and features of those activities – that those activities can be understood as *realisations* of our shared planning. And it is by understanding those activities as realisations of our shared planning that allows for the shared planning to rationally guide the ongoing activities. By moving the pan just so, and cracking the egg just so, we can, providing we have a hierarchical plan in place with sub-plans and sub-sub-plans and sub-sub-sub plans (etc.), place those very activities into that hierarchical structure. Our respective attention to the ongoing situation and environment allows us to coordinate our actions so that those actions – both before and after execution – can be recognised as elements of a plan that we are executing together, and that we must track and update together.

One might object at this point that *joint* attention – in the full-blown 'mutually manifest' sense – is not needed. Each agent needs to fulfil their own part of the joint action, and so needs to attentively monitor what is necessary for playing that part. Perhaps that might mean monitoring the other's *actions* at times, and the results of those actions, but why need there be anything more substantial than that? There is no need here, it might be thought, for our respective perceptual attention to be *out in the open* in any robust sense. Perhaps all that is needed is parallel attention to the ongoing action.

In response to this objection, it is worth considering what I shall call the *guided rationality of joint agency.* Suppose we are engaging in a joint intentional action *with the shared intention that we make a Hollandaise Sauce*. For this joint action to be intentional under this description it must be being guided by that shared intention; or, at least, that has been the thrust of the argument thus far. This is to say – once decomposed into its constructivist Bratmanian elements – I must be being guided by *your* intention that we make Hollandaise Sause, just as much as I'm guided by mine. What would it mean for me to be guided by your intention? How can your intention get a grip on me such that it *guides* my actions? If we were to think of your intention guiding my actions in precisely the way that my intention sometimes *causally controls* mine, then we truly would be violating an 'own action' condition on intention in an implausible way. For it is not as if I can get inside your skin, so to speak, and make your limbs move.

One way your intention might be thought to rationally guide my actions would be for my actions to be visually guided by your observable limb movements and

their effects. I observe you shifting the pan this way, causing me to pour in that way. Given that your actions and movements are themselves guided by your intention that we make the Hollandaise Sauce, then, it might be argued, if we suppose that guidance is transitive, I will thereby be guided by your intention. But this isn't rational guidance. It is, rather, a matter of reactively responding to your guided intentions. Rational guidance, as I shall explain shortly, is something more than this.

Another way in which your intention might be thought to guide my actions would be if I were to act in a way that I thought was responsive to your wishes and intentions, such that I adopted your prior intentions as if they were my own. You shift the pan this way, causing me to pour in that way because I am concerned to fulfil what you had intended. But this also isn't rational guidance – your intentions are here normatively *influencing* my intentional actions (they provide me with a reason to act in a certain way perhaps), but they are not rationally guiding my actions. They are not rationally guiding my actions because your intention, in this case, is not poised to control, intervene, correct or respond to my actions in pursuit of executing the prior intention.

Guidance, then, is not merely having one's actions being causally affected by a prior event such as an intention (Frankfurt 1978), nor is it a matter of merely being rationally responsive to the intentions of another, as if satisfying another's intentions were a reason for my action. Rather, as already noted, rational guidance is a matter of having a multi-faceted capacity and skill to control, intervene in, respond to, or even 'go with the flow' with regards to, one's actions in pursuit of the end specified in the content of the prior intention, such that one is able to anticipate effects, intervene to prevent mishaps, or respond to correct errors. If I am to be partially *rationally guided* by your intentions (as aspects of a shared intention) then your intentions must be poised to partially control, intervene in, correct, step back from, or be responsive to, my actions in pursuit of our shared intention. That is to say: your intention <that we make a Hollandaise Sauce together by means of each of our intentions and meshing sub-plans> must be poised to have that kind of guiding influence over my actions.

The present suggestion is that joint attention – in its full-blown 'mutually manifest' guise – plays the functional role of enabling your intention, in combination with my intention, to have that guiding role over both your actions, and my actions, (which together constitute the joint action) by making mutually manifest the elements of the environment that contribute to fulfilling the corresponding elements of the prior plan of which our shared distal prior intention was the governing element. By making features of the environment mutually manifest, our actions will be directed, in meshing, complimentary ways, towards fulfilling the shared intentions and the shared plan. Without us being mutually aware of the relevant features of the environment in which we are operating (the objects, the processes, the actions and the 'shared', not purely ego-centric space in which this occurs), then we would be unable to update, adapt and consult the course of our shared plan – what needs to be done and what has been done – in the necessary way.

Part of what makes joint attention important here is that plans are partial and are flexible in means. As a consequence, the execution of some sub-intentions will be

on the fly. Suppose, for instance, that we have a partial plan involving numerous fairly fine-grained tasks that could be performed by either one of us when the time comes. If we are to operate broadly in synch with each other then we need to be aware of who has done what, who is doing what, and who should be doing what, when and how (given the present lay-out of the environment and present locations of the participants). The features of the environment dictate our actions insofar as they are features that will fulfil our plans and are located within our field of attention. Mere parallel attention will not allow us do this because, without knowing what the other agent is aware of, we will not be able to anticipate their actions; we will be unable to determine the layout of the environment that is available to them and at what stage of the shared plan they think we are at.

It might be thought that the agents need only have awareness of each other's attention to the environment, but not that it be 'mutually manifest'.[15] But mutual manifestation has a role to play here which mere awareness of each other's attention would not fulfil. If we are *just* aware of each other's attention then, ex hypothesis, we are not aware of the other's awareness of our attention. That is to say, I might just think that you think that I am an inattentive fool. But this would play havoc with our activity. If a functional role of attention is to implement a multi-faceted skill of guidance, then presumed inattentiveness on either of our parts should lead to the other constantly intervening and correcting for our predicted bumbling, inattentive actions. Consequently, if I suspect that you think that I am an inattentive fool, then I should expect you to constantly interfere in my actions, and I will be compensating for that.

Joint attention – in its full-blown mutually manifest sense – allows us to be guided by a prior shared intention and plan. It is because we are *jointly* attending to the environment and the features in it, that your intention and my intention (which comprise in the appropriate Bratmanian sense our shared intention) together rationally guide the joint action. We are together attentively monitoring the on-going activity – what needs to be done, what has been done, what needs correcting, what needs intervention, what needs skilled, automatic activity – in a way that is open. This openness is crucial. It enables both of us to anticipate, intervene, control and respond to the ongoing activity *as a manifestation of a shared intention*, rather than merely reactively responding to the other. These are manifestations of a shared intention because they are actions that are fulfilling and forwarding what is required in order to fulfil that intention as an element of a hierarchically ordered, structured and coordinated shared plan that, in its details, will be flexible and constantly updated. It is not merely the ongoing action, or the relevant elements of the environment, that are perceptually open to us as a manifestation of a shared intention, but also, when they occur, our respective tendencies to correct and intervene (e.g. I tell you to bring the pan closer, or you tell me to 'watch out' as I am about to drop the sauce).

[15] Blomberg (2016) argues for a point related to this with regards to Common Knowledge and Joint Action

If we are both attentive to the objects, locations, actions and events in front of us – and know that we are both so attentive – then we can immediately conceptualise and frame the actions that are performed (be they controlled, intervening or correcting) as actions that are directed towards, because guided by, a prior shared intention. Joint attention, because it interfaces with prior intentions, allows us to *mutually recognise* both the actions which are being intervened on, as well as the intervening actions, *as manifestations of a prior shared intention*.

8.3 Natural Intersubjectivity and Development

I have argued that minimal cooperation can interface with shared intentions by way of joint attention. Joint attention allows for an ongoing joint activity to be rationally guided and coordinated by a prior shared intention. The claim is that in such cases, there is no need to appeal to a more minimal concept of a 'shared goal'.

A question which opens up about this, however, is what to say about three kinds of cases where there would appear to be joint activity, but where shared intentions are not operative (Tollefsen 2014). These are: (a) spontaneous, 'flash mob' joint actions; (b) interactions involving young children and infants; (c) interactions between animals. I shall focus on infant interaction, although more would need to be said about the others.[16]

8.3.1 The Development of Shared Intentions

I do not have space to fully develop the claim but I shall argue that joint attention and joint action play a developmental role in allowing children to engage in structured activities that, although not actually governed by shared intentions, are nevertheless enacted in a way that partially assumes the existence of shared intentions and shared plans. The capacity for joint attention allows infants to conceptualise the joint activities that they are engaged in as having a place within a wider normative structure, even when they are not yet capable of fully grasping that normative structure or the content of some of the elements of that structure.

[16] Very briefly – I would argue that Flash Mob actions are defined in terms of their being self-consciously *un*planned. To the extent that this is part of their motivation and definition, then they are defined and motivated (negatively) in terms of planning. With regards to animal interaction, I would follow Tomasello et al. (2005) – or perhaps more specifically, Henrike Moll's *cultural transformation* version of that hypothesis (Moll 2016, 2018; Kern and Moll 2017) –that shared intentionality is a defining feature of human uniqueness, and so animal interaction, unlike infant interactive cognition, is to be understood in a minimal way, along the lines suggested by Apperly and Butterfill (2009).

It would be useful, I think, to begin with the metaphorical picture I sketched in Sect. 8.1 and that was operative in the background of the discussion in the previous section. We can think of plans as a kind of hierarchical diagram, with lines drawn from intentions to sub-intentions and from sub-intentions to sub-sub-intentions, and so on. As Bratman stresses, some sub-intentions will be bargained over (who does what when and how), so we can think of each of these boxes as being filled in with further content as and when the time comes, and boxes being added and taken away. Further, we can think of joint attention as allowing agents to (a) identify how to execute the sub-intention which appears in a sub-intention box, and then, (b) when the sub-intention has been executed, mentally 'ticking' the box, so as to mentally record its successful execution. Now, if, as I suggested before, shared plans are shared in a fairly robust way – they operate, functionally at least, like a wall-chart or online schedule which we can both edit – then we can see, as argued, in the previous section, why Joint Attention must be something more than parallel; it must be playing a public role to ensure that we are updating our copies of the plan together.

With this metaphor in hand, I want to suggest that infants and children often engage in activities with a sense that they are doing something as part of a larger, more structured prior shared plan, albeit a shared plan that they don't fully understand or have access to. They might engage in games with rules for instance, yet will not know the rules beforehand, and learn them through engaging in the activities themselves (Wittgenstein 1956; Bruner 1993). Similarly, children might engage in more informal plans, but not have a full idea of what the plan is. They understand that what they are doing has a role, and they can begin to piece it together as they proceed through the activity. So, it is a kind of reverse picture of the planning theory: they might have a sense that there are 'boxes' to be ticked off and filled in, but not have an idea of how these boxes are structured, how they relate, or what their content is. Children will work from the activities to an understanding of the sub-plans to an understanding of the whole plan. In many cases, it might well be that the infants and children have an initial idea of a shared prior intention ('that we make a cake together'), without yet having an understanding of either what it means for it to be genuinely shared, or what is required to fill it out. Their understanding of what is required for a plan to be shared comes about, on this picture, by means of engaging and 'pitching in' (Paradise and Rogoff 2009) in collaborative activities. Only when they have mastered these activities, and the way that actions join with sub-intentions and sub-intentions with intentions, will they begin to see what a robust shared intention is.

By learning smaller activities involved in baking cakes, for instance, children will, as they go along, be learning how to fit those activities into a wider plan, and will come to see the activity as an intentional activity with the intention that *p*. In coming to engage in the activities, they might also eventually come to have an understanding of the complex structure of a shared intention; that is, of what it is to share an intention such that the content of one's own plan is intertwined with the content of the other's plan, and where one is responsive to the wishes and plans of the other and how they might mesh with one's own.

In some sense, this is illusory. There is no genuine *shared* intention between the caregiver and the infant. But the claim here is that the pedagogical attitude of a caregiver is (sometimes, in structured activities) one of a caregiver making a shared plan that is constructed *as if* it were equally the intention of the child, even when the child is not yet able to conceptualise the complex intentions that go into a shared intention, or the complex hierarchy by which it takes shape.

As with the adult case, joint attention plays a key role. By jointly attending, the participants are able to monitor, intervene, control and respond to the ongoing activity. At a very early stage, infants will almost always be the recipients of this guidance. The guidance will come almost fully from the caregiver's ersatz shared intention.

This view is related to the idea of *Natural Pedagogy* as developed by Gegerly and Csibra (2006), Csibra and Gegerly (2009, 2011), which postulates a "well-organised package of biases, tendencies and skills" (Gergely and Csibra 2006: 8) for learning on the side of children and infants, and a natural tendency towards structured teaching on the side of caregiver adults (Csibra and Gegerly 2011). Csibra and Gegerly's focus is on the often cognitively opaque *cultural knowledge* that natural pedagogy helps develop; knowledge which children often take to be generic, holding across contexts and people (Egyed et al. 2013; Topál et al. 2008), and which is often imparted within a formal context. This suggests a relation between Minimal Cooperation and Massively Shared Agency. The focus here, however, has been on the development of understanding others' plans and how those plans interact with one's own actions and how this might develop into a more sophisticated capacity for shared intentions. This is a relation between Minimal Cooperation and the contingent and local plans that arise in Modest Sociality, and is perhaps more closely related to what Barbara Rogoff calls 'informal learning' (Rogoff et al. 2016).

8.4 Concluding Remarks

To summarise, I have argued that we need an understanding of how Bratmanian shared intentions might interface with joint intentional actions, such that joint intentional actions can be seen as executions of shared intentions, which track those prior shared intentions in a way that would allow us to say that the agents are jointly φ-ing with the intention that p. That this is something of a lacuna in traditional accounts of collective intentionality, is not a new complaint (see, e.g. Tollefsen and Dale 2012, who call this the 'execution problem'). But I have attempted to answer this question by arguing that joint attention plays a pivotal functional role in this by enabling the agents' actions to be coordinated in a way that is rationally guided by and responsive to their prior shared intentions. Finally, I argued that this might help provide a framework for explaining the development of interactive skills within an informal learning environment.

References

Apperly, I., & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*(4), 953–970.

Anscombe, G. E. M. (1957). *Intention*. Cambridge, MA: Harvard University Press.

Barwise, J. (1988). Three views of common knowledge. *Proceedings of the 2nd conference on theoretical aspects of reasoning about knowledge* (pp. 365–379).

Blomberg, O. (2016). Common knowledge and reductionism about shared agency. *Australasian Journal of Philosophy, 94*(2), 315–326.

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

Bratman, M. (1999). *Faces of intention: Selected essays on intention and agency*. Cambridge: Cambridge University Press.

Bratman, M. E. (1992). Shared cooperative activity. *The Philosophical Review, 101*(2), 327–341.

Bratman, M. E. (2007). *Structures of agency: Essays*. New York: Oxford University Press.

Bratman, M. E. (2014). *Shared agency: A planning theory of acting together*. Oxford: Oxford University Press.

Brownstein, M. (2014). Rationalizing flow: Agency in skilled unreflective action. *Philosophical Studies, 168*(2), 545–568.

Brozzo, C. (2017). Motor intentions: How intentions and motor representations come together. *Mind & Language, 32*(2), 231–256.

Bruner, J. (1993). *Act of meaning*. Cambridge, MA: Harvard University Press.

Butterfill, S. A. (2012). Joint action and development. *The Philosophical Quarterly, 62*(246), 23–47.

Butterfill, S. A. (2013). Interacting Mindreaders. *Philosophical Studies, 165*(3), 841–863.

Butterfill, S. A., & Sebanz, N. (2011). Joint action: What is shared? *Review of Philosophy and Psychology, 2*(2), 137–146.

Butterfill, S. A., & Sinigaglia, C. (2014). Intention and motor representation in purposive action. *Philosophy and Phenomenological Research, 88*(1), 119–145.

Campbell, J. (2002). *Reference and consciousness*. New York: Oxford University Press.

Campbell, J. (2005). Joint attention and common knowledge. In N. Eilan et al. (Eds.), *Joint attention: Communication and other minds* (pp. 287–297). Oxford: Oxford University Press.

Chant, S. R., Hindriks, F., & Preyer, G. (Eds.). (2014). *From individual to collective intentionality*. Oxford: Oxford University Press.

Christensen, W., & Michael, J. (2016). From two systems to a multi-systems architecture for mindreading. *New Ideas in Psychology, 40*, 48–64.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153.

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transaction of the Royal Society B, 360*, 1149–1157.

Csíkszentmihályi, M. (1975). *Beyond boredom and anxiety*. San Francisco: Jossey-Bass Publishers.

Csikszentmihályi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.

Davidson, D. (1963). Actions, reasons, and causes. *The Journal of Philosophy, 60*(23), 685–700.

Dickie, I. (2015). *Fixing reference*. New York: Oxford University Press.

Egyed, K., Kirly, I., & Gergely, G. (2013). Communicating shared knowledge in infancy. *Psychological Science, 24*(7), 1348–1353.

Eilan, N. (2007). Consciousness, self-consciousness and communication. In T. Baldwin (Ed.), *Reading Merleau-Ponty*. London: Routledge.

Evans, G. (1982). In J. McDowell (Ed.), *The varieties of reference*. New York: Clarendon Press Oxford.

Fantasia, V., De Jaegher, H., & Fasulo, A. (2014). We can work it out: An enactive look at cooperation. *Frontiers in Psychology, 5*, 874.

Ferrero, L. (2013). Can I only intend my own actions? *Oxford Studies in Agency and Responsibility, 1*, 70–94.
Fiebich, A., & Gallagher, S. (2013). Joint attention in joint action. *Philosophical Psychology, 26*(4), 571–587.
Frankfurt, H. G. (1978). The problem of action. *American Philosophical Quarterly, 15*(2), 157–162.
Gergely, G., & Csibra, G. (2006). Sylvia's recipe: The role of imitation and pedagogy in the transmission of cultural knowledge. In: *Roots of human sociality: Culture, cognition, and human* interactions (pp. 229–255).
Gilbert, M. (2009). Shared intention and personal intentions. *Philosophical Studies, 144*(1), 167–187.
Grice, H. P. (1974). Method in philosophical psychology (from the Banal to the Bizarre). *Proceedings and Addresses of the American Philosophical Association, 48*, 25–53.
Heal, J. (1997). Indexical predicates and their uses. *Mind, 106*(424), 619–640. Reprinted in J. Heal (2003) *Mind, Reason and Imagination.* Cambridge University Press.
Kern, A., & Moll, H. (2017). On the transformative character of collective intentionality and the uniqueness of the human. *Philosophical Psychology, 30*(3), 315–333.
Luthra, Y. (2016). Non-rational aspects of skilled agency. *Philosophical Studies, 173*(8), 2267–2289.
McDowell, J. (2009). Acting as one intends. In J. Dancy & C. Sands (Eds.), *Philosophy of action: An anthology*. Oxford: Wiley-Blackwell.
Moll, H. (2016). Tension in the natural history of human thinking. *Journal of Social Ontology, 2*(1), 65–73.
Moll, H. (2018). The transformative cultural intelligence hypothesis: Evidence from young children's problem solving. *Review of Philosophy and Psychology., 9*(1), 161.
Mylopoulos, M., & Pacherie, E. (2017). Intentions and motor representations: The interface challenge. *Review of Philosophy and Psychology, 8*(2), 317–336.
Pacherie, E. (2006). Towards a dynamic theory of intentions. In: *Does consciousness cause behavior* (pp. 145–167).
Pacherie, E. (2011). The phenomenology of joint action: Self-agency versus joint agency. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience*. Cambridge, MA: MIT Press.
Pacherie, E. (2013). Intentional joint agency: Shared intention lite. *Synthese, 190*(10), 1817–1839.
Paradise, R., & Rogoff, B. (2009). Side by side: Learning by observing and pitching in. *Ethos, 37*(1), 102–138.
Peacocke, C. (2005). Joint attention: Its nature, reflexivity, and relation to common knowledge. In N. Eilan, C. Hoerl, T. McCormack, & J. Poessler (Eds.), *Joint attention: Communication and other minds: Issues in philosophy and psychology* (pp. 298–324). Oxford: Clarendon Press.
Rogoff, B., Callanan, M., Gutirrez, K. D., & Erickson, F. (2016). The organization of informal learning. *Review of Research in Education, 40*(1), 356–401.
Schear, J. K. (Ed.). (2013). *Mind, reason, and being-in-the-world: The McDowell-Dreyfus debate*. New York: Routledge.
Searle, J. (1990). Collective intentions and actions. In P. R. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 401–415). Cambridge, MA: MIT Press.
Seemann, A. (2012). Joint attention: Towards a relational account. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 183–202). Cambridge, MA/London: MIT Press.
Shapiro, S. (2014). Massively shared agency. In M. Vargas & G. Yaffe (Eds.), *Rational and social agency: The philosophy of Michael Bratman* (pp. 257–293). Oxford: Oxford University Press.
Shepherd, J. (2017). Skilled action and the double life of intention. *Philosophy and Phenomenological Research,* Online First.
Sinigaglia, C., & Butterfill, S. A. (2015). On a puzzle about relations between thought, experience and the motoric. *Synthese, 192*(6), 1923–1936.

Smithies, D. (2011). Attention is rational-access consciousness. In C. Mole, D. Smithies, & W. Wu (Eds.), *Attention: Philosophical and psychological essays* (pp. 247–273). New York: Oxford University Press.

Tollefsen, D. (2014). A dynamic theory of shared intention and the phenomenology of joint action. In Chant et al. (Eds.), *From individual to collective intentionality: New essays* (pp. 13–33). New York: Oxford University Press.

Tollefsen, D. (2005). Lets pretend! Joint action and young children. *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.

Topál, J., Gergely, G., Miklósi, Á., Erdőhegyi, Á., & Csibra, G. (2008). Infants' perseverative search errors are induced by pragmatic misinterpretation. *Science, 321*(5897), 1831–1834.

Tomasello, M., Carpetner, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioural and Brain Sciences, 28*, 675–735.

Tuomela, R. (2007). *The philosophy of sociality: The shared point of view*. Oxford: Oxford University Press.

Velleman, J. D. (2008). The way of the wanton. In C. Mackenzie & K. Atkins (Eds.), *Practical identity and narrative agency*. New York: Routledge.

Watzl, S. (2017). *Structuring the mind: The nature of attention and how it shapes consciousness*. Oxford: Oxford University Press.

Wilby, M. (2010). The simplicity of mutual knowledge. *Philosophical Explorations, 13*(2), 83–100.

Wittgenstein, L. (2001) [1956]. *Philosophical investigations*. Oxford: Blackwell Publishing.

Wu, W. (2011). Confronting many-many problems: Attention and agentive control. *Noûs, 45*(1), 50–76.

Chapter 9
Solving the Hi-lo Paradox: Equilibria, Beliefs, and Coordination

Francesco Guala

Abstract Game theorists have tried for decades to explain how rational or quasi-rational individuals are able to coordinate in situations in which there is more than one way to satisfy their preferences. In this chapter I focus in particular on the formation of common beliefs that supposedly sustain coordination in Hi-lo games. I review some attempts to solve the problem, such as bounded rationality, team reasoning, and solution thinking. Following their lead, I suggest that successful coordination is belief-less coordination, and that simple means-ends rationality explains how coordination problems may be solved using techniques of minimal mindreading.

Keywords Coordination · Team reasoning · Solution thinking · Beliefs · Focal point

9.1 Introduction

The extraordinary success of *homo sapiens* seems to be due in large part to our capacity to work together in coordinated, highly efficient groups. Biologists of course remind us that we are not the only hyper-social creatures on Earth. But we seem to be unique in the way in which we implement coordination: while the behaviour of bees and ants is largely genetically determined, *homo sapiens* is able to form strong bonds with genetically unrelated individuals, to flexibly enter and exit social relations, to imagine and implement different types of organization as required by new circumstances and needs. No other creature, as far as we know, is able to do this – not at least on such a scale and with such a level of sophistication.

F. Guala (✉)
Dipartimento di Filosofia, Università degli Studi di Milano, Milano, Italy
e-mail: francesco.guala@unimi.it

A. Fiebich (ed.), *Minimal Cooperation and Shared Agency*, Studies in the Philosophy of Sociality 11, https://doi.org/10.1007/978-3-030-29783-1_9

In light of all this, it is tempting to infer that our capacity to coordinate must be due to our unique cognitive skills. But this claim turns out to be problematic. There are reasons to believe that being too clever may actually *hamper* rather than help coordination.

A powerful argument comes from the theory of strategic games. Game theory is the branch of mathematical social science that analyses behaviour in interactive situations. Individuals are typically modelled as rational players (decision-makers) who behave optimally in response to external events and in light of their expectations regarding the behaviour of other players. The theory predicts that in strategic contexts the players should be able to converge on Nash equilibrium states – behavioural patterns such that no player can obtain a better outcome by changing her action unilaterally.

The cognitive skills required to converge on the Nash equilibria of some games, however, seem extremely demanding: coordination for example may require the capacity to examine several orders of nested beliefs ('I believe that you believe that I believe that you believe …'). And in many cases, it is not at all clear where such beliefs come from – how they are formed in the first place.

An obvious way to solve this puzzle is to revise the original assumption: perhaps being too clever does not help. Perhaps we are able to coordinate smoothly *in spite* of our remarkable cognitive skills, not because of them. But this raises a number of questions: what kind of cognitive mechanisms facilitate coordination? What mix of sophisticated and naïve thinking is involved? Is it possible to build a general account of the cognitive basis of coordination?

I will argue that minimal mindreading techniques offer plausible answers to some of these questions. I will first set the stage describing a simple coordination problem (the 'Hi-lo' game) that cannot be solved using the standard tools of game theory (Sect. 9.2). In the third section I will introduce a classic account of coordination, based on the concept of salience. In the fourth and fifth sections I describe two reasoning strategies that people may use to converge on salient equilibria, based respectively on bounded rationality and team reasoning. While the first strategy seems to be ruled out by experimental data, the second one points in an interesting direction. Section 9.6 shows that team reasoning is formally analogous to a reasoning strategy called solution thinking, and that in both cases coordination is reached by disregarding other people's beliefs. Section 9.7 backs up this insight using evidence from cognitive and developmental psychology, suggesting that the capacity to attribute beliefs to other agents plays a limited and peculiar role in our everyday interactions. Section 9.8 discusses evidence gathered in experiments with infants, indicating that from an early age humans are able to anticipate people's behaviour using simple principles of means-ends rationality in which belief attribution plays no role. Section 9.9 concludes: successful coordination is belief-less coordination, since reasoning about beliefs would very frequently lead us astray.

9.2 Coordination and the Hi-lo Paradox

A group of people face a *coordination problem* every time they have the opportunity to attain an outcome that is in some way beneficial for all of them – it is better than lack of coordination, at least – but that cannot be achieved by each individual acting independently from the other members of the group. Crucially, for coordination to be a 'problem', it is necessary that the beneficial outcome can be achieved in more than one way. For example: to prepare for a dinner party, Ann might go shopping while Bob picks up the kids at nursery school, or the other way around; to score a goal, Jane might take the corner kick while Jill heads the ball into the net, but they might switch roles; and so on and so forth.

Sometimes, as we will see shortly, a particular solution to a coordination problem is clearly better than any other solution, and would be preferred by everyone involved. It is, in the jargon of economics, the only efficient solution. But efficient outcomes are not the only rational solutions. Suppose that Bob's office is closer to the nursery school, Ann's office is closer to the superstore, and neither of them want to get stuck in peak-hour traffic. The most efficient solution is that Bob picks the kids while Ann goes shopping. But if Bob for some reason believes that Ann will go to the nursery school, then he should better go shopping. That pair of actions is also a rational solution (an equilibrium), albeit not an optimal one in light of agents' preferences.

A useful way to represent such problems of coordination is by means of game matrices. The matrix in Fig. 9.1 for example represents a coordination game with two players and two possible actions. Throughout the paper I will follow the usual conventions of game theory: the rows represent the possible actions of the first player (Ann), the columns those of the second player (Bob). The cells are the outcomes of the interactions of the two players, and are described by two numbers or payoffs (the first number for Ann, the second one for Bob). The payoffs indicate the order of preference of each player: the highest number is associated with the most preferred outcome, and so on until the lowest payoff is associated with the least preferred outcome. The configuration of payoffs in Fig. 9.1 indicates that any failure to coordinate would lead to the worst outcomes for both players: (Hi, Lo) and (Lo, Hi). There are also two ways to coordinate successfully, but one of them (Hi, Hi) is clearly better than the other (Lo, Lo).

Each of these two solutions is a *Nash equilibrium*, a profile of actions or strategies such that each action is a best response to the actions of the other players. In

Fig. 9.1 A simple coordination game (Hi-lo)

	Hi	Lo
Hi	2, 2	0, 0
Lo	0, 0	1, 1

equilibrium, in other words, no player can do better by changing her strategy unilaterally. If the others do their part in the equilibrium, no player has an incentive to deviate.[1]

During the last decade coordination games with uniquely efficient equilibria have become prominent in the philosophical and scientific debate thanks mainly to the work of Michael Bacharach (1999, 2006).[2] Bacharach calls them 'Hi-lo games', due to the structure of incentives (there is a 'high-payoff' equilibrium and at least one 'low-payoff' equilibrium).

There are various reasons to pay attention to these games: the first one is that many social institutions seem to solve coordination problems of this kind, helping people to converge on high-payoff equilibria.[3] Thus coordination games may hold the key to understand a wide range of important mechanisms in social ontology and social cognition generally.

The second one is more esoteric – I will only mention it here without elaborating further: in many situations, even games with several efficient equilibria can be transformed into Hi-lo games – making them easier to solve. This is in fact one of the motivations that brought Bacharach and others to the analysis of Hi-lo games.[4]

The third reason is the paradoxical nature of Hi-lo games: such games appear trivial at first sight, because we all have the irresistible intuition that it would be *irrational* to converge on the Lo equilibrium. And in fact the overwhelming majority of people choose the Hi strategy, when they play it in the laboratory. But in spite of the appearances, it is difficult to explain coordination in Hi-lo games. Choosing Lo is not irrational: as a matter of fact, it is optimal if the other player chooses Lo, and neither player can rule that out a priori. Ann cannot exclude that Bob will choose Lo because he is facing the same problem that she is facing. Since Ann hasn't decided what to do yet (she is still deliberating), Bob cannot have decided what to do either. But if we cannot prove that choosing Lo is irrational, then we cannot prove that choosing Hi is uniquely rational either.

So why do we think that choosing Lo is silly? Here is a typical – but invalid – piece of reasoning. Ann has no idea what Bob will do. So she assigns an equal probability to Bob choosing either strategy ($p = 1/2$). Given this probability, the expected value of choosing Hi is

$$EV_{\mathrm{H}} = \left(2 \times 1/2\right) + \left(0 \times 1/2\right) = 1.$$

The expected value of choosing Lo is

[1] There is also a third equilibrium in this game, where the players choose each strategy with a certain probability, and they have no incentive to change this probability. Such 'mixed-equilibrium strategies' are highly implausible in coordination games, and therefore I will ignore them from now on. I trust the reader to be able to figure out, before the end of the chapter, why they are so implausible.

[2] Previous discussions can be found in Hodgson (1967), Gauthier (1975), Sugden (1993).

[3] On how institutions help solve coordination problems, see e.g. Guala (2016).

[4] Gauthier (1975), Sugden (1993), Bacharach and Bernasconi (1997).

$$EV_{\mathrm{L}} = (1 \times 1/2) + (0 \times 1/2) = 1/2.$$

So it seems rational for Ann to choose Hi, because it maximizes expected value.

The problem is that Bob should be able to see this, and should be able to predict that Ann will choose Hi. His best response, given this belief, is to play Hi for sure. Ann however should be able to anticipate this, too. But then she cannot simultaneously believe that Bob will play Hi with $p = 1/2$, and that he will play it with $p = 1$: her beliefs would be inconsistent.

One may point out that in many cases of real-life coordination, there seems to be an obvious way out of this conundrum: the players can communicate, make proposals, and agree on a rule of conduct. Surprisingly, however, communication does not solve the problem. Suppose that Ann and Bob state a rule such as 'play Hi' (they write it on a blackboard that everyone can see, for example). Each one of them should follow the stated rule only if she thinks that the other will follow it. But the other will follow it only if she thinks the first one will follow it. But how can she believe *that*? After all, the first player will not follow it unless the second one will follow it. In a nutshell, stating the rule does not change the structure of the coordination problem because it does not change the incentives of the game. Ann's reason to follow the rule is not that the rule has been stated. It is the fact that Bob has an incentive to follow the rule. But Bob has an incentive only if Ann follows the rule. So we are still trapped in the same circle of reasoning.

There are other ways to try to justify the intuition that converging on the Lo equilibrium is irrational – reviewed by Bacharach (2006) – and they all fail. The Hi-lo paradox is that both equilibria are rationalisable, in the sense that each one is implied by a certain configuration of initial beliefs. But, for each of them, we lack an explanation of how the players could come to have the beliefs that are necessary to play that equilibrium. This is particularly disturbing for the Hi equilibrium, which seems intuitively to be the right choice.

9.3 Salience and Focal Points

The problem of belief formation is central in *epistemic game theory*. The ambitious goal of epistemic game theory is to explain how people reason their way towards equilibria, by modelling players as rational decision makers in the classic expected utility sense.[5] Because epistemic game theory can be extremely sophisticated, both conceptually and mathematically, it is tempting to blame economists for building their models upon unrealistic, idealized mathematical assumptions. But this would be mistaken. The problem of belief formation is not a problem that is peculiar to mathematical social science. In fact, it affects *any* theory that assumes that people

[5] For an overview, see e.g. Pacuit and Roy (2015).

are able to detect and interpret each other's beliefs. The latter assumption is so common in fact that it is considered a pillar of so-called folk psychology – the way in which we spontaneously explain and predict behaviour during everyday interactions.

Epistemic game theory shares the folk-psychological perspective, assuming that players are mental agents or mindreaders.[6] So perhaps there is something wrong with the assumption of mental agency in the first place. Perhaps people do not reason about others' beliefs, especially if the reasoning ought to take place at the level of second- or third-order beliefs (beliefs about beliefs about beliefs). Perhaps they follow a much simpler procedure to predict each other's behavior.

One of the striking aspects of real-life coordination problems is that we rarely pause to think about them. In most cases the solution looks obvious, indeed so obvious that it does not look like a problem at all. Suppose that John is a football fan, while his partner Jane likes to watch reality TV shows. Unfortunately, they own only one television set. If John's favourite team plays on Wednesday, and Jane's favorite show is on Thursday, then it is obvious that John should watch TV on Wednesday and Jane should do it on Thursday. This is the most natural arrangement, the one that first comes to mind and that 'stands out' from the crowd of the possible solutions.

A solution that 'stands out' from the crowd is a *focal point*. This expression was coined many years ago by Thomas Schelling in *The Strategy of Conflict* (1960), a book that has laid down the foundations for much contemporary work on coordination. Schelling noticed that individual decisions are facilitated by the existence of focal points. He also remarked that a focal point is often salient for reasons that have little to do with the strategic features of the game, but depend on properties of the environment that should be irrelevant from a purely logical point of view. And yet people often make better decisions if they do not engage in complex logical reasoning – just as they do when they instinctively choose the efficient equilibrium in the Hi-lo game.

Schelling's suggestion sounds basically right. However, it is at best only the beginning of an explanation. Without a theory of salience and focal points, several crucial questions remain unanswered. What is it that makes 'Hi' special in the Hi-lo game? What sort of reasoning does lead people to converge on the salient solution? What kind of amendments to game theory should we make in order to accommodate Schelling's insight?

[6] I will use these terms as roughly synonymous: mindreading is the activity of detecting and interpreting the mental states of other agents. A mental agent uses mental states attributions to explain and predict the behaviour of others.

9.4 Bounded Rationality

David Lewis was the first philosopher to explore the depth of Schelling's insight, in his ground-breaking book on *Convention* (1969). Among several important contributions, he also proposed a simple interpretation of the cognitive process that may underlie spontaneous coordination of the Schelling type.

Lewis postulated a 'brute' disposition to converge on focal points: failing any reason not to do so, people choose the option that is most salient. Let us call it the 'primary salience' principle: when you face a coordination problem, choose the first solution that comes to mind – the one that is most salient.[7] If we know what is salient for them, then we can predict other people's behaviour quite reliably. This idea is broadly consistent with the hypothesis of bounded rationality: people are not fully rational in the expected utility sense, because they often rely on 'fast and frugal heuristics' to save time and cognitive effort.[8] In the case of coordination, bounded rationality seems all the more plausible because a group of fully rational agents would struggle to find the solution to a coordination problem. They would be outperformed by a group of boundedly rational individuals who follow the primary salience principle.

Notice that for primary salience to work it is not necessary that *everyone* is boundedly rational. Different people may have different cognitive skills, and employ them differently depending on the circumstances. For Lewis' explanation to go through, we only need to assume that *some* people follow the salience principle – i.e. choose the option that first comes to mind. While these naïve players engage in 'level-0 reasoning' (Stahl and Wilson 1995), other players may be able to do better: they may figure out for example that the naïve players are going to choose the salient option, engaging in level-1 reasoning. Following the same logic, we can consider also level-2, level-3 (etc.) players, until we reach a limit (a level of reasoning) beyond which no one goes.[9] The nice thing is that, no matter how sophisticated the most sophisticated players are, their behaviour will be determined by the choices of level-0 players.

To see why, suppose that several individuals are playing anonymously a series of one-shot coordination games, pairwise, with other individuals who are drawn randomly from the same population. Suppose that Ann chooses the salient option because she's a level-0 player. Bob, a level-1 player, will understand that it is in his interest to choose the salient option, because *if* he is matched with Ann, choosing salience will guarantee that they successfully coordinate. Ann's naïve behaviour, in other words, increases the probability of coordinating on that option for every other member of the population. Carol, who is a level-2 player, will also understand that

[7] The expression 'primary salience' has been coined by Mehta, Starmer and Sugden (1994) in a paper that will be discussed shortly.

[8] See e.g. Simon (1982), Gigerenzer and Selten (2002).

[9] The evidence indicates that very few people actually go beyond level-2 reasoning (cf. Nagel 1995; Camerer et al. 2004).

the probability of coordinating on that option is higher, since Ann and Bob are both going to choose it – and so on with Dave, Eve, until the highest level of cognitive sophistication has been reached. The presence of even a minority of level-0 players triggers a cascade, ensuring that everyone will converge on the salient option.

This account is known among behavioural game theorists as 'level-k reasoning' or, more generally, as 'cognitive hierarchy theory'. The theory has several interesting applications, many of them outside the realm of coordination games. For example, it works well in some games of pure competition, where the goals are not aligned and the players try to out-guess each other. As an explanation of coordination, however, it has a major weakness: it seems to be empirically false.

Judith Mehta, Chris Starmer and Robert Sugden (1994) have shown in a classic experiment that equilibrium selection is *not* driven by the choices of level-0 players. Recall that for a naïve player the salient option is the first one that comes to mind. If level-1, level-2, etc. players can anticipate her choice, they should all converge on what the level-0 player finds primarily salient. To discover what is actually primarily salient, Mehta and colleagues simply asked people to indicate the first option that comes to mind in a series of open questions. For example, they asked their subjects to pick a year ('any year, past, present or future'); to name any flower; to name any British town or city; to write down any positive number; and so on. As one would expect, they obtained a wide range of answers. Part of the variety can be explained by the fact that what comes to mind depends on personal factors that are only meaningful to each individual responder. (For example, many experimental subjects picked the year they were born.) This, intuitively, may constitute a problem for coordination – primary salience may be too private, as we shall see shortly.

In these trials the elicitation of salience did not take place in an interactive setting. That is to say: the subjects were not given an incentive to give the same answer. Nonetheless, according to cognitive hierarchy theory, primary salience should drive equilibrium selection also in coordination games – when people are aware that they have an incentive to converge on the same solution. But this is not what Mehta and colleagues found. When Mehta and colleagues asked a second group of subjects the same questions, explaining that they would get a prize if the same answer was given by an anonymous partner, they discovered that most people followed a completely different strategy: they chose an answer with *general* significance. For example, the most frequently chosen year was 1990, the year the experiment took place. The most frequently chosen number was 1 (instead of 7, 2, or 10, which were the most frequently chosen numbers in the first trial). The most frequently chosen colour was red, instead of blue, and so on. These choices were driven by what Mehta, Starmer, and Sugden call '*Schelling salience*'.

As opposed to primary salience, Schelling salience presupposes an awareness that there is a common problem to be solved. When they coordinate, people do not just pick the first option that comes to mind, but the first one that will help solve the problem they are facing. Cognitive hierarchy theory therefore seems to be on the wrong track. Perhaps (some, or many) people do spend time thinking about the mental states of others; but if so, they do not simply try to predict the first option that comes to mind. They are more sophisticated: they look for an option that they think

will help solve the very specific interactive problem they are facing (see also Sugden and Zamarròn 2006: 614–5).

Another way to put it is to say that Schelling salience is a form of *engaged* reasoning. It presupposes that the players take a first-person perspective, the perspective of someone who is participating in a joint task. In contrast, primary salience is the perspective of a solipsistic decision-maker, and secondary salience is the perspective of an observer who is trying to predict the behaviour of a solipsistic decision-maker. But contrary to cognitive hierarchy theory, few people seem to take such a detached perspective when they are playing coordination games.

9.5 Team Reasoning

Schelling says that coordination requires a 'meeting of minds' (1960: 83). It is tempting to infer from remarks like this that the reasoning that takes place during coordination is a kind of *collective* reasoning, in which Ann and Bob jointly identify the solution that is best 'for them'. Several philosophers and social scientists over the last two decades have proposed theories that develop this idea in various ways. Some theories start from the premise that people are able to think in 'collective mode'. They can, for example, represent a problem of coordination such as Hi-lo as a situation in which *they* prefer, as a group, the Hi equilibrium to the Lo equilibrium, and *they* think that choosing Hi is the best thing to do.[10]

Most theories of this kind focus on already formed intentions to act, sidestepping the reasoning processes that may lead to their formation. A notable exception is *team reasoning*, a theory of coordination that has been developed by Michael Bacharach, Robert Sugden, and Natalie Gold.[11] According to Gold and Sugden, the theory

> seeks to extend standard game theory, where each individual asks separately "What should I do?" to allow teams of individuals to count as agents and for players to ask the question "What should we do?" This leads to team reasoning, a distinctive mode of reasoning that is used by members of teams, and which may result in cooperative actions. (Gold and Sugden 2007: 110–111)

The key idea is that mental states formulated in we-mode facilitate the solution of problems of coordination that would not be satisfactorily addressed within an individualistic framework. Hi-lo is a typical example. Team reasoning theorists argue that the game looks trivial when it is perceived as a parametric problem for a single agent, the team constituted by the two players, rather than as a problem of strategic interaction. The first step to solve the problem, then, is to look at it from the point of view of a single agent, the team.

[10] See e.g. Gilbert (1989), Hurley (1992), Searle (1990), Tuomela (1995, 2002), Bratman (2014).

[11] See Sugden (1993, 2000, 2003), Bacharach (2006), Gold and Sugden (2007).

	Hi	Lo
Hi	4	0
Lo	0	2

Fig. 9.2 Hi-lo seen from the point of view of a team

The team must have a single preference ordering about the outcomes. To make the point in the simplest manner, let us suppose that team payoffs are obtained by simply adding the payoffs of the two players. From the perspective of the team, Hi-lo then looks like the transformed matrix in Fig. 9.2.

The transformation of payoffs is a *framing* process. Team reasoning theorists have offered different accounts of this process, which I will skip for the purposes of this paper.[12] What matters is that once the game has been framed as a problem for the team, rather than for two separate individuals, the coordination problem becomes trivial.

The intention to choose Hi can be derived by means of an inference from premises formulated in we-mode. The conclusion is a set of prescriptions for the team members:

(T1)

1. Our goal is to maximize the team's payoffs.
2. T is the obvious way to achieve this goal.
3. T implies that I choose Hi and you choose Hi.
4. I will choose Hi and you will choose Hi.

I have called 'T' the solution of the maximization problem for the team. T is a profile of strategies (a set of actions), in this case (Hi, Hi). The inference T1 is a simple case of instrumental reasoning. It is a valid inference, because the coordination problem has been transformed from a strategic into a parametric form. Since there is only one decision-maker (the team), it is not necessary to consider the beliefs of other individual players. What the team members ought to do – their rational actions – can be deduced from the goal and from the constraints the team is facing. The only premise that we need is that all players are reasoning as team members.

Notice that although beliefs do not figure as inputs of the reasoning process, they can in principle be derived using this inferential scheme. For example: Ann could predict that Bob will choose Hi, and that Bob believes that she will choose Hi. This only requires an attribution of instrumental rationality, and the premise that Ann and Bob are reasoning as team members. Notice that the procedure is opposite to the one

[12] According to Bacharach (2006), the process is automatic and largely sub-conscious. According to Sugden (2011), each individual can make a deliberate decision whether to reason in team mode or not.

followed in standard game-theoretic reasoning: it is not the case that we coordinate because we have certain expectations; we have certain expectations because we want to coordinate.

9.6 Solution Thinking

Collective intentions, desires, and preferences (formulated in 'we-mode') have attracted a lot of attention in the philosophical literature. This is understandable, for they seem to imply a major departure from the individualism of standard rational choice theory. Nonetheless, they do not constitute the most significant aspect of team reasoning theory, in my view. I suggest that we should move the spotlight onto a different aspect of the theory – away from its content and onto its *form*. This is not entirely original. According to Elisabeth Pacherie, for example, "team reasoning is simply a matter of using certain patterns of inference" (2013: 1834).[13] If this is right, then how we formulate the premises of team reasoning is of secondary importance. What is essential, in order to solve coordination problems, is the way in which goals, means, and beliefs are arranged in the inferential scheme.

I will explain this point gradually. First, I will illustrate a theory that exploits essentially the same pattern of inference as team reasoning, but without using we-mode premises. The theory goes under the name of 'solution thinking', and has been proposed by Adam Morton (2003). Then, I will generalise from team reasoning and solution thinking, with the aim of identifying the pattern of inference that allows, in both cases, to solve problems of coordination such as Hi-lo.

Morton's idea is that in many interactive situations we achieve coordination by asking: what is the easiest or most natural way to tackle this problem? What is the obvious solution? If there is a clear answer, then the same reasoning is attributed to the other player by default.

> One first thinks of an outcome which one can imagine the other person or persons both would want to achieve and would believe that one would try to achieve. One then thinks out a sequence of actions by all concerned that will lead to it. Lastly, one performs the actions that fall to one's account from this sequence […] and expects the other(s) to do their corresponding actions. (Morton 2003: 120)

The reasoning process may be reconstructed as follows:

(S1)

1. My goal is to maximize my payoff, your goal is to maximize your payoff.
2. S is the obvious way to achieve these two goals.
3. S implies that I choose Hi and you choose Hi.
4. I will choose Hi and you will choose Hi.

[13] See also Pacherie (2018). I have discussed the similarities between team reasoning and solution thinking in Guala (2016, 2018).

I have formulated S1 using the same format as T1, to highlight the similarity with the inferential pattern of team reasoning (in the previous section). What used to be called 'T' is now called 'S', i.e. the (Hi, Hi) equilibrium or solution. Since the payoffs of the two players are perfectly aligned in the Hi-lo game, team reasoning delivers the same result as simulation thinking. But this is not a matter of coincidence, for these two reasoning strategies have very similar features. In both cases the inference proceeds backwards from the identification of a solution, to the individual strategies that help attain that solution, and in both cases the game is transformed into a parametric decision. Moreover, in both cases belief attribution plays no role: like team reasoning, solution thinking is an example of *belief-less coordination*.

In team reasoning T is the obvious solution because it is the best outcome for the team. The rational solution is the focal point. In simulation thinking S is the obvious solution because it is the only outcome that maximizes my payoff and your payoff simultaneously. It is the instrumentally rational solution to the problem. But notice that if beliefs are ignored, whether we describe the goals in a collectivistic or an individualistic manner (in the first premise) is irrelevant. In the end, the style of reasoning is what really matters for coordination; whether we start from an "I" or a "we" is a secondary detail.

One might say that my diagnosis is slightly unfaithful to Morton's theory, at this point. Although they do not play a key role, beliefs play *some* role in solution thinking. Morton is one of the pioneers of 'simulation', an approach to mindreading that has been extremely influential in contemporary research on folk psychology. According to the simulation theory of mindreading, when we attribute mental states to other agents we do not apply a theoretical framework with preferences and beliefs as main variables. We rather replicate the reasoning of other agents using ourselves (our own cognitive apparatus) as a simulation device.[14] Solution thinking, says Morton, is a special kind of simulation:

> First of all it results in an understanding of others: you have an idea what the other is going to do, and you have the materials for putting together an account of why they might do it. And the understanding is got by reproducing rather than representing their thinking. (Morton 2003: 120)

To capture Morton's claim that solution thinking is simulation, the scheme may be modified as follows:[15]

(S2)

1. My goal is to maximize my payoff, your goal is to maximize your payoff.
2. S is the obvious way to achieve these goals.
3. You also think that S is the obvious way to achieve these goals.
4. S implies that I choose X and you choose Y.
5. You also think that I must choose X and you must choose Y.
6. I will do X and you will do Y.

[14] See e.g. Morton (1980), Gordon (1986), Goldman (1989).

[15] This is a modified version of the scheme in Guala (2016: 97). That scheme suffers from a confusion between means and goals that I have tried to correct here.

Simulation occurs in the third and in the fifth steps. In steps one and two, I look at the problem and identify the solution. Step three replicates the procedure for the other player: she identifies the same solution because she is just like me. But once the solution has been identified, I can derive my own actions and the actions of the other player (step four). Using the same procedure ('she reasons in the same way') finally I can predict what she will do and what she believes that I will do (steps five and six).

Morton says that solution thinking is a *special* kind of simulation. What makes it distinctive is that people do not try to identify the beliefs of the other players at the outset, as a preliminary step to the identification of the equilibrium of the game. Rather, they first identify the solution, and then derive players' beliefs from it. In this sense, solution thinking breaches a key assumption of traditional folk psychology and epistemic game theory, namely, that behaviour is predicted from beliefs and desires. Instead, in solution thinking beliefs are inferred from goals and means.

But this entails that beliefs are idle in S2. Steps 3 and 5 are unnecessary. Coordination may be attained using a much leaner reasoning scheme, like S1, where there is no belief attribution at all: the players may well be *reasoning without beliefs*.

I suggest that belief-less reasoning does not betray Morton's main insight. Morton argues (correctly) that the attribution of beliefs is often an effect of our desire to coordinate. In many circumstances we do not coordinate because we have the same beliefs, but rather we have the same beliefs (we attribute one another the same mental representations) because we want to coordinate. The trick is to infer others' beliefs from a common target and from some features of the environment, assuming tacitly that the other players are just like us. Beliefs then will mesh: everyone will end up attributing to others the same beliefs that she attributes to herself.

Still, the point is: why should beliefs be attributed at all? Beliefs can be inferred by simulation thinking (from our desire to coordinate) *if it is necessary*. But for the sake of coordination it is probably unnecessary. Belief-less reasoning will do the job just as well. This claim clearly is at odds with traditional views about the function of beliefs, which have dominated philosophy and rational choice theory for decades. It will be necessary then to reassess the role of beliefs, with a little help from recent research on folk psychology.

9.7 Coordination Without Beliefs

Empirical evidence and theoretical considerations suggest an intriguing hypothesis: perhaps people coordinate smoothly in everyday interactions because they do not reason about beliefs at all. The coordination that takes place in ordinary circumstances may be *belief-less coordination*.

This suggestion may sound incredible, especially to those who have been trained in the tradition of epistemic game theory. But those philosophers who accept the belief-desire model of folk psychology will also find it odd. In the remaining sections of this paper I will try to make the idea of belief-less coordination appear less idiosyncratic, showing first of all that it does not come out of the blue. On the

contrary, it is consistent with an important body of research that is making headways in cognitive science and the philosophy of mind. This research programme challenges a fundamental assumption of earlier theories of mindreading, namely, that we typically explain and predict behaviour by attributing 'full blown' beliefs and desires. Following the terminology of this volume, I will call it the 'minimal mind reading' (MMR) programme.[16]

What is a 'full blown' belief or desire, to begin with? There is a core of characteristics that philosophers and psychologists typically associate with such mental states: a belief (desire) is (i) a propositional attitude, (ii) that causes behaviour; (iii) that is directly unobservable; and (iv) that is holistically related with other unobservable mental states. Propositional attitudes are mental states directed toward the content of propositions – such as, for example, a desire that a certain event occurs ("John wants to have some cash in his pockets"). The holism principle states that the content of a mental state depends on its role or place in a network that includes other mental states, and implies as a consequence that it is impossible to identify a single mental state from behaviour alone. We cannot, for example, infer that John wants to have some cash from the fact that he is going to the bank, unless we assume that John believes that there is money in the bank and that he can withdraw it using the ATM.

Notice that this approach clearly sets the bar high for folk psychology: if beliefs and desires are defined this way, then many interpretations and predictions of everyday behaviour do not seem to require the attribution of beliefs and desires. For example: we can infer that Bob is looking for an ATM from the fact that Bob has run out of cash. We do not have to think explicitly that Bob has formed a representation (a belief) of the state of affairs 'there is no money in my pocket', of the hypothetical state 'there is cash in the ATM', or of the counterfactual state 'if I withdraw cash, then there will be money in my pocket'. Similarly, we can predict that Ann tomorrow will go to London, from the fact that her boss has scheduled a meeting, without attributing full-blown beliefs and desires to Ann.

One may say that that we *implicitly* attribute these mental states to Bob and Ann when we predict their behaviour. But this temptation should be resisted. Although any prediction can be enriched and rationalized a posteriori, there is no evidence that these rationalizations are faithful reconstructions of what goes on when we interact seamlessly and spontaneously. In fact, we have good reasons to believe that belief-desire accounts of behaviour are radically *un*faithful accounts of our ordinary, spontaneous mindreading practices.

We do not engage in belief-desire attribution on a routine basis, partly because it is unnecessarily complicated. Inferring behaviour from sheer factual knowledge (Bob has run out of cash; a meeting has been scheduled in London) allows to make predictions more efficiently than engaging in full mental state attribution. In fact, as

[16] I take MMR to be a cluster of approaches sharing a family resemblance, rather than a unified body of theory. Relevant contributions include e.g. Gallagher (2001), Gallagher and Hutto (2008), Vesper et al. (2010), Andrews (2012), Butterfill and Apperly (2013), Pacherie (2013), Fiebich and Coltheart (2015).

every teacher of game theory knows, most students find reasoning about others' beliefs and desires extremely counterintuitive and difficult. And in most cases such reasoning is not even worth the effort: fast-and-frugal heuristics work well enough. Actually, in situations like Hi-lo, belief-desire attributions are *harmful*: they create inferential loops that can block us in situations that seem otherwise unproblematic. All things considered, it is not surprising that we are not 'natural mindreaders' in the belief-desire sense.

Fiebich and Coltheart (2015) state this point using the familiar distinction between System 1 and System 2 reasoning processes.[17] Most everyday social interaction is governed by fast, unreflective, computationally cheap cognitive processes (System 1). Belief-desire attributions in contrast are involved in theoretically complex, computationally more expensive forms of reasoning (System 2). In general, a MMR approach does not need to deny that people engage in System 2 reasoning – MMR is fully compatible with a pluralistic stance on this matter.[18] But if belief-desire mindreading is not our default form of reasoning, then it is probably activated only in peculiar and relatively rare circumstances.

According to an important strand of MMR theory, the main function of belief-desire accounts is not to predict or explain behaviour. Belief-desire attributions play a totally different role: their main function is to *justify* behaviour, by attributing to agents a set of *reasons* for doing what they did.[19] A prominent piece of evidence in favour of this hypothesis is that beliefs and desires are typically invoked in anomalous, problematic situations – when Bob returns from the ATM with empty pockets, for example; or when we discover that Ann has just missed an important meeting with her boss. In these occasions we feel that it is important to say that Bob *believed* that there was cash in the machine, to make sense of his behaviour. Similarly, we feel compelled to say that Ann *thought* that the meeting had been re-scheduled, to justify her being tardy. But in routine situations, in contrast, attributions of 'full blown' beliefs are rare. What is Bob doing now? He's going to the bank because he has no cash. Where will Ann be tomorrow morning? She will go to London because she has a meeting with her boss. No 'full-blown' belief attribution is deemed necessary in cases of this sort.

9.8 Thinking with Rationality Principles

So far I have treated beliefs and desires symmetrically – as if they played similar roles in folk psychology. But casual observation suggests that goals, intentions, desires, play a more prominent role in our everyday reasoning about behaviour,

[17] See e.g. Evans (1984), Sloman (1996), Kahneman (2011).

[18] See, again, Andrews (2012), Fiebich and Coltheart (2015).

[19] The idea that attributions of 'full blown' mental agency play primarily a normative, justificatory function can be traced back to Bruner (1990). This insight has been further developed and articulated by McGeer (1996, 2007), Hutto (2008), Zawidzki (2008, 2013), Andrews (2012, 2015).

compared to beliefs. Consider Bob once again: to predict what he is about to do (going to the bank), it is just as likely that we say that 'he has run out of cash' (describing an objective state of affairs) or that 'he wants to get some cash' (mentioning a subjective desire). His belief that there is an ATM at the bank (which is probably full of cash) is not worth mentioning and is rarely mentioned in ordinary situations of this kind.

Generalizing, we may say that in normal circumstances beliefs tend to stay in the background. To anticipate people's behaviour, it is usually sufficient to monitor the environment and to assume that everyone will do what is appropriate in the circumstances. If explicit attributions of beliefs become necessary, then we often move from the tacit assumption that the other agents know more or less what we know. We often apply the simple principle that 'seeing is believing', for example. But in most cases we just skip this intermediate step and infer behaviour from the circumstances. Such an inference is fallible, of course, like any other heuristic. But it works most of the time, sparing the trouble of attributing beliefs from a limited informational basis.

This heuristic works mainly because the external world is one and the same for everyone, and a lot of information about the external world is shared. In contrast, people's desires are not only private, but are also diverse enough to make it difficult to infer them directly from the environment (see e.g. Steglich-Petersen and Michael 2015). What Bob will do next depends very much on his desires: maybe he does not want to look for cash; if we see him going to the bank, we cannot immediately tell whether he intends to deposit a cheque, close his bank account, or apply for a loan – who knows?

Another reason why we pay more attention to desires is that in order to *interact* successfully with others (as opposed to merely observe their behaviour) it is crucial to figure out whether our goals are aligned or not. If there is conflict (or just mismatch) between desires, we need to make sure that the conflict is resolved prior to action; otherwise, we must switch from cooperative to Machiavellian reasoning, which requires a completely different attitude. This of course does not apply only to desires: *all* conflicting mental states are problematic for coordination. The point is merely that in ordinary situations – for the reasons mentioned earlier – it is more likely that there is a conflict between desires, rather than beliefs.

The desire-belief asymmetry has a *developmental* aspect.[20] Children seem to develop the capacity to attribute conative states (goals, intentions, desires) earlier than the capacity to attribute beliefs. This asymmetry may persist in adulthood in a different form – as an augmented fluency and automaticity in desire-attribution than in belief-attribution (Apperly et al. 2006; Steglich-Petersen and Michael 2015). Although children and adults are able to engage in full-blown belief attribution roughly from age six, they do it only when it is worthwhile – when Machiavellian reasoning is required, or because other interpretative methods fail.

Experiments on infants suggest that our earliest folk-psychological procedures exploit inferential patterns that are remarkably similar to the ones we use later in

[20] See Wellman and Woolley (1990), Perner (1991).

adulthood. In a number of papers, backed up by a large amount of experimental data, Gergely Csibra and Gyorgy Gergely have argued that pre-verbal infants interpret and predict others' behaviour by means of a strategy called *teleological thinking* (see e.g. Csibra and Gergely 1998; Gergely and Csibra 2003).

Because infants below the age of two cannot communicate verbally, their interpretative procedures must be elicited using ingenious experimental designs. A widely used technique consists in measuring the amount of time an infant spends looking at an event or situation, as an indirect estimate of her expectations. (Longer looking times reflect surprise when the infant is faced with an unexpected event.) The experiments of Csibra and Gergely – and of many other psychologists – suggest that infants are more surprised when people perform inefficient actions, i.e. do not choose the best (simplest, most direct) means to attain a given outcome. Infants, in other words, predict behaviour using a simple efficiency principle: they assume that people will perform the most direct, least expensive action to achieve their goal. Since this interpretative strategy appears earlier than infants' capacity to form full-blown beliefs, Csibra and Gergely conclude that teleological thinking is a minimal technique that precedes and perhaps lays the foundations for full-blown mindreading.

Csibra and Gergely draw a distinction between goals on the one hand, and desires or intentions on the other. While a desire is a cause (i.e. an antecedent condition) of behaviour, a goal is an outcome (i.e. a consequence) of behaviour. While a desire is a subjective state of the agent, a goal is an objective state of the world. Being a state of the world, a goal is more directly and easily represented than a desire (which is a state about a state, i.e. has propositional content). If they are right, then the efficiency principle may be employed in the context of an entirely non-mentalistic reasoning strategy, which exploits purely behavioural patterns and does not involve the attribution of any subjective mental state at all.

The claim that infants employ non-mentalistic strategies to anticipate people's actions is currently controversial. Some psychologists are willing to attribute mindreading capacities to infants at a much earlier stage of development – before 18 months and in some cases as early as 6 months of age (Baillargeon et al. 2016). If this were true, then it would be appropriate to speak of a 'rationality principle', rather than merely of an efficiency principle: predictions would be based on a capacity to identify people's representations of their goals – rather than mere states of affairs.

Be that as it may, we do not have to take a position on this controversy here. The crucial point is that beliefs do not necessarily play a role when we think with teleological or with rationality principles: the means can be identified looking directly at the environment; mental states (people's representations of the environment) can then be inferred cheaply using seeing-is-believing heuristics. But such mental state attributions are often unnecessary – what matters is that we correctly identify the means, that is, the 'obvious' (objectively efficient) ways to achieve the goals.

In light of all this, it seems likely that thinking with rationality principles is an important skill for adults and infants alike. Even when it is complemented by full-blown mindreading, it is probably never replaced completely: it survives as a heuristic that governs social interactions in ordinary, unproblematic situations (see e.g. Fiebich 2014). This is all the more plausible if, as we have seen, there are situations

in which social interaction is *hampered* by full blown mindreading. It is precisely in coordination games that we should expect belief-less thinking to be used most frequently by adults – because it is most effective in those situations.

9.9 To Sum up

Let us return to the Hi-lo game. The goal, recall, is to maximize individual payoffs. In theory, if Ann and Bob were full-blown mindreaders, they should try to identify each other's beliefs, in order to figure out what are the best means to attain their goals. But this leads nowhere, as we have seen. A more effective procedure is to ignore beliefs altogether, and ask what is the most rational way to attain the goal – which strategy is most efficient. In Hi-lo, the means and goals of the two players happen to coincide, so the profile of strategies (Hi, Hi) is the obvious solution. From this, it follows that Ann should choose Hi and Bob should do the same.

This is the same pattern of reasoning that we have detected in team reasoning and solution thinking. Thinking with rationality principles is more general, to be sure – it applies to spectatorial as well as to interactive mindreading. In both cases it simplifies the inference from goals and environment to behaviour. But in interactive situations it also allows to bypass the problem of belief formation. Thinking about what Bob has in mind does not always improve Ann's interpretive and predictive skills. In Hi-lo, thinking about what Bob believes that Ann believes that Bob believes (etc.) would positively *harm* their capacity to coordinate. It is much better if each one of them focuses on the environment and the way it constrains their goals, without thinking too much about the way the other thinks (and thinks they think …) about the coordination problem.

Acknowledgments I have presented this paper at the University of Rostock and at the University of Milan. I'm grateful to seminar participants and to Anika Fiebich for making detailed comments that have helped me to improve the paper. All the remaining mistakes are obviously mine.

References

Andrews, K. (2012). *Do apes read minds?* Cambridge, MA: MIT Press.

Andrews, K. (2015). The folk psychological spiral: Explanation, regulation, and language. *Southern Journal of Philosophy, 53*, 50–67.

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science, 17*, 841–844.

Bacharach, M. (1999). Interactive team reasoning: A contribution to the theory of co-operation. *Research in Economics, 53*, 117–147.

Bacharach, M. (2006). *Beyond individual choice*. Princeton: Princeton University Press.

Bacharach, M., & Bernasconi, M. (1997). The variable frame theory of focal points: An experimental study. *Games and Economic Behavior, 19*, 1–45.

Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology, 67*, 159–186.
Bratman, M. E. (2014). *Shared agency*. Oxford: Oxford University Press.
Bruner, J. S. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.
Butterfill, S., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language, 28*, 606–637.
Camerer, C. F., Ho, T. H., & Chong, J. K. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics, 119*, 861–898.
Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science, 1*, 255–259.
Evans, J. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 75*, 451–468.
Fiebich, A. (2014). Mindreading with ease? Fluency and belief reasoning in 4- to 5-year-olds. *Synthese, 191*, 929–944.
Fiebich, A., & Coltheart, M. (2015). Various ways to understand other minds: Towards a pluralistic approach to the explanation of social understanding. *Mind & Language, 30*, 235–258.
Gallagher, S. (2001). The practice of mind: Theory, simulation, or primary interaction? *Journal of Consciousness Studies, 8*, 83–108.
Gallagher, S., & Hutto, D. (2008). Understanding others through primary interaction and narrative practice. In J. Zlatev et al. (Eds.), *The shared mind* (pp. 17–38). In: John Benjamins Publishing.
Gauthier, D. (1975). Coordination. *Dialogue, 14*, 195–221.
Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Science, 7*, 287–292.
Gigerenzer, G., & Selten, R. (Eds.). (2002). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.
Gilbert, M. (1989). *On social facts*. London: Routledge.
Gold, N., & Sugden, R. (2007). Collective intentions and team agency. *Journal of Philosophy, 104*, 109–137.
Goldman, A. I. (1989). Interpretation psychologized. *Mind & Language, 4*, 161–185.
Gordon, R. M. (1986). Folk psychology as simulation. *Mind & Language, 1*, 158–171.
Guala, F. (2016). *Understanding institutions*. Princeton: Princeton University Press.
Guala, F. (2018). Coordination, team reasoning, and solution thinking. *Revue d'économie politique, 128*, 355–372.
Hodgson, D. H. (1967). *Consequences of utilitarianism*. Oxford: Clarendon Press.
Hurley, S. (1992). *Natural reasons*. Oxford: Oxford University Press.
Hutto, D. (2008). *Folk psychological narratives*. Cambridge, MA: MIT Press.
Kahneman, D. (2011). *Thinking, fast and slow*. London: Macmillan.
Lewis, D. K. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.
McGeer, V. (1996). Is 'self-knowledge' an empirical problem? Renegotiating the space of philosophical explanation. *Journal of Philosophy, 93*, 483–515.
McGeer, V. (2007). The regulative dimension of folk psychology. In D. Hutto & M. Ratcliffe (Eds.), *Folk psychology re-assessed* (pp. 137–156). Dordrecht: Springer.
Mehta, J., Starmer, C., & Sugden, R. (1994). The nature of salience: An experimental investigation of pure coordination games. *American Economic Review, 84*, 658–673.
Morton, A. (1980). *Frames of mind*. Oxford: Clarendon Press.
Morton, A. (2003). *The importance of being understood*. London: Routledge.
Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review, 85*, 1313–1326.
Pacherie, E. (2013). Intentional joint action: Shared intention lite. *Synthese, 190*, 1817–1839.
Pacherie, E. (2018). Solution thinking and team reasoning: How different are they? *Philosophy of the Social Sciences, 48*, 585–593.

Pacuit, E., & Roy, O. (2015). Epistemic game theory. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. URL: https://plato.stanford.edu/archives/sum2017/entries/epistemic-game/.
Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.
Schelling, T. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.
Searle, J. (1990). Collective intentions and actions. In P. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication*. Cambridge, MA: MIT Press.
Simon, H. A. (1982). *Models of bounded rationality*. Cambridge, MA: MIT press.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3–22.
Stahl, D. O., & Wilson, P. W. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior, 10*, 218–254.
Steglich-Petersen, A., & Michael, J. (2015). Why desire reasoning is developmentally prior to belief reasoning. *Mind & Language, 30*, 526–549.
Sugden, R. (1993). Thinking as a team: Toward an explanation of nonselfish behavior. *Social Philosophy and Policy, 10*, 69–89.
Sugden, R. (2000). Team preferences. *Economics and Philosophy, 16*, 174–204.
Sugden, R. (2003). The logic of team reasoning. *Philosophical Explorations, 6*, 165–181.
Sugden, R. (2011). Mutual advantage, conventions and team reasoning. *International Review of Economics, 58*, 9–20.
Sugden, R., & Zamarròn, I. E. (2006). Finding the key: The riddle of focal points. *Journal of Economic Psychology, 27*, 609–621.
Tuomela, R. (1995). *The importance of us*. Stanford: Stanford University Press.
Tuomela, R. (2002). *The philosophy of social practices*. Cambridge: Cambridge University Press.
Vesper, C., Butterfill, S., Koblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks, 23*, 998–1003.
Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition, 35*, 245–275.
Zawidzki, T. W. (2008). The function of folk psychology: Mind reading or mind shaping? *Philosophical Explorations, 11*, 193–210.
Zawidzki, T. W. (2013). *Mindshaping*. Cambridge, MA: MIT Press.

Chapter 10
Proprietary Reasons and Joint Action

Abe Roth

Abstract Some of the reasons one acts on in joint action are shared with fellow participants. But others are proprietary: reasons of one's own that have no direct practical significance for other participants. The compatibility of joint action with proprietary reasons serves to distinguish the former from other forms of collective agency; moreover, it is arguably a desirable feature of joint action. Advocates of "team reasoning" link the special collective intention individual participants have when acting together with a distinctive form of practical reasoning that purports to put individuals in touch with group or collective reasons. Such views entail the surprising conclusion that one cannot engage in joint action for proprietary reasons. Suppose we understand the contrast between minimal and robust forms of joint action in terms of the extent to which participants act on proprietary reasons as opposed to shared reasons. Then, if the team reasoning view of joint intention and action is correct, it makes no sense to talk of minimal joint action. As soon as the reason for which one participates is proprietary, then one is not, on this view, genuinely engaged in joint action.

Keywords Joint action · Proprietary reasons · Group reasons · Acting for a reason · We-mode · Team reasoning

10.1 Proprietary Reasons

Joint action involves the sharing of reasons and aims amongst those engaged in it. This is a natural thought, at least if we understand intentional action in terms of the reasons or aims that rationalize or make sense of it.[1] For example, the raising of my arm is intentional insofar as it is done toward some end, or for a reason – such as

[1] Davidson (1980), Anscombe (1963).

A. Roth (✉)
Department of Philosophy, Ohio State University, Columbus, OH, USA
e-mail: roth.263@osu.edu

A. Fiebich (ed.), *Minimal Cooperation and Shared Agency*, Studies in the Philosophy of Sociality 11, https://doi.org/10.1007/978-3-030-29783-1_10

getting the attention of the speaker in order to ask a question. The same would seem to be true for joint or collective action: our φ-ing is intentional insofar as it is directed toward an end, or done for a reason. And since we're φ-ing *together*, the reason or aim that makes the φ-ing intentional presumably is something that we share. Thus, there is in joint action a meeting of minds, a sharing of the reasons and goals that makes what we're doing together the intentional act that it is.

But how much sharing of reasons can there be? More importantly for my purposes here, how much sharing *must* there be? How little overlap or convergence of reasons is necessary for joint action of a minimal sort?

You might think that some of the *subsidiary* aims that an individual has in connection with a joint endeavor – those she adopts as part of how she goes about making her contribution to it – must inevitably be shared with fellow participants. For example, if we share a ride downtown, we cannot each aim to get there using incompatible routes. At least, we would seem to be subject to some sort of rational criticism were we engaged in the joint action and yet failing to reconcile our differences about how to go about it.[2]

What about our reasons or aims for going downtown in the first place? It would seem that here there is room for significant divergence of aims. Maybe I want to get to the museum, and you want to get to the ballpark, both of which are downtown. Sharing a ride doesn't seem to require that we share reasons for doing so. Call a reason or aim had in connection with engaging in joint action *proprietary* when it is not shared with other participants. My proprietary reason (getting to the museum) has immediate or direct normative significance for what I do; it is, after all, my reason for action. But insofar as it is proprietary, it doesn't have that sort of immediate significance for you.[3]

Joint action does not preclude proprietary reasons on the part of some or all of the participants. Bratman articulates this point forcefully when he says

> The sharing of intention need not involve commonality in each agent's reasons for participating in the sharing. You and I can have a shared intention to paint the house together, even though I participate because I want to change the color whereas you participate because you want to remove the mildew…though we participate for different reasons, our shared intention nevertheless establishes a shared framework of commitments; and this can happen even if these differences in our reasons are out in the open…much of our sociality is *partial* in the sense that it involves sharing in the face of – in some cases, public – divergence of background reasons for the sharing. (Bratman 2014, 91)

Bratman points out that such a conception of shared intention and joint action resonates with a liberal pluralism:

> It is an important fact about our sociality that we manage to share intentions and act together in the face of substantial differences of reasons for which we participate. We work together,

[2] We might allow for what Bratman (1992) calls the mesh of sub-plans. Participants need not have the same sub-aims, so long as carrying out the differing sub-plans is compatible with the success of both.

[3] Of course, it might be relevant in other ways. For example, it might have a moral significance for you; and that's something that you would presumably have to take into consideration.

> we play together, and we engage in conversations together even given substantial background differences in our reasons for participation and our reasons for various sub-plans. This is especially characteristic of a pluralistic, liberal culture. This is the *pervasiveness of partiality in our sociality*. It is a virtue of the basic thesis that it makes room, in a theoretically natural way, for this pervasiveness of partiality. (Bratman 2014, 91)

From a more narrowly action-theoretic perspective, it seems that the compatibility of proprietary reasons with joint action is a mark of the phenomenon that is our focus: that of distinct individual agents coming together to act jointly. It stands in contrast to a form of agency exercised by some highly integrated group – something that is arguably more like an individual agent, albeit on a large scale. Precluding proprietary reasons would, if anything, undermine joint action and leave in its stead some more monolithic form of individual agency.

So there are things to be said in favor of the compatibility (and perhaps even the necessity) of proprietary reasons for joint action. These reasons notwithstanding, I think it will be illuminating to explore some considerations that could lead us to think otherwise. Might there be at least some tension between joint action and the presence of proprietary reasons?

10.2 Against Propriety

For a start, consider how Tuomela criticizes Bratman for neglecting the *collective* reason that is necessary for joint action. He says (2007, 100–101) that on Bratman's analysis "The agents have their personal intentions on the basis of their private reasons." Tuomela objects that this "does not entail a full-blown group reason, namely, that the agents have their intentions of the form "I intend that we J" (for each participant) necessarily because of a group reason satisfying the Collectivity Condition." Tuomela says "Formulated for the special case of goal satisfaction, [the Condition] necessarily connects the members as follows. Necessarily (as based on group construction of a goal as the group's goal), the goal is satisfied for a member if and only if it is satisfied for all other members." (Tuomela, 4).

It's clear enough that Tuomela objects to the absence in Bratman's account of a common or shared reason as a requirement on joint action; indeed, Tuomela speaks of a 'group reason' requirement on joint action (4, 13). It's less clear, however, what his attitude is toward proprietary reasons and whether their presence is somehow antithetical to joint action. Do Tuomela's group reason and collectivity conditions entail anything about what proprietary reason an agent may or may not have? Must proprietary reasons be set aside in the 'we'-mode, which for Tuomela is the perspective of participants in robust, paradigmatic forms of joint agency? Or does adopting the perspective of the we-mode simply add group considerations to the personal or proprietary reasons one may already have for participating? How easily do shared or collective reasons for our φ-ing sit along side of proprietary reasons for engaging in joint action?

Some of what Tuomela says suggests that proprietary reasons are *not* compatible with joint action. Thus, his collectivity condition indicates that "the depersonalization that occurs in social groups shows up in the basic structure of group life (in its we-mode content) and thus is not a mere contingent feature of groups" (Tuomela 2007, 50; see also 4, 10, 48). It's not entirely clear what is meant by 'depersonalization', but it connotes that personal or proprietary reasons are distinct from group reasons (4), and set aside in joint action. If so, Tuomela would be insisting not only on the necessity of group reason, but also the exclusion of proprietary reasons. Tuomela confirms this sort of reading with a passage that contrasts someone acting in we mode and hence for some collective reason with someone who outwardly behaves that way but acts for a very different sort of reason. Thus,

> "a member can instrumentally function (or, better, quasi-function) as a group member just by doing his "work," even if not for the ethos-serving reason (that satisfies the Collectivity Condition). He would then, so to speak, be exhibiting the right actions, but his *reason* for performing a group-task T would be, roughly, the I-mode reason expressible by "I perform T because it is conducive to my personal interests" and not the we-mode reason expressible by "I perform T at least in part because it is my duty and furthers, or at least does not contradict, the group's ethos." (Tuomela 2007, 50)

What seems to be important for Tuomela is that in *paradigmatic* joint action, one is *acting for* a group reason. And this is contrasted with other cases where one is acting for some personal reason. Perhaps one can possess personal reasons, but as far as robust forms of joint action are concerned, one is not acting for them and but from the group reason instead.

But, having put forth the strong thesis seemingly excluding acting from proprietary reasons in paradigmatic joint action, Tuomela concedes that "in real life one often acts both for group reasons and private reasons when performing an action" (98; see also 130). What are we to make of this? Is the suggestion that group action is compatible with proprietary reasons? Or is it rather that there is a spectrum of cases? On the latter proposal, there are many instances of joint action involving both group and proprietary reasons, and to the extent that the motives are mixed, the case is not one of fully collective or joint action. If that's the view, then it's not quite a compatibilist thesis. Tuomela says,

> To think (e.g., believe, intend) or act in the we-mode is to think or act as a group member in a full sense, thus for a group reason. Thinking and acting in the we-mode expresses collective intentionality in its full sense. In contrast, to think or act in the I-mode is to think or act as a private person—even if a group reason might contingently be at play. (Tuomela 2007, 7)

It's the group reason and *not* the proprietary reason that has authority over what the agent does in joint action:

> Thinking and acting in the we-mode basically amounts to thinking and acting for a group reason, that is, to a group member's taking the group's views and commitments as his authoritative reasons for thinking and acting as the group "requires" or in accordance with what "favors" the group (namely, its goals, etc.). (Tuomela 2007, 14)

> …the we-mode was said to involve the idea of one's functioning because of a group reason *rather than* a private reason. (Tuomela 2007, 47 emphasis added; see also a passage on 44-5 that similarly draws a contrast between acting on group vs. private reasons.)

For Tuomela, then, any concession that allows for proprietary reasons would seem to be a falling away from some ideal of joint action. This is not really a *proper* compatibilism of the sort we find in Bratman – one that allows that some element of proprietary reason that *does not in any way* compromise joint action. (Indeed, a fully compatibilist view would insist that without some proprietary reason, there is the risk of losing the sense that we*,* a *plurality*, share agency; joint action is not the doings of some monolithic individual agent.) Tuomela, in contrast, seems to favor a mixed view that allows that one might have proprietary reasons, but to the extent that one does, the action is less than fully joint.[4]

What is it about a proper compatibilism that might put one off? Why does Tuomela think that to the extent that one is acting for the group reason, this is *at odds* with acting on a proprietary reason? There doesn't seem to be anything in the very notion of acting for a reason that would rule out acting for *multiple* reasons. Why couldn't one, in a case of joint action, have a proprietary reason, as well as a group reason, and act for both of them? To get a better sense of the sort of consideration that might move one away from compatibilism, let me turn to team reasoning.

10.3 Incompatibilist Team Reasoning

Team reasoning provides a solution to certain games (especially Hi-Lo) that pose difficulties for orthodox game theory. The details have been discussed extensively; what I want to emphasize here is that the proposed solution involves departing from the standard view that assumes that the only practical or deliberative perspective one can take is that of the individual: the only relevant locus of agency is the *individual*. Whereas, on the team reasoning proposal, each player thinks, not from her own perspective – where she asks, "What should I do, given what other players do?" – but from a larger, collective perspective, where she asks what *we together* should do. From this perspective it can be obvious what should be done; that is, it will be clear what the *collective* should do. If one can take on this point of view, then the thought is that this reasoning will give a straightforward answer regarding what one should do: namely, one's part in the collective activity that the reasoning recommends.[5]

---

[4] See a passage at Tuomela (2007), 101 that suggests the possibility of private or proprietary reasons. I take this not as evidence for a proper compatibilism but for the possibility of mixed cases that are less than fully ideal instances of joint action.

[5] There are many discussions of team reasoning, and some variety in what exactly is encompassed by the term. See Bacharach (2006), Sugden (1993), Gold and Sugden (2007), Anderson (2001) and Hurley (1989), as well as more recent discussion such as Pacherie (2018), Hakli, Miller and Tuomela (2010), and Roth (2012).

I have some reservations regarding how team reasoning is implemented. In particular, I feel that standard presentations of it owe us more of an explanation of exactly when it is reasonable to use team reasoning. Moreover, the answer to this question had better not collapse into a form of individual reasoning.[6] It will not be necessary, however, to address such worries here. I will take it for granted that some form of team reasoning is viable as a rational or normative theory, and that it's distinctive – i.e. doesn't collapse into a form of individual rationality.

What is important for my purposes is the characterization of the intentions of participants in joint action in terms of the distinctive team reasoning that issues those intentions. Gold and Sugden say

> Team reasoning was originally introduced to explain how, when individuals are pursuing collective goals, it can be rational to choose strategies that realize scope for common gain. But it also provides an account of the formation of collective intentions…It is natural to regard the intentions that result from team reasoning as collective intentions." (Gold and Sugden 2007, 126)

> "It is an implication of our analysis that a given pattern of behavior can be intended either individually or collectively, depending on the reasoning which led to it." (Gold and Sugden 2007, 128)

So far, I've mentioned team reasoning and how collective intentions at the heart of joint action might be characterized in terms of such reasoning. It is not far now to the *Exclusion Thesis* – the thought that joint action is at odds with proprietary reasons. One idea we still need is that team reasoning is a way of gaining access to a type of reason for action – one that is had primarily by a group. Although the literature of team reasoning tends to present itself as a theory of reasoning rather than of reasons, it seems that when you take on the more expansive deliberative perspective demanded by this form of reasoning, you gain access and sensitivity to reasons or considerations that you otherwise would not have.[7] For Gold and Sugden, an individual who engages in team reasoning starts from premises and concludes with an intention to perform his or her component of the joint action. Crucially, an individual "who accepts [these premises] and draws the inference 'has a reason to form the intention to choose that component' and presumably has a reason so to act." (Gold and Sugden 2007, 126).

Let me elaborate on this last remark. When one acts on an intention, one will be acting for the reasons that went into forming it. Take the case of an ordinary intention. When I decide to go for a bike ride this weekend, I usually have reasons for doing so: getting exercise, enjoying the weather, etc. The matter is now settled, so when the appropriate time comes, I don't normally have to re-deliberate. I just go for the ride. Since I don't have to deliberate at the time of action, I don't necessarily have to think about the reasons in favor of going for a ride. Nevertheless, when I go for the ride, I'm doing so for reasons: normally, the reasons that went into my decision several days ago. Intentions thus serve to *preserve* reasons from a prior episode of decision making so that when one acts on the intention, one acts for those reasons.

[6] Roth (2012).

[7] For example, Jackson (1987).

According to Gold and Sugden, collective intentions are formed by a distinctive kind of reasoning undertaken by an individual, one that involves taking on the perspective of a larger group of which the individual is a part. In so doing, one departs from the individual perspective; the reasons that one has access to in team reasoning needn't correspond to those available in individual practical reasoning. There may be some overlap. Some considerations might be relevant from both perspectives. However, *proprietary* reasons won't be retained at the collective level. (This just follows from their being proprietary.) If we follow Gold and Sugden in thinking that engaging in team reasoning leads to forming the relevant collective intention, then when one subsequently acts on that intention one is acting for the distinctive collective or group reasons one accessed through team reasoning. Insofar as one is acting on the collective intention, one is *not* acting on proprietary reasons.

In sum, if in joint action one is acting on collective intentions, then the distinctiveness of team reasoning ultimately explains why it is that in joint action one acts on group reasons to the exclusion of proprietary or personal reasons. Given the preservative role of intention and its origin of the intention in team reasoning, to act on the collective intention is to act for group reasons accessed through team reasoning, and not for proprietary reasons.[8]

Suppose we agree at this point that when one acts on a collective intention stemming from team reasoning that one acts for the corresponding group or collective reasons. Might it be possible nevertheless to be acting on a proprietary reason as well? Even if the team reasoning offers a deliberative perspective that is different from individual perspective, and puts the agent in touch with reasons she would otherwise not have, why must this entail that she lose her handle on her proprietary

[8] On my understanding, intentions formed when one undertakes team reasoning are quite distinct from ordinary individual intentions. The origin of the former attitude makes an important difference, such as the fact that when one acts on the intention one acts for group reasons – reasons that are not necessarily available as such to one from the individual perspective. For a contrasting view of intentions resulting from team reasoning, (see Kutz 2000). Kutz argues that joint action should be understood in terms of participatory intentions – an intention each participant has regarding his or her contribution to the collective action or outcome. Kutz argues for his participatory intention account of joint action in part by invoking team reasoning. But Kutz characterizes participatory intention as an ordinary individual intention, albeit with collective subject matter. I think that the upshot of my discussion is that Kutz is not entitled to this characterization of participatory intention; at the very least he needs to say more to justify it. Downstream causal/functional role might have tempted Kutz to disregard the difference in provenance between individual and collective intention. Consider for example snap judgment vs. judgment arrived at through deliberation and weighing of reasons. Both count as judgments presumably because of their downstream roles, despite their very different origins. However, both snap and deliberative judgments are subject to critical assessment by the same sorts of reasons or considerations. Whereas, individual and collective intentions arrived at through individual and team reasoning respectively seem not to be subject to assessments by the same standards or reasons; indeed, the point of team reasoning was to take on a perspective quite distinct from the individual perspective; indeed, to take up the individual perspective and to act from reasons from within that perspective, would undermine the commitment embodied in the collective intention stemming from team reasoning. (And if we were to ensure that the individual consideration were merely to ratify the collective intention, then it's not clearly an efficacious consideration.)

reason? It's quite possible that on some occasion, I can recognize that to engage in team reasoning and to act in accord with its recommendation can satisfy some proprietary reason of mine. (After all, in the Hi-Lo case, it is certainly in one's *personal* interest to be able to coordinate on Hi-Hi.)[9] So what stops me from acting on the proprietary reason as well?

The worry is going to be something like this. To regard the matter of what to do from the deliberative perspective within which one's proprietary reason has an exigent presence – and, moreover, to *act* from that perspective – is to *distance* oneself from the joint perspective from which the matter was supposedly settled. It is, perhaps, fine to recognize in the abstract that one may have proprietary reasons for engaging in joint activity. But if one is taking on the individual perspective to seek out the relevant proprietary reasons and to act for those reasons, it seems that one's grip on the collective reasons and team reasoning has become tenuous. In taking up the question of what, from an individual perspective, might be said in favor of or against acting in accord with team reasoning, is to re-litigate a matter that was supposed to have been settled through team reasoning and collective intention. It's in this sense, then, that acting from proprietary reasons is in tension with acting on a collective intention derived from team reasoning.

One might, nevertheless, wonder why one cannot *on the basis of a proprietary reason* undertake team reasoning and form the corresponding intention. And, in acting on it, couldn't one also be acting *for* the proprietary reason that prompted the undertaking in the first place? I agree that proprietary reasons can be part of what explains much of what we do, including the undertaking of certain forms of reasoning. But a consideration that figures in the history that leads up to an action in this way is not necessarily a consideration that is a reason for so acting. It might help in explaining the conditions or circumstances under which one acts, but not necessarily the action itself. That's to say that unlike the reasons-preserving process of acting on a prior intention, the process of undertaking team reasoning on the basis of proprietary reasons is not preservative, but transformative.

Perhaps a familiar example would help. I might recognize that my selfish egoistic outlook is actually hindering my relationships with people and working against my own interests. So I decide for selfish reasons to give up on normative egoism, and become more altruistic. A selfish motivation might have led me to transform myself into someone who is not selfish. This does not mean that my altruistic acts are selfishly motivated. It means that selfishness had a role to play in my becoming non-selfish.[10] Another example: I may have practical reasons for engaging in some form of theoretical reasoning. But this doesn't mean that those practical considerations can count as reasons for the beliefs generated by that theoretical reasoning.

[9] I'm setting aside the debate as to whether team reasoning might collapse into a form of individual reasoning. I will assume that it does not. Nevertheless, one might wonder whether one can for personal reasons become a team reasoner.

[10] Parfit (1984) argues that the fact that an egoist might have reasons to become altruistic is not an argument against egoism. This may be right. However, I don't think that this shows that one's actions subsequent to the transformation are done for egoistic reasons.

If we are on the right track with this line of thought, then it would appear that an understanding of joint action that ties collective intention to team reasoning points to an incompatibility between joint action and proprietary reason.

10.4 Accommodating Subordinate Proprietary Reasons

Is there, nevertheless, a way for the team reasoning conception of joint action to countenance participants acting on proprietary reasons? In one sense, the answer is easy, as we've already seen: the mixed view allows that there are proprietary reasons – with the upshot, however, that the action is less than *fully* joint. But there remains the intuition that joint action at least in some circumstances is perfectly compatible with proprietary reasons.[11] Moreover, it's unclear how joint action could be sustained or stable if *nothing* could be said in favor of the joint effort from the personal or proprietary point of view. It is likely, after all, that participants in joint activity – especially when it is long term – will on some occasion fall into an individual perspective regarding what they are doing. It is unrealistic and likely undesirable (stifling and oppressive) to think that the team or group perspective of the collective intention can *never* loosen its grip on the agent's practical point of view.[12] Indeed, joint activity is often engaged in precisely because it serves the individual interests of the individuals involved. Given the ever-present possibility of seeing things from the perspective of the individual agent, we might welcome the possibility that sometimes that perspective could be taken within the context of joint activity. So it would be worthwhile for one who places team reasoning at the center of her account of joint action to investigate this possibility.

The case of competitive games is suggestive.[13] This is a form of joint activity, and yet when you and I engage in it, much of what I do is aimed at winning against you. This is a reason not shared with you, so it seems to be proprietary.[14] But it is a proprietary reason that is, nevertheless, very much a part of engaging in that sort of joint activity. We might speak of such proprietary reasons as *constitutive* of the particular joint yet competitive activity in question.[15]

---

[11] And indeed, proprietary reasons might be necessary if we're to avoid turning joint action into monolithic individual agency.

[12] This is not to say that sometimes you don't want this perspective, as in the case of temptation.

[13] Searle (1990).

[14] If you are also aiming for me to win, then you are not in the fullest sense playing this game.

[15] Some might be tempted to think that one cannot even act on such a proprietary reason (defined in terms of the aim of winning against you) outside the framework established by some joint activity. But can I not have this reason and act on it outside such a framework? After all, might it explain why I seek to play the game with you? This is indeed a proprietary reason, but it's not clear that this is the reason one is acting on when one is really playing the game. I want to beat you at a game. And this might explain why I seek out the opportunity, and why I engage in it. But when we play the game – really play it together – then it's no longer clear that this is the relevant reason that explains my particular moves within the game. It seems now that I have a new *framework defined*

But what about non-competitive activity? Is it possible to act on proprietary reasons in such contexts? Notice that someone might be engaged in a non-competitive joint action, and yet have his own reasons for what he's doing. For example, how well he contributes to some joint effort will have reputational effects (how good a team player is he?), and open up possibilities for collaborations not only with current partners, but with *other* partners as well. Another example: a particular task falls to me in the context of joint activity; there are several ways of doing it all perfectly adequate, but I choose one because it is particularly enjoyable for me, or because doing it that way will afford me an opportunity to develop skills or know-how useful for me in projects outside of the joint activity. Thus, an agent might have reasons for how he conducts himself within the joint action that are not shared with other participants in joint activity: it needn't be one of your aims in joint action that I gain a good reputation and secure future opportunities.[16] So long as the current activity is not undermined or compromised by being motivated by such perks, it seems that this is an instance of a proprietary reason for what I'm doing in joint action.

Does this contradict what I said earlier about the incompatibility of team and individual reasoning? I don't think so. The proprietary reason in question only figures as a reason for how one acts *within* a framework established by the team reasoning that leads to collective intention and joint activity. In particular, one needn't think that in participating in joint action in the first place one is acting on that proprietary reason. Moreover, the proprietary reason is one that promotes one's engagement in the joint activity in question; after all, it's a consideration for being a good team player, or encourages one to do one's part with greater relish. Such a proprietary reason is not constitutive in the sense introduced above in connection with competitive games. But it would be natural to regard such proprietary reasons as permitted or sanctioned and, along with those that are constitutive, entirely compatible with joint action.

In sum, the fundamental idea here is that team reasoning conception of collective intention is supposed to settle the matter of one's participation in joint action. This committal element of the collective intention (along with the idea that intentions serve to transmit reasons) entails that when one acts on the collective intention, one is acting for collective reasons, not the proprietary reasons. To act on the individual intention and for proprietary reasons implies that the matter of whether to engage in the joint action has been reopened even though no relevant defeater has been triggered; this would undermine the committal element of the collective intention.

The only sort of proprietary reason that one could act on in a way that is compatible with the committal force of collective intention would be those that are *subordinate* to the collective reason. In contrast, one cannot act for what we might refer

goal and reason for action. It's a mistake to run together reasons inside and outside the framework of the game, even if they are superficially similar.

At least, this is true for joint activity engaged in through team reasoning. There may be forms of joint action that are more individualistic and don't require team reasoning.

[16] Though I might explicitly take on this end.

to as *coordinating* proprietary reasons – that is, those proprietary reasons that are not subordinate, but stand alongside of the collective or group reason (in favor of the joint activity). Thus, although the team reasoning conception of joint action is fully compatible with *some* proprietary reasons, this accommodation is significantly circumscribed and limited to subordinate reasons.

10.5 Concluding Remarks

I hope to have articulated certain underappreciated aspects of the team reasoning conception of collective intention that suggest that acting on non-subordinate proprietary reasons is incompatible with fully joint action. Some might find this to be a problematic upshot of the team reasoning view of collective intention. I will not try to assess whether this amounts to a serious critique of the view. At the very least, it points toward the distinctiveness of the team reasoning view. There is a significant chasm between this view and theories of joint action that are more reductive in seeing collective intentions as a form of individual intention – and thus are friendlier to the possibility of proprietary reasons.

I want to turn finally to implications for more minimal joint action. There are different respects in which one might imagine joint action or shared agency to be minimal. For example, one might imagine participants that are quite incompetent in coordinating with one another. Although each is very much committed to a joint project, one can imagine that their collective bumbling would at best merit the label of *minimally* joint action if it counts as joint at all.

Another sense of minimality is closer to the concerns of this paper. Joint action might be minimal insofar as it is very opportunistic. Each participant finds it in her interest to cooperate because of a highly contingent circumstances. Individuals might have very divergent goals, but circumstances happen to be such that some immediate aim serves the purposes of each. As soon as circumstances change, cooperation ceases in part because no other aim is shared so as to serve as a further focal point for sustained interaction. Joint action can thus be minimal in the sense of being more contingent, more opportunistic, and less stable. Correspondingly, one might imagine joint action being more robust, stable, and sustained the more reasons are shared, with fewer proprietary reasons that figure as reasons for joint action.

That, I think, is an intuitive way to understand at least *one* way in which joint actions might range from the minimal to the robust. But if what we've been saying about the team reasoning conception of collective intention and joint action is on the right track, then this view doesn't offer a dimension along which we find a gradient from minimal to robust forms of joint action. It would therefore make no sense to talk of minimal forms of joint action. As soon as proprietary reasons of the coordinating, non-subordinate variety figure as reasons for which one is acting, one is no longer genuinely acting *with* others. On the team reasoning view, joint action, at least in the ideal, is all or nothing.

References

Anderson, E. (2001). Unstrapping the straightjacket of 'preference': A comment on Amartya Sen's contributions to philosophy and economics. *Economics and Philosophy, 17*, 21–38.

Anscombe, G. E. M. (1963). *Intention* (2nd ed.). Ithaca: Cornell University Press.

Bacharach, M. (2006). In N. Gold & R. Sugden (Eds.), *Beyond individual choice: Teams and frames in game theory*. Princeton: Princeton University Press.

Bratman, M. (1992). Shared cooperative activity. *Philosophical Review, 101*, 327–341.

Bratman, M. (2014). *Shared agency: A planning theory of acting together*. New York: Oxford University Press.

Davidson, D. (1980). *Essays on actions and events*. Oxford: Oxford University Press.

Gold, N., & Sugden, R. (2007). Collective intentions and team agency. *Journal of Philosophy, CIV*(3), 109–137.

Hakli, R., Miller, K., & Tuomela, R. (2010). Two kinds of we-reasoning. *Economics & Philosophy, 26*, 291–320.

Hurley, S. (1989). *Natural reasons*. New York: Oxford University/Cambridge University Press.

Jackson, F. (1987). Group morality. In P. Pettit, R. Sylvan, & J. Norman (Eds.), *Metaphysics and morality* (pp. 91–110). Oxford: Blackwell.

Kutz, C. (2000). Acting together. *Philosophy and Phenomenological Research., 61*, 1–31.

Pacherie, E. (2018). Solution thinking and team reasoning: How different are they? *Philosophy of the Social Sciences, 48*(6), 585–593.

Parfit, D. (1984). *Reasons and persons*. Oxford: Oxford University Press/Clarendon Press.

Roth, A. (2012). Team reasoning and shared intention. In A. Ziv & H. Schmid (Eds.), *Institutions, emotions, and group agents* (pp. 279–295). Dordrecht: Springer.

Searle, J. (1990). Collective intentions and actions. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 401–415). Cambridge: MIT Press.

Sugden, R. (1993). Thinking as a team: Towards an explanation of nonselfish behavior. *Social Philosophy and Policy, 10*, 69–89.

Tuomela, R. (2007). *The philosophy of sociality: The shared point of view*. Oxford: Oxford University Press.

Chapter 11
Motor Representation and Action Experience in Joint Action

Corrado Sinigaglia and Stephen A. Butterfill

Abstract Acting together with a purpose is a familiar feature of everyday life. We jump together, play music together and move tables together. But what do we experience of action in acting together? It is perhaps tempting to suppose that there is a special way in which we can experience our own actions, and that we cannot experience the actions of others in this way. This view would imply that in acting together, our own actions are experienced in a way that our partners' actions are not. However recent research on motor representation suggests that, in observing another act, it may be possible to experience her actions in whatever sense we can experience our own actions. This makes it at least conceivable that in acting together we can experience the actions each of us performs in the same way. But the occurrence of a joint action involves more than merely the occurrences of two individual actions. Are there experiences of joint actions which involve more than merely two or more experiences of individual actions? In this chapter we defend a positive answer. In some cases, experiences associated with joint action are experiences of action in whatever sense experiences of acting alone are.

Keywords Joint action · Motor representation · Action experience · Collective goals · Shared agency

11.1 Introduction

Many of the things we do are done together with others. We play duets, move pianos together and drink toasts together. We also fill rooms with noise together, damage furniture together and spill drinks together. As these examples hint, acting together is sometimes but not always done with a purpose. Filling a room with noise is some-

C. Sinigaglia
Dipartimento di Filosofia, Università degli Studi di Milano, Milan, Italy
e-mail: corrado.sinigaglia@unimi.it

S. A. Butterfill (✉)
Department of Philosophy, Social Sciences Building, University of Warwick, Coventry, UK
e-mail: s.butterfill@warwick.ac.uk

A. Fiebich (ed.), *Minimal Cooperation and Shared Agency*, Studies in the Philosophy of Sociality 11, https://doi.org/10.1007/978-3-030-29783-1_11

thing that we typically do together (neither you nor I alone is speaking quite loudly enough to fill the room with noise), but it is not usually something done with a purpose. By contrast, playing a piano duet and drinking toasts are paradigm cases of acting together with a purpose.

What is it to act together with a purpose? Minimally, it seems necessary that there should be a single outcome to which the actions we perform in acting together with a purpose are directed. Further, our actions being directed to this outcome should not be, or should not only be, a matter of each of our actions being directed to that outcome. After all, acting together with a purpose is not merely a matter of me doing one thing and you doing another.

Of course philosophers have offered various more elaborate ways to characterise forms of acting together with a purpose (e.g. Bratman 2014; Gilbert 2013). It is possible that our characterisation needs to be elaborated along some such lines. But for our purposes the above minimal characterisation of acting together with a purpose will be sufficient.

For the purposes of this chapter, a *joint action* is an event involving two or more agents acting together with a purpose; similarly, an *individual action* is an event involving a single agent acting alone with a purpose. Elsewhere the term 'joint action' has been defined in a variety of ways, some broader, others narrower (e.g. Sebanz et al. 2006; Ludwig 2007). Our definition can be treated as a terminological stipulation. Its purpose is only to fix the range of cases of interest for our questions about experience.

We face an immediate difficulty in talking about experiences associated with (joint) action. What are the contents of such experiences? Or, if experiences lack contents (we rely on no assumption about this in what follows), what do such experiences present to their subjects? Do they present actions in whatever sense some visual experiences present colours and shapes? Or are experiences associated with action phenomenologically no different from experiences associated with events that are not actions? As things stand, we know of no way to answer such questions (see Sinigaglia and Butterfill 2016 for some difficulties involved in deciding between the two views). But we do know a way to duck these questions. Experiences of action enable their subjects to make judgements about which action is being observed or performed—for example, about whether the goal of the action is to pull or push an object (Cattaneo et al. 2011). This suggests that at least some experiences associated with actions provide their subjects with reasons for judgements about those actions. Let us stipulate that for an experience to *reveal* an action is for it to provide the subject with a reason for judging that this action is being performed, and, further, that an *experience of action* is one that reveals an action. The existence of experiences of action (in this sense) is both relatively uncontroversial and neutral on different views about the contents of experiences associated with actions. Despite this, it is far from uninteresting for it raises deep and difficult issues. In this chapter, our concern is with a question about experiences of joint actions. The occurrence of a joint action involves more than merely the occurrences of two individual actions. But what about experiences? Are there experiences of joint actions which involve more than merely two or more experiences of individual actions?

To answer this question, we first need to consider some background issues concerning experiences of individual action. Our plan is to consider how comparatively well-established views about experiences of individual action might generalise to joint action.

11.2 A Preliminary Distinction

As a preliminary, let us distinguish our question about experiences of actions from questions about experiences of acting. Experiences of action reveal what is being done whereas experiences of acting reveal who is acting.[1]

A body of research focuses on experiences which enable people to identify whether or not they are effective as agents of events. Here it is common to talk about a sense of agency, which is sometimes defined as the 'experience of controlling one's own actions' (Haggard and Chambon 2012, 390). Pacherie (2014) connects research on the sense of agency to joint action. However, our concern is different. The sense of agency is associated with experiences of acting, whereas our concern is with experience of action.

To see that this distinction matters, consider a patient with anarchic hand syndrome whose two hands appear to be fighting against each other (that is, there is intermanual conflict). Imagine that one is fastening buttons on her shirt while the other undoes them.[2] There is clearly a disruption in her sense of agency: she has a sense of being in control of some actions but not others. But both actions are clearly purposive, and she can readily identify the goals to which they are directed. As far as anyone can tell, what she experiences of action is barely, if at all, affected by the disruption to her sense of agency. So while the two may be linked in all kinds of ways, cases of intermanual conflict in anarchic hand syndrome demonstrate the importance of distinguishing experiences of agency from experiences of action. And whatever we experience of action, it cannot be reduced merely to our having experiences of agency.

11.3 What Are the Causes of Experiences of Action?

Having distinguished experiences of action from experiences of acting, a question arises. What are the causes of such experiences? What factors could you change in order to alter those of an agent's experiences which reveal what is being done?

[1] As mentioned in the Introduction, we stipulate that for an experience to *reveal* an action is for it to provide the subject with a reason for judging that this action is being performed.

[2] Fisher (2000, 197) cites this case from Bogen (1973).

Our ultimate aim to understand experiences associated with acting together. But it will be helpful to start with ordinary, individual action as there is a far wider body of evidence covering this case.

One initially tempting assumption is that the causes of experiences of action are the configurations and movements of a body involved in acting, and perhaps also some of the perceptible consequences of these. On this view, experiences of action associated with brushing your hair (say) are a consequence of the ways your hand grips the brush and moves it through your hair.

This assumption appears to be wrong, however, because there are individuals who appear to experience actions without actually performing them at all. Patients with anosognosia for hemiplegia are characterised as being unaware of suffering from paralysis (e.g. Berti et al. 2005). When requested to perform actions involving a limb which is in fact paralysed, they will report having done so. Further, indirect measures (which exploit bimanual interference effects and brain responses) indicate that this is not mere confabulation (Berti et al. 2005, 2008; Garbarini et al. 2012). Although not decisive, this is a strong reason to suppose that experiences of action cannot all be merely consequences of bodily configurations and movements. After all, the patient with anosognosia for hemiplegia reports experiences like those involved in actually acting in the absence of relevant bodily configurations and movements.

The idea that experiences of actions do not depend on actual bodily configurations or movements is strengthened by reflection on imagining acting. There is a way of imagining acting which is not simply a matter of thinking about acting but which involves motor imagery (Jeannerod 1994, 1995). Imagining acting in this sense is phenomenally closer to actually acting than merely thinking about acting is. The similarity between experiences associated with actually acting and with imagining acting is a further reason to hold that experiences of action cannot all be merely consequences of bodily configurations and movements.

Given the experiences associated with imagining acting, and the experiences of patients with anosognosia for hemiplegia, it may be tempting to assume that the causes of experiences of action are intentions. After all, you can have an intention whether or not you act. This can make it seem that intentions are common to all cases in which there is experience of action.

Against this possibility, consider the experiences of patients with anarchic hand syndrome. Such patients may readily detect goals to which their various actions are directed even when these goals are in conflict with what they avowedly intend to achieve. And they appear to have whatever experiences are characteristic of actions rather than, say, experiencing the movements of an anarchic hand as mere events, as they would if, for example, they were suffering from a tremor or other loss of control over the body.[3] So even when actions run counter to intentions, experiences of action can persist. (At least, this follows unless we are willing to suppose that all such patients have manifestly conflicting intentions, contrary to what they them-

[3] According to Gallese and Sinigaglia (2010, 750), 'the anarchic hand is still felt as being part of the experiencing and acting body and its actions'; even if these actions occur 'outside the agent's will, [they] are still lived by the patient as potentialities of his/her body.'

selves say.) These unusual cases reveal that whatever we experience of action, it cannot all be a consequence of our intentions.

Whatever causes the experiences of action, it cannot be bodily manifestations of action only because some experiences of action occur in the absence of such manifestations; and it cannot be intentions only because some experiences of action are contrary to what is intended. This motivates considering a third possibility: experiences of action are (at least in part) a consequence of motor representations.

Motor representations are the representations involved in control of very small-scale actions such as playing a chord, dipping a brush into a can of paint, placing a book on a shelf or cracking an egg. Attention to the ways these actions unfold reveals that, often enough, the early parts of an action anticipate the future parts in ways that cannot be determined from environmental constraints alone. This is a sign that even very small-scale actions are a consequence of representations concerning how actions will unfold in the future. These are what we refer to as *motor representations*.[4]

Motor representations resemble intentions in some key respects. Like intentions, some motor representations specify outcomes, provide for the coordination of action, and normally do so in such a way as to increase the probability that the specified outcome will occur (Rizzolatti and Sinigaglia 2010, 2016). This makes it easy to conflate motor representations with intentions. However, the two should be distinguished on several grounds. One is that motor representations exclusively concern present actions rather than potentially actions at some unspecified future time. Another is that motor representations, unlike intentions, are tied to the body and its capacities to act. But perhaps most importantly, intention and motor representation differ with respect to representational format and are therefore not inferentially integrated (Butterfill and Sinigaglia 2014). The claim that motor representations are a cause of experiences of action is therefore distinct from any claim about intentions.

Motor representations are good candidates for causes of experiences of action. They are present in ordinary cases as well as those involving anarchic hand syndrome and anosognosia for hemiplegia (on the latter, see Berti et al. 2005). They also underpin imagining acting (Jeannerod 2003). So whatever exactly we experience of action, at least some component of this experience is likely to be a consequence of how actions are represented motorically. We can therefore be confident that motor representations are among the causes of experiences of action.

11.4 Motor Representations Shape Experiences of Action

In the previous section we considered causes of experiences of action. In discussing these causes we were neutral on the nature of such experiences. So far they have been characterised simply as experiences associated with actions which reveal what

[4] On what motor representations are and why they are necessary, key sources include Rosenbaum (2010), Prinz (1990), Wolpert et al. (1995), Jeannerod (2006) and Rizzolatti and Sinigaglia (2008).

is being done. Our aim in this section is to take a small further step without getting entangled in complicated issues about the contents of experience.

There are experiences which provide their subjects with reasons for making judgements about actions (or so we have been assuming from the start). Suppose, for example, that you have been lying in bed for some time worrying about various problems. At some point you may find yourself getting up. Your experiences may provide you with reasons to judge that you are doing this, and that you are performing various actions associated with getting up. What things can influence whether your experience provides you with these reasons rather than with some quite different reasons (or with no reasons at all)? We will say that these things, whatever they are, *shape* your experience. So for something to shape an experience of action is just for it to influence which reasons this experience provides.

Note that a thesis concerning what shapes experiences of actions can be neutral on their contents. The contents of such experiences may only ever involve bodily configurations, joint displacements, sounds and other perceptible consequences of acting; or they may involve actions in some richer sense (see Sinigaglia and Butterfill 2016). Our discussion of what shapes experiences of action is neutral between these and other possibilities.

In the previous section we argued that motor representations are a cause of experiences of action. In this section we shall argue, further, that motor representations also shape experiences of action.

As a first step, consider that motor representations can influence judgements about actions and their consequences. This is true both for actions you are performing (for example, Costantini et al. 2011) as well as for actions you are merely observing (for example, Casile and Giese 2006; Cattaneo et al. 2011; see Blake and Shiffrar 2007, 56–57 for a concise review of earlier research). To illustrate, consider a pair of experiments by Repp and Knoblich. In each case, subjects were asked to make a judgement about the relative pitches of two tones. These tones were always played in the same order. However, the tones were carefully selected to be an ambiguous pair: that is, they could be heard either as ascending in pitch or as descending in pitch. Repp and Knoblich investigated how judgements were influenced by motor representations by exploiting the fact that expert pianists' motor representations of certain keyboard actions specify not only finger movements but also chords (Haslinger et al. 2005). As they found, performing (Repp and Knoblich 2009) or observing (Repp and Knoblich 2007) soundless keyboard actions that expert pianists would typically represent in relation to a pair of tones which increase (or decrease) in pitch resulted in the expert pianists being correspondingly biased to judge that the tones they heard were increasing (or decreasing) in pitch. That is, the way you represent your own or another's actions motorically can influence your judgements about those actions and their consequences.

Second, the influence of motor representation is thought to have a function: it enables you to make judgements which are more accurate or detailed than would otherwise be possible (Wilson and Knoblich 2005; Shiffrar 2010, 72).

Third, it is sometimes assumed that the effects of motor representation on judgements goes via experience. For example, in her discussion, Shiffrar (2010) shifts

without explicit argument from motor processes to 'motor experience' (her term). We take this assumption to be correct. After all, motor representations and judgements are inferentially isolated (perhaps because of differences in their representational formats; see Butterfill and Sinigaglia 2014). The existence of any facilitatory connection between them stands in need of explanation. We suppose that such connections depend in one way or another on experience. That is, motor representations influence experiences associated with actions which in turn influence judgements. And if this is right, we have arrived at the conclusion that motor representations are not only causes, but also shape, experiences of action.

11.5 Experiences of Joint Action

Up to this point we have considered individual action only. This is because so much research on action focuses on this case. But our aim is to understand experiences associated with joint action. Suppose two people are performing a very small-scale joint action. For example, they are clicking glasses, passing an object between them or playing a chord together. What could either of them, or an observer, experience of the joint action?

We have already removed one potential obstacle to answering this question. Antecedent of any discoveries about motor representation and experience, it would perhaps have been tempting to suppose that there is a special way in which we can experience our own actions, and that we cannot experience the actions of others in this way. However, the twin discoveries that motor representations shape experiences of action and that motor representations occur in observing others act indicate that this may be incorrect. Instead, it appears (in the absence of further discoveries about what shapes experiences of action with contrary implications) that experiences of your own and of another's actions have a common element. Further, if we focus on how experiences provide their subjects with reasons for judgements about actions, there appears to be no ground for assuming that any difference in kind between your experience of an action of your own and your experience of the same action performed by another. So the fact that a joint action involves more than one agent does not straightforwardly entail that there cannot be experiences of joint action.

Questions about experiences of joint action are nevertheless complicated by the fact that a joint action is not simply a composite of two ordinary, individual actions: it is the event of two or more people acting together with a purpose (see Introduction). This requires, minimally, that there is a single outcome to which the individual actions are directed and this is not, or not only, a matter of each individual action being individually directed to that outcome. It follows if experiences associated with joint action consisted in no more than experiences associated with component individual actions, then experiences associated with joint action would not reveal any difference between multiple actions being performed in parallel and genuinely joint actions. We should therefore ask whether experiences of action only ever provide their subjects with reasons for judgements concerning the goals of individual

actions? Or do they sometimes provide reasons for judgements concerning the goals of joint actions where this is not, or not only, a matter of them providing reasons for judgements concerning the goals of individual actions?

We stipulated earlier (in the Introduction) that an experience of action is one that reveals the action; that is, one which provides its subject with reasons for judging that the action is being performed. Our question, then, is simply whether there are any experiences of joint actions which are not merely experiences of individual actions?

Given that motor representations shape experiences of action, an important step towards answering this question is to consider whether motor representations could also shape experiences of joint action.

11.6 Motor Representation in Joint Action

What role do motor representations play in joint action? In pursuing this question it is helpful to introduce some further terminology. A joint action is an event involving two or more agents acting together with a purpose (as we stipulated in the Introduction). In acting together with a purpose, there is an outcome to which the agents' actions are directed where their being so directed is not, or not only, a matter of each individual action being directed to this outcome. In what follows we will need to refer to outcomes with this property. Let us therefore stipulate that for some actions to be *collectively directed* to an outcome is for them to be directed to this outcome and for their being so directed not to be, or not only to be, a matter of each action being so directed. Further, a *collective goal* is an outcome to which two or more agents' actions are collectively directed. Note this notion of collective goal is neutral on mechanisms: it says nothing at all about the psychological (or other) states in virtue of which actions have collective goals.

Now consider a joint action in which you draw a line and another draws a circle where these actions, yours and the others, are collectively directed to producing a single design. What is represented motorically in you?

Consider three possibilities: (1) the only outcomes ever represented motorically in you are those to which your own actions are directed; (2) the only outcomes ever represented motorically in you are those to which the other's actions are directed and the outcomes to which your own actions are directed; and (3) the outcomes represented motorically in you sometimes include the collective goal of producing a single design. To distinguish the first two possibilities experimentally, we can compare situations in which you are acting alone with situations in which you are performing a joint action (e.g. Sebanz et al. 2003). Indeed, a range of evidence suggests that the first possibility can be excluded.[5] But to distinguish the second and

[5] See, for example, Kourtis et al. (2013), Meyer et al. (2011), Meyer et al. (2013), Janeen D. Loehr and Vesper (2015), Janeen D Loehr et al. (2013), Ménoret et al. (2014), Baus et al. (2014), Schmitz et al. (2017), Novembre et al. (2014). For complications and opposing interpretations of some of the evidence, (see Dolk et al. 2014; Dittrich et al. 2016; Wenke et al. 2011; Constable et al. 2017).

third possibilities we need a further comparison: between joint action and parallel but merely individual actions.

To this end, della Gatta et al. (2017) had participants draw lines with their right hands while observing circles being unimanually drawn by a confederate. To create a minimal contrast between acting together with a purpose and merely acting in parallel, participants were divided into two groups with different instructions. In the 'acting-together group', participants were instructed to perform the task together with the confederate, as if their two drawing hands gave shape to a single design. In the 'acting-in-parallel group', participants were given no such instruction. Importantly, the groups differed only in the instructions given before the drawing started. If participants were to follow the instructions, their actions would be collectively directed to the outcome of drawing a circle and a line in the acting-together group only. Earlier research has established that when people have to perform incongruent movements simultaneously, such as drawing lines with one hand while drawing circles with the other, each movement interferes with the other and line trajectories tend to become ovalized (Franz et al. 1991). This ovalization has been described as a bimanual coupling effect, suggesting that motor representations for drawing circles can affect motor representations for drawing lines (Garbarini et al. 2012). Accordingly, if the collective goal of producing the line-circle drawing is represented motorically (as only possibility 3 above allows), there should be an interpersonal motor coupling effect in the acting-together group only. This would result in greater ovalization of the lines drawn in the acting-together group than in the acting-in-parallel group, which is actually what della Gatta et al. found.

The findings of della Gatta et al. indicate that collective goals can be represented motorically. We should be cautious about relying on evidence from a single source, of course: positive results from replications and extensions of their paradigm would strengthen confidence. Nevertheless, for the purposes of this chapter we will assume that when people act together with a purpose, collective goals are sometimes be represented motorically. What might this tell us about experiences of joint action?

11.7 Motor Representations Shape Experiences of Joint Action

Our question (see Experiences of Joint Action) is whether there are any experiences of joint actions which are not merely experiences of individual actions. An indirect way to answer this question is suggested by combining two pieces of evidence. First, motor representations are not only causes of experiences of action (see What Are the Causes of Experiences of Action?) but also shape experiences of action (see Motor Representations Shape Experiences of Action). So when experiences provide reasons for giving a particular answer to a question about which goal an action is

directed to, which answer the experiences provide reasons for giving can depend on which goal is represented motorically. Second, as we have just seen (in Motor Representation in Joint Action), the collective goals of joint actions can be represented motorically. We therefore conjecture that, when observing or performing a joint action, motor representations of a collective goal of the joint action sometimes both cause and shape your experiences. This conjecture goes beyond the evidence discussed so far and, to our knowledge, is yet to be tested. Were it correct, it would follow that experiences can provide reasons for judgements about the overall goal to which a joint action is directed in whatever ways experiences do for judgements about individual actions. And, further, that experiences associated with joint action do not, or do not only, provide reasons for judgements concerning the goals of individual actions: they can also provide reasons for judgements concerning the collective goals of joint actions.

In short, there are motor representations of collective goals, and if (as we conjecture) these can shape experiences then there are experiences of joint action which are not merely experiences of individual actions.

It may be useful to note a limit on the conclusion we have drawn. Can experiences of action also enable their subjects to make judgements not just about the goals of actions but also about whether they are performing or observing an individual or a joint action? The ideas and discoveries we have reviewed in this paper do not provide an answer to this question either way. Consider the motor representations which (we conjecture) shape experiences of joint actions. For all we know, these are agent-neutral representations of outcomes.[6] Now consider that an individual action and a joint action may be directed to just the same type of goal: filling a glass or playing a chord, say. (In the joint action, the work is divided between two agents.) Agent-neutral motor representations of the goals of such actions may in principle not differ between the individual action and the joint action. It follows that the experiences shaped by such motor representations may in principle not provide reasons for judging that the action observed or performed is a joint or individual action.

This neutral conclusion makes sense given earlier reflection on the consequences of the agent-neutrality of motor representations (see Motor Representations Shape Experiences of Action). In the case of ordinary, individual action, the fact that experiences are shaped by motor representations may explain none of their powers (if any) to provide reasons for judgements about who is acting. It is only a small further step to the view that being shaped by motor representations may also fail to explain any power that experiences have to provide reasons about the number of those acting.

[6] For a representation to be *agent-neutral* is for its content to not specify any particular agent or agents. Some motor processes appear to involve agent-neutral representations (Ramsey and Hamilton 2010; Jeannerod 2003). Gallese (2001) and Pacherie and Dokic (2006) argue for the agent-neutrality of some motor representations.

11.8 Conclusion

In this chapter our aim was to investigate whether there are experiences of joint actions which involve more than merely experiences of individual actions. One complication we faced was uncertainty over the phenomenology of experiences associated with action. To avoid relying on any premise about the contents of experiences, we argued for a parallel between individual and joint action. Just as when acting individually there are sometimes motor representations of the outcome to which the action is directed (see What Are the Causes of Experiences of Action?), so also when acting jointly there are sometimes motor representations of the collective goal (see Motor Representation in Joint Action). In the individual case, these motor representations can shape experiences of action (see Motor Representations Shape Experiences of Action). That is, which actions the experiences reveal to their subject depends, at least in part, on which outcomes are represented motorically. We conjecture that the same is true in the case of joint action. If this is right, experiences associated with acting together are experiences of action in whatever sense experiences of acting alone are. Not all experiences associated with joint action are merely experiences of individual actions: there are irreducible experiences of joint action.

References

Baus, C., Sebanz, N., de la Fuente, V., Branzi, F. M., Martin, C., & Costa, A. (2014). On predicting others' words: Electrophysiological evidence of prediction in speech production. *Cognition, 133*(2), 395–407.

Berti, A., Bottini, G., Gandola, M., Pia, L., Smania, N., Stracciari, A., Castiglioni, I., Vallar, G., & Paulesu, E. (2005). Shared cortical anatomy for motor awareness and motor control. *Science, 309*(5733), 488–491.

Berti, A., Spinazzola, L., Pia, L., & Rabuffetti, M. (2008). Motor awareness and motor intention in Anosognosia for hemiplegia. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Sensorimotor foundations of higher cognition* (pp. 163–181). Oxford: Oxford University Press.

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology, 58*(1), 47–73.

Bogen, J. E. (1973). The Callosal syndromes. In K. M. Heilman & E. V. Valenstein (Eds.), *Clinical neuropsychology* (pp. 308–359). Oxford: Oxford University Press.

Bratman, M. E. (2014). *Shared agency: A planning theory of acting together*. Oxford: Oxford University Press.

Butterfill, S. A., & Sinigaglia, C. (2014). Intention and motor representation in purposive action. *Philosophy and Phenomenological Research, 88*(1), 119–145.

Casile, A., & Giese, M. A. (2006). Nonvisual motor training influences biological motion perception. *Current Biology, 16*(1), 69–74.

Cattaneo, L., Barchiesi, G., Tabarelli, D., Arfeller, C., Sato, M., & Glenberg, A. M. (2011). One's motor performance predictably modulates the understanding of others' actions through adaptation of premotor Visuo-motor neurons. *Social Cognitive and Affective Neuroscience, 6*(3), 301–310.

Constable, M. D., de Grosbois, J., Lung, T., Tremblay, L., Pratt, J., & Welsh, T. N. (2017). Eye movements may cause motor contagion effects. *Psychonomic Bulletin & Review, 24*(3), 835–841.

Costantini, M., Ambrosini, E., Sinigaglia, C., & Gallese, V. (2011). Tool-use observation makes far objects ready-to-hand. *Neuropsychologia, 49*(9), 2658–2663.

della Gatta, F., Garbarini, F., Rabuffetti, M., Viganò, L., Butterfill, S. A., & Sinigaglia, C. (2017). Drawn together: When motor representations ground joint actions. *Cognition, 165*, 53–60.

Dittrich, K., Bossert, M.-L., Rothe-Wulf, A., & Klauer, K. C. (2016). The joint flanker effect and the joint Simon effect: On the comparability of processes underlying joint compatibility effects. *The Quarterly Journal of Experimental Psychology, 0*(0), 1–16.

Dolk, T., Hommel, B., Prinz, W., & Liepelt, R. (2014). The joint flanker effect: Less social than previously thought. *Psychonomic Bulletin & Review, 21*(5), 1224–1230.

Fisher, C. M. (2000). Alien hand phenomena: A review with the addition of six personal cases. *Canadian Journal of Neurological Sciences, 27*(3), 192–203.

Franz, E. A., Zelaznik, H. N., & McCabe, G. (1991). Spatial topological constraints in a bimanual task. *Acta Psychologica, 77*(2), 137–151.

Gallese, V. (2001). The 'shared manifold' hypothesis. From Mirror neurons to empathy. *Journal of Consciousness Studies, 8*(5–6), 33–50.

Gallese, V., & Sinigaglia, C. (2010). The bodily self as power for action. *Neuropsychologia, 48*(3), 746–755.

Garbarini, F., Rabuffetti, M., Piedimonte, A., Pia, L., Ferrarin, M., Frassinetti, F., Gindri, P., Cantagallo, A., Driver, J., & Berti, A. (2012). 'Moving' a paralysed hand: Bimanual coupling effect in patients with Anosognosia for hemiplegia. *Brain, 135*(5), 1486–1497.

Gilbert, M. P. (2013). *Joint commitment: How we make the social world*. Oxford: Oxford University Press.

Haggard, P., & Chambon, V. (2012). Sense of agency. *Current Biology, 22*(10), R390–R392.

Haslinger, B., Erhard, P., Altenmüller, E., Schroeder, U., Boecker, H., & Ceballos-Baumann, A. O. (2005). Transmodal sensorimotor networks during action observation in professional pianists. *Journal of Cognitive Neuroscience, 17*(2), 282–293.

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences, 17*(2), 187–245.

Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia, The Neuropsychology of Mental Imagery, 33*(11), 1419–1432.

Jeannerod, M. (2003). The mechanism of self-recognition in humans. *Behavioural Brain Research, 142*(1–2), 1–15.

Jeannerod, M. (2006). *Motor cognition: What actions tell the self*. Oxford: Oxford University Press.

Kourtis, D., Sebanz, N., & Knoblich, G. (2013). Predictive representation of other People's actions in joint action planning: An EEG study. *Social Neuroscience, 8*(1), 31–42.

Loehr, J. D., & Vesper, C. (2015). The sound of you and me: Novices represent shared goals in joint action. *The Quarterly Journal of Experimental Psychology, 0*(ja), 1–30.

Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., & Knoblich, G. (2013). Monitoring individual and joint action outcomes in duet music performance. *Journal of Cognitive Neuroscience, 25*(7), 1049–1061.

Ludwig, K. (2007). Collective intentional behavior from the standpoint of semantics. *Nous, 41*(3), 355–393.

Ménoret, M., Varnet, L., Fargier, R., Cheylus, A., Curie, A., Portes, V. d., Nazir, T. A., & Paulignan, Y. (2014). Neural correlates of non-verbal social interactions: A dual-Eeg study. *Neuropsychologia, 55*, 75–97.

Meyer, M., Hunnius, S., van Elk, M., van Ede, F., & Bekkering, H. (2011). Joint action modulates motor system involvement during action observation in 3-year-olds. *Experimental Brain Research, 211*(3–4), 581–592.

Meyer, M., van der Wel, R. P. R. D., & Hunnius, S. (2013). Higher-order action planning for individual and joint object manipulations. *Experimental Brain Research, 225*(4), 579–588.

Novembre, G., Ticini, L. F., Schutz-Bosbach, S., & Keller, P. E. (2014). Motor simulation and the coordination of self and other in real-time joint action. *Social Cognitive and Affective Neuroscience, 9*(8), 1062–1068.

Pacherie, E. (2014). How does it feel to act together? *Phenomenology and the Cognitive Sciences*, 25–46.

Pacherie, E., & Dokic, J. (2006). From Mirror neurons to joint actions. *Cognitive Systems Research, 7*(2–3), 101–112.

Prinz, W. (1990). A common coding approach to perception and action. In *Relationships between perception and action, edited by Odmar Neumann and Wolfgang Prinz* (pp. 167–201). Berlin: Springer.

Ramsey, R., & de C. Hamilton, A. F. (2010). Understanding actors and object-goals in the human brain. *NeuroImage, 50*(3), 1142–1147.

Repp, B. H., & Knoblich, G. (2007). Action can affect auditory perception. *Psychological Science, 18*(1), 6–7.

Repp, B. H., & Knoblich, G. (2009). Performed or observed keyboard actions affect pianists' Judgements of relative pitch. *The Quarterly Journal of Experimental Psychology, 62*(11), 2156–2170.

Rizzolatti, G., & Sinigaglia, C. (2008). *Mirrors in the brain: How our minds share actions, emotions*. Oxford: Oxford University Press.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the Parieto-frontal Mirror circuit: Interpretations and misinterpretations. *Nature Reviews: Neuroscience, 11*(4), 264–274.

Rizzolatti, G., & Sinigaglia, C. (2016). The Mirror mechanism: A basic principle of brain function. *Nature Reviews Neuroscience Advance Online Publication*.

Rosenbaum, D. A. (2010). *Human motor control* (2nd ed.). San Diego, CA: Academic.

Schmitz, L., Vesper, C., Sebanz, N., & Knoblich, G. (2017). Co-representation of others' task constraints in joint action. *Journal of Experimental Psychology. Human Perception and Performance, 43*(8), 1480–1493.

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like One's own? *Cognition, 88*(3), B11–B21.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and mind moving together. *Trends in Cognitive Sciences, 10*(2), 70–76.

Shiffrar, M. (2010). People watching: Visual, motor, and social processes in the perception of human movement. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(1), 68–78.

Sinigaglia, C., & Butterfill, S. A. (2016). Motor representation in goal ascription. In Y. Coello & M. H. Fischer (Eds.), *Foundations of embodied cognition 2: Conceptual and interactive embodiment* (pp. 149–164). Hove: Psychology Press.

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.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin, 131*(3), 460–473.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. (1995). An internal model for sensorimotor integration. *Science, 269*(5232), 1880–1882.

Chapter 12
From Collective Memory … to Collective Metamemory?

Santiago Arango-Muñoz and Kourken Michaelian

Abstract Our aim in this chapter is to delineate the form of shared agency that we take to be manifested in collective memory. We argue for two theses. First, we argue that, given a relatively weak conception of episodicity, certain small-scale groups display a form of emergent (i.e., genuinely collective) episodic memory, while large-scale groups, in contrast, do not display emergent episodic memory. Second, we argue that this form of emergent memory presupposes (high-level and possibly low-level) metamemorial capacities, capacities that are, however, not themselves emergent group-level features but rather strictly individual-level features. The form of shared agency that we delineate is thus revealed as being minimal in three senses. First, the relevant groups are themselves minimal in terms of their size. Second, the form of memory in question is minimally episodic. And finally, the cognitive capacities attributed to the relevant groups are minimal, in the sense that they need not themselves be capable of metacognition.

Keywords Collective memory · Collaborative remembering · Metacognition · Extended cognition · Distributed cognition

The authors contributed equally to this work. SA-M's contribution was supported by the research project "Metacognitive Feelings and the Cognitive Phenomenology Thesis" funded by the Comité para el Desarrollo de la Investigación (CODI), Cód. 20167047, Acta 01 de 2016, Universidad de Antioquia. KM's contribution was supported by grant 16-UOO-016 from the Marsden Fund, administered by the Royal Society of New Zealand (this grant was hosted at his previous institution, the University of Otago). Work on the chapter was completed during a visit by SA-M to the Université Grenoble Alpes, supported by the IDEX conférenciers invités programme. The authors would like to thank Anika Fiebich for her invitation to write this chapter and for her insightful comments, as well as Francesco Guala, Corrado Sinigaglia, John Sutton, and Francesco Ferreti for their feedback.

S. Arango-Muñoz (✉)
Instituto de Filosofía, Universidad de Antioquia, Medellín, Colombia
e-mail: santiago.arango5@udea.edu.co

K. Michaelian
Centre de philosophie de la mémoire, Université Grenoble Alpes, Grenoble, France
e-mail: kourken.michaelian@univ-grenoble-alpes.fr

A. Fiebich (ed.), *Minimal Cooperation and Shared Agency*, Studies in the Philosophy of Sociality 11, https://doi.org/10.1007/978-3-030-29783-1_12

12.1 Introduction

Philosophical research on agency has focused primarily on agency as manifested in bodily action. Agency is, however, manifested equally in *mental* action. Whereas bodily action aims to effect change with respect to the agent's body or material environment, mental action aims to effect change with respect to his mind or knowledge. This chapter considers one particular manifestation of mental agency, that involved in *remembering*, which, we will assume, qualifies as a variety of mental action insofar as the rememberer aims to retrieve or reconstruct a past episode in a controlled manner (Arango-Muñoz and Bermúdez 2018).[1] Each of us is, of course, capable of remembering on his own, but we are also capable of remembering together with others. There is a growing body of research on this kind of *collective* remembering, and it is the minimal form of shared mental agency that is manifested in collective—as opposed to individual—memory that is our concern in the chapter.

The nature of the groups at issue in collective memory research varies significantly, with some researchers, especially those based in the social sciences, investigating remembering in groups composed of very large numbers of widely dispersed individuals, including whole nations and societies (see Barash 2017; Olick et al. 2011), while others, especially those based in psychology, investigate remembering in groups comprising much smaller numbers of more intensely interacting individuals, such as mother-child dyads and married couples. We will have something to say here about both large-scale and small-scale groups, but, reflecting a tendency in recent philosophy of memory (e.g., Theiner 2013; Huebner 2016; Kirchhoff 2016) to build on the results of empirical research on transactive memory (see Ren and Argote 2011 for an overview), our interest is primarily in groups of the latter sort.

We will defend two theses regarding collective memory, which we understand as exemplifying a form of shared mental agency that is "minimal" in three senses. The first thesis is that transactive memory systems, which, as Wegner (1987: 191) defined them, arise under certain conditions through "the operation of the memory systems of the individuals and the processes of communication that occur within the group", enable the emergence of a genuinely collective form of memory. Although we will argue that it is unlikely that such emergence occurs in large-scale groups, given the relative lack of cohesion of the latter, we maintain that findings on collaborative inhibition (Rajaram and Pereira-Pasarin 2010) and both quantitative (Meade et al. 2009) and qualitative (Harris et al. 2017) varieties of collaborative facilitation suggest that transactive memory does give rise to genuinely collective memory—specifically, to genuinely collective episodic memory—in some long-married couples and potentially in other small-scale groups. This is a first sense in which the form of shared mental agency with which we are concerned is minimal: it pertains to groups that are themselves minimal in terms of their size, typically consisting of only two

[1] In addition to this sort of voluntary (or deliberate) memory, memory is often involuntary (or spontaneous) (Mace 2007). Involuntary memory may not qualify as a form of mental action, and our discussion here pertains to voluntary memory only.

individuals. The claim that this form of shared mental agency is involved in collective episodic memory presupposes a relatively weak characterization of episodicity, and this is a second sense in which the form of shared agency in question is minimal: it pertains to a form of memory that is only minimally episodic.

Given Wegner's definition, a transactive memory system necessarily includes a metacognitive or metamemorial level at which "who knows what" in the group is tracked. The claim that transactive memory enables the emergence of genuinely collective memory thus leads us to pose the question whether the form of metacognition at work in collective memory is to be understood as an individual-level capacity or whether it is, instead, itself an emergent group-level capacity. Our ultimate aim in the chapter is thus to determine whether *collective memory* implies *collective metamemory*, and the second thesis that we defend is that metacognition is, in the relevant groups, a strictly individual-level capacity. This is a third (and final) sense in which the form of shared mental agency with which we are concerned is minimal: it is minimally demanding in terms of the sophistication of the cognitive capacities that are attributed at the group level.

12.2 Episodic Memory

It is easy enough to give a rough operational definition of memory: memory refers to the modification of an agent's behavioural tendencies on the basis of its past experience. This rough definition, however, encompasses a number of very different kinds of memory, and a group capable of one of these might or might not be capable of the others. We therefore begin by singling out the particular kind of memory in which we are interested here.

It is standard, when describing kinds of memory, to distinguish between *procedural* and *declarative* memory (see, e.g., Werning and Cheng 2017). The former, which aligns—perhaps approximately—with what epistemologists refer to as "knowledge how", is the kind of memory that is at work when an agent remembers how to perform an action of a given type and as such need not involve representations, or at least representations with consciously articulable content. The latter, which aligns—again, perhaps approximately—with "knowledge that", is the kind of memory that is at work when an agent retrieves or reconstructs a representation with consciously articulable content. There are certainly interesting questions to be asked about procedural memory in groups, but the idea of collective procedural memory remains largely unexplored (see Manier and Hirst 2008 for a preliminary exploration), and our argument here builds on extant work on *collective declarative memory*.

Given that it is declarative memory that is at issue, the modification of the agent's behavioural tendencies to which the operational definition given above refers amounts to the acquisition of an ability to entertain a representation of one or another sort. But the definition, so understood, still encompasses different kinds of memory. It is standard to distinguish, within declarative memory, between *semantic* and *episodic* memory. The former is the kind of memory that is at work when an

agent retrieves a representation that does not essentially refer to the occasion on which learning occurred. For example, when one remembers that Kiev is the capital of Ukraine, one need not represent the occasion on which one learnt that Kiev is the capital of Ukraine. The latter is the kind of memory that is at work when the agent retrieves a representation that does essentially refer to the occasion on which learning occurred. For example, when one remembers visiting Kiev, one necessarily represents one's visit to Kiev. Given that it is episodic memory that is at issue, the modification of the agent's behavioural tendencies to which the operational definition refers amounts to the acquisition of an ability to represent a particular past event. Competing accounts of this ability are available, but episodic memory is, on any account, a more demanding capacity than semantic memory. Thus, if a group is capable of episodic memory, it is also likely to be capable of semantic memory. The focus of our argument will therefore be on *collective episodic memory*.

There is, as just noted, a choice to be made among competing accounts of episodic memory. In fact, there are two distinct choices to be made here, corresponding to the distinction between *mnemicity* and *episodicity* (Michaelian and Sutton 2017a). The first is between competing conceptions of the nature of declarative memory in general. The second is between competing conceptions of the distinguishing feature of episodic memory in particular. We take these choices in turn.

12.2.1 Mnemicity

The two main conceptions of the nature of declarative memory are the archival conception and the constructive conception (Robins 2016). The *archival* conception aligns roughly with preservationism in the philosophy of memory, the core claim of which is that remembering is essentially a matter of the transmission of content—through stages of encoding, consolidation, storage, and retrieval—from the original experience to the retrieved representation. The *constructive* conception, in contrast, aligns with generationism, the core claim of which is that, while remembering may involve the transmission of content, it is essentially a matter of the construction of a representation that is more or less adequate with respect to the original experience (see Michaelian and Robins 2018).

It has sometimes been suggested that there is no fundamental incompatibility between the archival and the constructive conceptions, as the construction of a retrieved representation would seem necessarily to draw on content transmitted from past experience. If so, there is room here for a *compromise* conception, the core claim of which would be that both transmission and construction are essential to remembering. A detailed version of the compromise conception has yet to be developed (see Robins 2016 for an initial attempt), but it is nevertheless with such a conception that we will work here, as taking into account both the features of remembering emphasized by the archival conception and those emphasized by the constructive conception will enable us to develop a treatment of collective memory that speaks to partisans of both conceptions.

12.2.2 Episodicity

Regardless of the conception of mnemicity with which one works, one must choose between competing conceptions of episodicity. The two main conceptions of the distinguishing feature of episodic memory are what we will refer to as the content-based conception and the phenomenological conception (see Perrin and Rousset 2014).

According to the *content-based* conception, episodic memory is to be distinguished from semantic memory in terms of its content: episodic memory is memory for the "what", the "when", and the "where" of experienced past events. Though Tulving, in his foundational work (1972), initially adopted the content-based conception, he soon abandoned it in favour of the phenomenological conception (Tulving 1985). According to the *phenomenological* conception, episodic memory is to be distinguished from semantic memory in terms of its phenomenology: episodic memory, unlike semantic memory, involves autonoetic consciousness—a sense of the self in subjective time. Tulving's adoption of the phenomenological conception was motivated in part by the observation that semantic memory, too, sometimes provides information about the what, the when, and the where of past events, including experienced past events, and, as this observation has gained wider recognition, the phenomenological conception has become increasingly popular (e.g., Klein 2015; Mahr and Csibra 2018).

The phenomenological conception of episodicity tends to be endorsed by those who endorse the constructive conception of mnemicity—especially by those who view episodic memory as a form of mental time travel (e.g., Suddendorf and Corballis 2007)—but it does not appear to be entailed by the latter. Indeed, there are views available on which, while autonoesis is indeed characteristic of episodic memory in healthy human subjects, it is not, strictly speaking, essential to episodic memory even in such subjects (Michaelian 2016). One motivation for such views is provided by the need to accommodate cases in which subjects appear to be able to remember normally despite having an impaired capacity for autonoesis (Klein and Nichols 2012). Another motivation is provided by research pointing to a capacity for "episodic-like" or what-when-where memory in nonhuman species despite a lack of evidence for a capacity for autonoesis in such species (see Malanowski 2016).

While a compromise between the archival and the constructive conceptions of mnemicity may be available, there can be no compromise between the content-based and the phenomenological conceptions of episodicity: either autonoesis is required for episodic memory or it is not. We acknowledge that the debate between the partisans of these conceptions is ongoing, but we nevertheless opt here for the content-based conception, for, as we will see below, opting for the phenomenological conception would all but trivially rule out the possibility of collective episodic memory. Partisans of the phenomenological conception are therefore likely to see any discussion of collective episodic memory as being simply misguided, and this chapter will make no attempt to convince them otherwise.

12.3 Emergence

Before turning to the evidence for the existence of genuinely collective—i.e., emergent—episodic memory, we set out the conception of emergence that we will employ.

12.3.1 Levels and Intensity of Interaction

As noted at the outset, there may be important differences between groups of different sizes with respect to whether they are capable of a given kind of memory. In principle, a group of any given size might be capable of any given kind of memory. In practice, however, the size of a group imposes constraints on the sorts of interactions that occur among its members and hence on its capacity for collective memory.

We begin by drawing a rough distinction among three levels of social interaction. The first is the level of the *individual*—the zero level of social interaction. At this level, the researcher investigates cognition as it is manifested in individual behaviour. The individual level is perhaps a theoretical idealization, but researchers nevertheless often build theories and models with this level in mind. The second level is that of the *small-scale group*. At this level, the researcher investigates cognition as it unfolds in relatively tight-knit, intensely interacting groups of two or a handful of individuals. Cognitive scientists working in the tradition of distributed cognition (Hutchins 1995) are often concerned with this level, as are cognitive psychologists working on collaborative remembering in mother-child dyads (Reese et al. 1993) or married couples (Harris et al. 2014), and it is with this level that we ourselves are primarily concerned here. The third level, finally, is the level of the *large-scale group*. At this level, the researcher investigates cognitive processes in loose-knit, weakly interacting groups on the scale of whole societies, that is, groups consisting of many thousands or millions of individuals. Memory at this level has so far been the province primarily of social scientists and historians but has recently begun to attract the attention of psychologists (Roediger and Abel 2015) and philosophers (Barash 2017), and we are to a lesser extent concerned with this level as well.[2]

Small-scale and large-scale groups characteristically differ not only in terms of their size but also in terms of the intensity of the interactions that take place among their members. Employing the vocabulary of the archival conception of mnemicity, Michaelian and Sutton (2018) have attempted to flesh out the intuitive notion of the

[2] The boundary between the small-scale group level and the large-scale group level may, of course, be vague, in the sense that there are intermediate cases, involving groups consisting of dozens, hundreds, or thousands of more or less intensely interacting individuals. Similarly, the boundary between the individual level and the small-scale group level may be vague, in the sense that individual behaviour always involves at least some degree of indirect interaction with others. The foregoing distinctions are thus meant only to be approximate; in our view, the development of a serious typology of the groups investigated in collective memory research is a promising area for future research in social ontology, though we will not attempt to make a case for this view here.

intensity of interactions in remembering groups. Both encoding and retrieval, they observe, might be either *parallel*, in the sense that group members encode/retrieve on their own, or *interactive*, in the sense that they encode/retrieve together, giving us four possibilities.[3] First, some groups are characterized by parallel encoding and parallel retrieval. The "nominal groups" employed in the collaborative recall paradigm (discussed below), for example, interact neither at encoding nor at retrieval and thus are not groups in any robust sense. Second, some groups are characterized by parallel encoding and interactive retrieval. This pattern of interaction is responsible for, inter alia, the influence of post-event misinformation on individual memory investigated by eyewitness memory researchers (see Loftus 2005). Third, some groups are characterized by interactive encoding and parallel retrieval. This pattern of interaction is illustrated by the case of a group of friends who experience an event together but later remember it individually (Sutton 2008). Groups characterized by parallel encoding and interactive retrieval or by interactive encoding and parallel retrieval are groups in a relatively robust sense, but it is the fourth pattern of interaction—interactive encoding and interactive retrieval—that is characteristic of the groups constituting transactive memory systems, durable groups whose members have established responsibilities for encoding and retrieving different kinds of information.

The examples given above are of remembering in small-scale groups, and the groups constituting transactive memory systems do indeed typically include at most a handful of individuals. Large-scale groups may involve indirect forms of interaction, but, due simply to the greater numbers of individuals involved, these groups are presumably always characterized by parallel encoding and parallel retrieval, since there is no way in practice for group members to interact directly with one another. We come back to this point below.

12.3.2 Criteria for Emergence

It is by no means obvious what conditions need to be satisfied before we are entitled to attribute emergent collective memory to a more or less intensely interacting small- or large-scale group. Nor is it obvious what such an attribution would mean.

Regarding the latter question, Wilson (2005; see also Barnier et al. 2008) has distinguished between three related but importantly different theses about the social or collective character of remembering. According to the *triggering thesis*, individual remembering is sometimes triggered by social interactions without strictly depending on those interactions. According to the *social manifestation thesis*, remembering is a social process in the sense that one person's remembering some-

[3] A more thorough treatment of intensity of interaction in groups would consider consolidation and storage in addition to encoding and retrieval, but Michaelian and Sutton's approach is sufficient for present purposes. Note that, while this approach begins from the archival conception rather than the constructive conception, it is intended merely to provide a way of zeroing in on the relevant types of interaction, and the constructive dimension of collective memory is considered below.

times depends for its occurrence on the presence of other people. According to the *group mind thesis*, remembering is a social process in the sense that groups themselves may sometimes remember. The triggering and social manifestation theses are relatively weak and should therefore be relatively uncontroversial. The group mind thesis is much stronger and correspondingly more controversial; our aim here is to explore the evidence for the thesis as it pertains to small- and large-scale collective memory. As Wilson remarks, philosophers have often been sceptical with respect to the group mind thesis, preferring to embrace one or another form of reductionism about collective memory, and the same goes for researchers in other disciplines. While many of the arguments in favour of reductionism about collective memory are persuasive, we will not respond directly to them here. Given the complexity of the relationship between emergence and reduction——both in general (O'Connor and Wong 2015) and in the special case of social entities (Zahle and Kaidesoja 2019)—doing so would simply take us too far afield. Rather than developing negative arguments in response to reductionism, therefore, we focus on making a positive case for the existence of emergent group-level memory.

Regarding the former question, we will adopt stringent criteria for the existence of group-level memory, in the hope that our argument will persuade even sceptics that some forms of collective memory are indeed an emergent property of the "social integrate" (Pettit 2003). In previous work (Michaelian and Arango-Muñoz 2018), we followed Theiner (2013) in applying Wimsatt's (1986) *mechanistic* criteria for emergence to the case of collective memory. On Wimsatt's approach, a property of a system is emergent to the extent that it fails to satisfy the following criteria. First, the property in question does not vary when some components of the system are replaced with others or with similar components from outside the system. Second, the property does not qualitatively change when components are added to or subtracted from the system. Third, the property does not vary when the system is disassembled or reassembled. Finally, the property is not affected by cooperative or inhibitory interactions among components of the system.

In virtue of its inclusion of this final criterion, Wimsatt's approach converges with a distinct mechanistic approach recently developed by Huebner (2014). On Huebner's approach, which both Huebner himself (2016) and Michaelian and Sutton (2017b) have applied to the case of collective memory, we are not to attribute collective mental states where, first, the relevant collective behaviour results from a top-down mechanism ensuring that the groups acts in accordance with the intentions of certain of its members, second, the collective behaviour straightforwardly results from simple rules governing individual behaviour, and, third, the members of the group have a mental capacity of the same kind as the mental capacity attributed to the group but the computations performed by the group are no more complex than those performed by its members. Huebner makes clear that what this final criterion requires is that the performance of the group be shaped by interactions among its members. Ultimately, then, both Wimsatt's approach and Huebner's emphasize *interaction* as a prerequisite for emergence, leading us to suspect that there may be grounds for attributing a capacity for memory to groups characterized by interactive encoding and interactive retrieval.

The two sets of mechanistic criteria considered so far are already fairly demanding, but the *rational* criteria proposed by Szanto (2014) are even more so. On Szanto's approach, the group mind thesis is plausible in a given case to the extent that the relevant group has a rationally unified point of view, where a group must satisfy three requirements in order to qualify as having such a point of view. First, it must be able to "form, hold and robustly entertain intentional states […] with representational and propositional content", and its behaviour must be explainable in terms of these states. Second, the group mind "holds holistically construed, relatively consistent, non-contradictory beliefs". Finally, the group mind must "integrate intentional states so as to constitute on overall rationally unified point of view, i.e., a unified set of reasons, in the light of which the group assesses its given beliefs, preferences, and intentions". What Szanto's approach adds to the two approaches considered above is a set of further criteria: *intentionality*, *coherence*, and *integration*. In the case of memory, we note, a group's rationally unified point of view is plausibly understood as being not the starting point for contemplating the past but rather a possible end point of an interactive process in which the members of the group draw on the information that each of them stores in order to construct a shared representation of the shared past. In what follows, we will make use of both Wimsatt's and Huebner's mechanistic approaches and Szanto's rational approach.

12.4 Emergent Episodic Memory

With a conception of emergence in place, we turn to the evidence for the existence of emergent collective episodic memory.

12.4.1 Memory in Small-Scale Groups

In view of the differences, highlighted above, between small-scale groups and large-groups, we treat these separately, beginning with the former.

12.4.1.1 Emergence

One key source of evidence regarding small-scale groups is provided by research demonstrating that *conversational remembering* can reshape the memories of both speakers and their audiences, with this reshaping being due in part to retrieval-induced forgetting (Hirst and Echterhoff 2008; Stone et al. 2012). *Within-individual retrieval-induced forgetting* occurs when retrieval of an item by a subject strengthens his memory for the retrieved item and causes forgetting of related items. *Socially shared retrieval-induced forgetting* occurs when retrieval of an item by a speaker similarly causes forgetting of related items in his listeners, an effect that appears to

occur primarily because listeners covertly retrieve the same information as the speaker, thus allowing the mechanisms that lead to within-individual retrieval-induced forgetting to reshape their memories. In groups—such as married couples—characterized by frequent, ongoing conversational remembering, socially shared retrieval-induced forgetting may lead to convergence on shared memory representations. However, though the mechanisms in question are activated by interactions among group members as they remember together, this convergence is ultimately driven by individual-level mechanisms. This form of shared memory thus appears not to satisfy mechanistic criteria for emergence, and it may lend support to the triggering thesis or the social manifestation thesis rather than the group mind thesis.

Another source of evidence is provided by research on *collaborative recall*, which has consistently identified two superficially opposed effects (Weldon 2000; Rajaram and Pereira-Pasarin 2010). On the one hand, the quantity of information recalled by a group is often greater than that recalled by any of its members individually. This first effect (not to be confused with the collaborative facilitation described below) occurs simply because group members recall nonoverlapping sets of items. On the other hand, the quantity of information recalled by a group of interacting individuals (a "real group") is often less than that recalled by a set of noninteracting individuals (a "nominal group"). This second effect—known as *collaborative inhibition*—occurs because individuals recall less when remembering together than when remembering alone, a tendency that appears to be due to retrieval disruption, in which incompatible retrieval strategies employed by different group members interfere with each other. Though retrieval disruption occurs only due to the interactions that take place among group members as they remember together, the retrieval disruption hypothesis ultimately appeals to the disruption of individual-level mechanisms. This form of shared memory thus likewise appears not to satisfy mechanistic criteria for emergence and may lend support to the triggering thesis or the social manifestation thesis rather than the group mind thesis.

The groups that figure in research on *transactive memory* (Wegner 1987; Wegner et al. 1991; Hollingshead et al. 2011), in contrast, appear to satisfy not only the mechanistic criteria but also the rational criteria for emergence reviewed above. Broadly speaking, a transactive memory system consists of two components (Theiner 2013). First, its representational component includes both the first-order (declarative and procedural) memories of its members and their metacognitive knowledge of each other's memories. Second, its procedural component includes the various (implicit and explicit) communication processes through which group members assign responsibility for and coordinate performance of the stages of the memory process. Crucially, transactive memory systems often perform better than individuals do on their own, at least when performing tasks that lend themselves to a division of cognitive labour. Theiner (2013) has argued that transactive memory system satisfy all four of Wimsatt's criteria. First, because the members of the system have nonoverlapping memories, they are not interchangeable and normally will not be interchangeable with individuals from outside the system. Second, if enough members of the system are removed, the system will fail, again due to its members'

non-overlapping knowledge. Third, the system's history makes a difference to its performance—in particular, teams trained together tend to perform better—so disassembly and reassembly may affect its ability to remember. Finally, since the system is welded together in part by its members' metacognitive knowledge of "who knows what" within the system, cooperative and inhibitory interactions among its members are critical to its functioning. Huebner (2016) has similarly argued that transactive memory systems satisfy his own mechanistic criteria, including the critical final criterion: since transactive memory systems outperform groups of noninteracting individuals, the interactions among the members of a transactive memory system are key to explaining its performance.

Transactive memory systems thus appear to satisfy both Wimsatt's and Huebner's mechanistic criteria for emergence. They appear, moreover, to satisfy Szanto's more demanding rational criteria. Each individual takes the other as a potential source of information concerning the aimed memory and both aim at retrieving or reconstructing the same intentional content; thus, they satisfy the intentionality criterion. Moreover, they aim to retrieve or reconstruct a representation that is consistent, coherent and non-contradictory; that is, if there are conflicts among the retrieved information, the subjects are committed to resolve it; thus, they satisfy the coherence criterion. These features of the transactive group interaction ensure the fulfilment of the integration requirement, according to which the group should share a unified point of view. However, this unified point of view is not the starting point of the reconstructive memory process, but the arrival point after joint construction. That is, starting from different points of view (different perspectives or versions about the past episode), and trying to reach a coherent version of the past episode out of the different versions, the subjects negotiate the details of a version of the past in which all then should agree. Wegner's experiments on transactive remembering clearly illustrates this aim at a unified point of view: when members of the group retrieve conflicting memories, they negotiate till they arrive to a version that is endorsed by all (Wegner 1987; Wegner et al. 1991). We consider an example in detail below.

We do not pretend that this brief review of the evidence is decisive (see Michaelian and Arango-Muñoz 2018 for a more detailed review), but we do take it to be sufficient to establish that there is a prima facie case to be made for emergence in small-scale groups, particularly in transactive memory systems. If transactive memory systems satisfy plausible criteria for emergence, however, it remains to be seen whether the emergent activity that they perform satisfies criteria for episodicity and mnemicity.

12.4.1.2 Episodicity and Mnemicity

It is useful, in this connection, to consider research on *collaborative facilitation*. Under conditions in which group members are likely to employ similar or complementary retrieval strategies, collaborative inhibition can be overcome or even reversed, enabling real groups to recall more than nominal groups. Thus collabora-

tive facilitation has been found with groups of experts in a given domain, such as airline pilots (Meade et al. 2009). In addition to the *quantitative* collaborative facilitation that has been the focus of research on groups of experts, Harris et al. (2014, 2017) have investigated what might be thought of as *qualitative* collaborative facilitation in married couples. Whereas, in quantitative collaborative facilitation, the members of the group are able to recall more items from a given domain when remembering together than when remembering alone, in qualitative collaborative facilitation they are able to recall more or different information about specific items. Harris et al. have demonstrated, in particular, that couples—specifically, long-married couples—have a tendency to "go episodic" when remembering together, in the sense that they tend to recall additional details of events that they experienced together, even when they have been explicitly instructed to recall as many events as possible. Interestingly, qualitative collaborative facilitation does not presuppose quantitative collaborative facilitation: with respect to the task instruction, the couples studied by Harris et al. experience collaborative inhibition, but in an important sense they nevertheless benefit from remembering together.

In line with our suggestion above that emergence depends on interaction and that intensity of interaction can be understood in terms of interactive vs. parallel encoding and retrieval, transcripts of couples' conversations indicate that qualitative collaborative facilitation occurs due to interactions that take place during collaborative recall—i.e., at the time of retrieval—and that the relevant interactions are only possible because the members interacted at the time of encoding. Consider the following representative case (Harris et al. 2014: 290–291).

> Interviewer: And how many more trips did you do? There's the Greek Islands.
> Wife: South America.
> Husband: We did South America, yes, we did Peru and Brazil and Argentina and Bolivia and The Andes. We went up to …
> H: Do you remember munching the coca leaf to try …
> W: Oh yes.
> H: We went up to The Andes at 5000 m, and munching coca leaf, and [wife] decided that she needed to have a pee.
> W: So we were on the road here, you see, but the little latrine was up on the top.
> H: It was about 50 m higher.
> W: So we had to climb up from the road.
> H: So I said, alright, I'll take you up there. By the time I got down, which at 5000 m climbing, I'd just about had it.
> W: Yes, we thought we were going to faint, but we didn't. But those coca leaves were very good, I rather liked them.

The emergent activity performed by this and similar couples would appear to satisfy our criteria for both episodicity and mnemicity.

Beginning with episodicity, what this couple does as it "goes episodic" would, given the content-based conception of episodicity, which characterizes episodic memory in terms of what-when-where information, appear to qualify as episodic. There is clearly representation of episodic details—information about the what, the when, and the where of the event in the Andes—at the level of the husband and wife taken individually. But there also appears to be representation of episodic details at the level of the couple taken as a group. The husband and wife agree on a representa-

tion about what happened when they were in the Andes, but they do not merely agree on the representation; that is, this is not a case of merely shared episodic thought. Instead, the representation on which they agree is itself the product of their interaction as they remember together; that is, this is a case of properly collective episodic thought. What we witness, in this transcript and in others provided by Harris et al., is the unfolding of a process in which additional episodic detail emerges as husband and wife each draw on their metacognitive knowledge of what the other knows in order to provide the other with cues that elicit new information and encourage agreement, allowing them to converge on a shared representation of a shared experience.

Indeed, Harris et al. (2014) argue that the qualitative collaborative facilitation revealed by their studies includes more than one kind of emergence. First, they identify emergence of *new details*, in which information that neither individual could recall alone becomes available due to interaction during retrieval. This is the kind of emergence that we have been emphasizing so far. Second, they identify emergence of *quality*, in which remembering is emotionally richer and more vivid when members remember together than when they remember alone. Finally, they identify emergence of *understanding*, in which members' interpretations of a given event are transformed when they remember together. It is important to note that not all couples go episodic or go episodic to the same degree, and, in line with Wegner's definition of transactive memory systems as being constituted by both "the operation of the memory systems of the individuals" *and* "the processes of communication that occur within the group" (1987: 191), Harris et al. (2014) observe that these differences between couples appear to be due to differences in interaction style, including differences in intensity and style of communication. The tendency of long-married couples to go episodic thus provides particularly clear evidence for a form of collective thought that is both emergent and, given the content-based conception of episodicity, episodic.

It is also important to note that we do not witness, in Harris et al.'s transcripts, the emergence of *phenomenology*. The phenomenological conception of episodicity would therefore classify the emergent activity performed by couples when they "go episodic" as not being genuinely episodic. Indeed, we suggested above that the phenomenological conception of episodicity all but trivially rules out the possibility of collective episodic memory. As Michaelian and Sutton (2017b) have argued, our reluctance to attribute mental states to groups may be explained in part by our reluctance to attribute phenomenal consciousness to groups. Some have suggested that this reluctance is unfounded, arguing that materialists are bound to admit that collective phenomenal consciousness is possible in principle (Schwitzgebel 2015). Others, however, have pointed out that there is no reason to think that any actual groups are organized in a manner sufficient to give rise to collective phenomenal consciousness in practice (List 2018). This certainly goes for the groups in question here: we have offered no reason to take the notion of collective autonoetic consciousness seriously, and we do not imagine that such a reason might be offered. Thus, given the phenomenological conception of episodicity, even couples who go episodic will not qualify as engaging, at the group level, in a form of episodic thought.

Turning to mnemicity, the activity performed by the couples in question would, given the compromise conception of mnemicity, which incorporates both the archival and the constructive conceptions, appear to qualify as memory. According to the archival conception, on the one hand, remembering is essentially a matter of the transmission of content from experience to retrieval. It is clear that transmission occurs at the level of the husband and wife taken individually; assuming that they are remembering, which is not at issue here, we know this simply because we know that remembering involves transmission. But transmission also appears to occur at the level of the couple taken as a group, in the sense that group-level mechanisms are responsible for the production of a portion of the content that was available at the time of experience. As Harris et al. stress, the couple is able to produce details that neither of its members is able to produce on his own; in other words, while the individual members of the group transmit information, they are able to transmit some of the information that they transmit only insofar as they are members of the group. According to the constructive conception, on the other hand, remembering is essentially a matter of the construction of a representation that is more or less adequate with respect to the original experience. It is clear that construction occurs at the level of the husband and wife taken individually—again, assuming that they are remembering, we know this simply because we know that remembering involves construction. But construction also appears to occur at the level of the couple taken as a group, in the sense that group-level mechanisms are responsible for the emergence of episodic detail, emotional richness, and understanding that were not available at the time of experience. Regarding episodic detail, it is likely that some of the details present at the time of retrieval were not present at the time of encoding. Regarding emotional richness, it will often be the case that the emotional tone of the memory does not correspond to the emotions that either of the members of the couple felt at the time of experience. And regarding understanding, this is something that is negotiated between members and often continues to evolve even when the event is well in the past, as the couple renegotiates its interpretation of the significance of the event.

Overall, then, given the content-based conception of episodicity and the compromise conception of mnemicity, we can conclude that some small-scale groups—in particular, transactive memory systems of the sort constituted by some long-married couples—are capable of a form of collective thought that is emergent, episodic, and mnemic, i.e., that they are capable of genuinely collective episodic memory.

12.4.2 Memory in Large-Scale Groups

Once one has come to this conclusion, one may be tempted to jump immediately to the further conclusion that large-scale groups are also capable of collective episodic memory. A brief analysis, however, reveals that large-scale groups are unlikely to be capable of memory, never mind episodic memory.

Adopting Szanto's rational approach to emergence, on which intentionality, coherence, and integration are key, it is doubtful that large-scale groups display emergent mental states, whether mnemic or otherwise. Even if societies can be said to remember in a loose sense, their members do not actively interact and collaborate with the aim of reconstructing a particular past event; thus they do not satisfy the intentionality criterion. Nor do they satisfy the coherence criterion, since their members are in general not committed to the resolution of inconsistencies or contradictions, leading to situations in which different subgroups within a given society remember significantly different or even outright incompatible versions of the same event. When, for example, Brescó and Wagoner (2016) studied the peace process that took place in the Basque Country in 2006, they found that three versions of the same fact circulated in the society, varying according to the political view of the individual or group of individuals. There was no integration among the three narratives, and this case therefore also does not satisfy the integration requirement. While we cannot decisively show here that Szanto's criteria are never satisfied, the literature abounds with similar cases, and this would seem to be the safest bet.

Adopting Wimsatt's and Huebner's mechanistic approach to emergence, large-scale groups may display a variety of emergent features, and possibly even emergent mental states, but there is, despite the ongoing "boom" in studies of large-scale collective memory (Blight 2009), little reason to take them to display emergent memory in particular. Some treatments of large-scale collective memory adopt purely archival conceptions of mnemicity. Such conceptions are, as noted above, inadequate, and we can therefore simply set these treatments aside. Others adopt a more constructive conception. The constructive conception is more promising, but, on closer inspection, the constructive activities performed by large-scale groups have little in common with those performed by individuals or small-scale groups. Michaelian (2014), for example, responding to Anastasio et al.'s (2012) argument for the existence, at the level of large-scale groups, of a form of collective memory consolidation analogous to individual memory consolidation, argues that, due in part to the role in large-scale collective remembering of external memory representations and in part to the role of the kind of conflict highlighted by Brescó & Wagoner (2016), collective consolidation differs dramatically from individual consolidation in that, while the latter is correctly understood as a transition from labile, short-term representations to stable, long-term representations, the former is more adequately understood as a transition from stable, long-term representations to labile, short-term representations.[4] While we cannot review the disanalogies between "memory" at the large-scale group level and memory at the small-scale group and individual levels in detail, the view that large-scale groups are not capable of memory would, again, seem to be the safest bet.

Thus while there may be merely shared memories at the level of a large-scale group—similar memories held by the individual members of the group—and while

[4] There has been, as far as we are aware, no investigation of group-level consolidation in small-scale groups, and we grant that such investigation might turn up disanalogies between individual consolidation and small-scale group-level consolidation.

the individuals in question may have such memories in part due to the effect of group-level processes on what they remember, exemplifying the sort of triggering or social manifestation described by Wilson, we can conclude that it would nevertheless be a mistake to see the group itself as remembering. This conclusion echoes Wegner's remark that the concept of transactive memory cannot usefully be applied to large-scale groups such as societies, since treating societies as transactive memory systems would simply "make transactive memory into a synonym for culture" (Wegner et al. 1985: 257); our argument here suggests that this goes not only for the concept of *transactive* memory but also for the more basic concept of *memory*.

If large-scale groups are incapable of *remembering*, they are a fortiori incapable of remembering *episodically*. Even if we were willing to countenance the existence of memory in large-scale groups, however, Michaelian and Sutton (2018) have pointed out that "[m]emory in large-scale groups is typically memory for events which are of concern to the individuals who make up the group but in which those individuals did not necessarily take an active part and of which they often have only indirect knowledge". They thus observe that large-scale memory thus appears to be semantic rather than episodic: individual group members "may (episodically) remember personal experiences which are linked to the events in question, but, to the extent that remembering is concerned with large-scale, public events, it lacks the characteristic features of episodic memory". Putting this in terms of the intuitive notion of intensity of interaction outlined above, large-scale groups typically involve no interaction at the time of encoding and at most indirect interaction at the time of retrieval; remembering in such groups is thus unlikely to display emergent episodicity.

Overall, then, it is relatively unlikely that large-scale groups display emergent mental states; if they do display such mental states, it is unlikely that they are capable of remembering; and if they are capable of remembering, it is unlikely that they are capable of remembering episodically.

12.5 Metamemory

With our defence of our first thesis—that small-scale but not large-scale groups are capable of a form of shared mental agency through which genuinely collective memory emerges—we turn to the defence of our second thesis, namely, that metacognition is, in the relevant groups, an individual-level capacity, beginning with some brief background on the nature of metacognition. Many researchers argue that metacognition is a prerequisite for group-level memory. Shea et al. (2014), for example, see metacognition as establishing a communicative interface that regulates collaborative work. In a similar vein, Heyes (2016) understands social learning strategies as being based on metacognition: knowing who knows what. And, as already noted, Wegner (1987) claims that metacognition is necessary for transactive memory systems. The question thus arises whether the form of shared agency manifested in collective memory implies *group-level* metacognition or only *individual-level* metacognition.

Metacognition can be defined as the capacity to *monitor* and *control* cognitive processes (Nelson and Narens 1990). Monitoring can occur either before or after the occurrence of the relevant cognitive process; the former sort of monitoring is a matter of *self-prediction*, the latter a matter of *post-evaluation* (Proust 2013). Self-prediction, as it pertains to memory, refers to an assessment of the likelihood that one will be able to retrieve or reconstruct an item of information, whereas post-evaluation refers to an assessment of the accuracy of a retrieved or reconstructed item. Control, in turn, refers to actions taken on the basis of monitoring: either accepting the outcome of the memory process or rejecting it. If, for example, someone asks me the address of the University of Antioquia, self-predictive metamemory assesses whether or not I know it. In some cases, metamemory determines that I can retrieve it (monitoring), and I therefore attempt to remember it (control), even if the address does not immediately come to mind. In other cases, metamemory determines that I cannot retrieve it (monitoring), and I therefore consider alternative strategies, such as asking someone or looking it up online (control). If I have retrieved an address, say, "Calle 67, Medellín", metamemory post-evaluates the retrieved item (monitoring), determining whether I should endorse it or instead reject it and try again (control).

The recent literature on metacognition distinguishes between two levels or types of metacognition: low-level and high-level (Arango-Muñoz 2011; Koriat 2007; Proust 2013; Shea et al. 2014). Each of these levels has a different structure, a different content, and a different function in the cognitive architecture. On the one hand, *low-level* metacognition is based on *metacognitive feelings*, such as the feeling of knowing, the tip-of-the-tongue state, and the feeling of forgetting. These experiences are used to monitor and control memory: the feeling of knowing motivates the subject to remember, whereas the feeling of forgetting motivates her to look for complementary strategies (Arango-Muñoz and Michaelian 2014). Thus, they guide memory retrieval in a direct, immediate manner. On the other hand, *high-level* metacognition employs psychological concepts and theory of mind to understand memory, to rationalize behavior, and sometimes to control memory; that is, it is *metarepresentational* in character, rather than feeling-based.

12.6 Emergent Metamemory?

In this section, we consider whether and how these two types of metacognition are implicated in collective memory.

12.6.1 Metamemory in Small-Scale Groups

Beginning with small-scale groups, we consider first low-level metamemory and then high-level metamemory.

12.6.1.1 Low-Level Metamemory

The metacognitive feelings described above are feelings that refer to the subject's own cognitive capacities and their outputs. In addition to *self-directed* feelings of this sort, there may be *other-directed* metacognitive feelings that refer to the cognitive capacities and outputs of other subjects, and such feelings may be involved in transactive memory systems. Wegner, in particular, has suggested that, when one subject trusts another subject's memory—i.e., when he defers responsibility for remembering a given item or category of information to the other subject—the relevant feeling of trust can be seen as the social counterpart of the feeling of knowing (1987: 198; Wegner et al. 1985: 266; cf. Huebner 2016). Along the same lines, he has pointed out that, following the dissolution of a transactive memory system of which a subject was a member, the subject may experience a feeling of indecision as he progressively becomes aware of the loss of access to information held by the other member or members of the system (Wegner et al. 1985: 273) Although other-directed metacognitive feelings are interesting in their own right and are certainly worthy of further study, they are clearly not group-level metacognitive feelings: other-directed feelings of knowing and indecision may in a sense refer to the group, but they are nevertheless unambiguously experienced by the individual. Thus they are most parsimoniously attributed to the individual rather than the group, whereas it is properly group-level feelings the existence of which would have to be established in order to establish the existence of low-level group-level metamemory.

It is doubtful that a convincing case can be made for the existence of such feelings, for their existence would entail the existence of group-level phenomenal consciousness, and, as we saw above in our discussion of the notion of collective autonoetic consciousness, group-level phenomenal consciousness can be ruled out as a practical reality, even if it cannot be ruled out as an in-principle possibility. It is worth noting here that even theorists who take the idea of collective mental states of other sorts quite seriously tend to be sceptical with respect to the possibility of collective phenomenal states (e.g., Clark 2009). And even those who take the idea of collective phenomenal states seriously tend to be sceptical with respect to their prevalence. Krueger (2014), for example, distinguishes between extended emotions and collective emotions. Examples of extended emotions are cases in which the stimulus causing the emotion is essential to the production of the experience, as sometimes occurs when we listen to music. Krueger takes it that there are bona fide cases of extended emotion. Examples of collective emotions are cases in which a single emotion is realized in more than one subject. Krueger suggests that infant-caregiver emotional interconnection provides one plausible case of collective emotion, but he acknowledges that it is likely to be the only exception to the rule that "there are non-transferable phenomenal aspects of emotional experience in adulthood that seem to preclude their being collectively realized" (551). Absent any reason to suppose that collective metacognitive feelings constitute an additional exception to this rule, we are entitled to rule out their existence.

12.6.1.2 High-Level Metamemory

In the individual case, high-level metacognition crucially involves a metarepresentational structure deployed by the subject to self-ascribe mental properties. As Carruthers describes it, this type of metacognition is a matter of "turning mindreading capacities upon oneself" (2009: 3). High-level metacognition also plausibly plays a role in the case of transactive memory: because it provides a representational format suitable for the construction of complex intentions and purposes that involve other subjects and their mental states, high-level metamemory would appear to be more directly relevant than low-level metamemory to the functioning of transactive memory systems. Theiner (2013: 72) describes the role of this metarepresentational structure in the collective case as follows. "The awareness of who knows what in a group constitutes a special kind of higher-order knowledge because they are about other memory structures, rather than about things as such. [T]ransactive memories are meta-memories about memories which one does not possess … but which are believed to be held by other group members". Thus the members of a transactive memory system have metarepresentational knowledge that allows them to coordinate their contributions to collective remembering with those of other members of the system.

Wegner understands transactive remembering as including three stages: transactive *encoding*, transactive *storage*, and transactive *retrieval*.[5] Transactive encoding involves a decision as to which member of the group is to be responsible for learning a given item of information or information belonging to a given category, a decision that presupposes knowledge of the encoding, storage, and retrieval abilities of each group member; thus group members must be represented not just as generic mnemic resources but rather as resources with specific capacities. Transactive storage requires a means of keeping track of which group member has learnt what and thus of which information or kind of information a given group member can be expected to provide; again, this presupposes a representation of group members not as generic stores of information but rather as sources of specific items or types of information. And transactive retrieval requires the integration of information provided by different group members to produce a consensus representation; at this stage, too, group members must be represented not as generic sources of information but rather as sources having particular domains of authority or competence. Thus metarepresentational knowledge plausibly plays a role at all three stages of transactive remembering. It would, however, be a mistake to attribute this metarepresentational knowledge to the system itself, just as it would have been a mistake to attribute metacognitive feelings to the system. Interaction among members of the system does not produce the high-level metacognition that is involved in transactive memory—instead, high-level (and possibly low-level) metacognition makes their interaction possible. In short, metarepresentational knowledge, like metacognitive feelings, is most parsimoniously attributed to the individual rather than the group.

[5] We noted above that there has been no discussion of small-scale group-level consolidation in general, and this goes for transactive consolidation in particular.

12.6.2 Metamemory in Large-Scale Groups

Though we have already ruled out the possibility of emergent processes in large-scale groups that, strictly speaking, satisfy criteria for mnemicity, we have acknowledged that such groups may, given mechanistic criteria for emergence, manifest emergent mental states of other sorts and that it may be possible to speak in a loose sense of large-scale collective memory. Before concluding, we therefore briefly discuss the possibility of emergent metamemory in large-scale groups.

In certain cases, it may seem that a subgroup within the larger society carries out metacognitive monitoring and control. We have in mind, for example, the role assigned by Anastasio et al. (2012) to what they refer to as groups of "opinion leaders", in which intellectuals, journalists, historians, and like individuals actively shape a society's memory for key events of its past, determining what is remembered and how it remembered. The role of groups of opinion leaders in directing large-scale collective consolidation is certainly interesting, but, even in cases where groups of opinion leaders play this role, we are not entitled to attribute the relevant monitoring and control processes to the society as a whole, even given mechanistic criteria for emergence. One of Huebner's criteria for collective mentality, recall, is that we are not to "posit collective mentality where collective behavior results from an organizational structure set up to achieve the goals or realize the intentions of a few powerful and/or intelligent people" (2014: 21). Since this sort of structure would seem to be at work in cases of the sort described by Anastasio et al., all that needs to be explained in these cases can be explained by appealing to the decision process of the members of the relevant subgroup, without any appeal to collective metacognition.

In sum, neither small-scale collective memory nor large-scale collective memory appears to involve either low-level or high-level group-level metacognition.

12.7 Conclusions

Our aim in this chapter has been to delineate a form of shared agency that we take to be manifested in collective memory. We have argued for two theses. First, we have argued that, given a relatively weak conception of episodicity, certain small-scale groups display a form of emergent (i.e., genuinely collective) episodic memory, while large-scale groups, in contrast, do not display emergent episodic memory. Second, we have argued that this form of emergent memory presupposes (high-level and possibly low-level) metamemorial capacities, capacities that are, however, not themselves emergent group-level features but rather strictly individual-level features. The form of shared agency that we have delineated is thus revealed as being minimal in three senses. First, the relevant groups are themselves minimal in terms of their size. Second, the form of memory in question is minimally episodic. And finally, the cognitive capacities attributed to the relevant groups are minimal, in the sense that they need not themselves be capable of metacognition.

References

Anastasio, T. J., Ehrenberger, K. A., Watson, P., & Jiang, W. (2012). *Individual and collective memory consolidation: Analogous processes on different levels*. Cambridge, MA: MIT Press.

Arango-Muñoz, S. (2011). Two levels of metacognition. *Philosophia, 39*(1), 71–82.

Arango-Muñoz, S., & Bermúdez, J. P. (2018). Remembering as a mental action. In K. Michaelian, D. Debus, & D. Perrin (Eds.), *New directions in the philosophy of memory* (pp. 75–96). London/New York: Routledge.

Arango-Muñoz, S., & Michaelian, K. (2014). Epistemic feelings, epistemic emotions: Review and introduction to the focus section. *Philosophical Inquiries, 2*(1), 97–122.

Barash, J. A. (2017). Collective memory. In S. Bernecker & K. Michaelian (Eds.), *The Routledge handbook of philosophy of memory* (pp. 255–267). London/New York: Routledge.

Barnier, A. J., Sutton, J., Harris, C. B., & Wilson, R. A. (2008). A conceptual and empirical framework for the social distribution of cognition: The case of memory. *Cognitive Systems Research, 9*(1), 33–51.

Blight, D. W. (2009). The memory boom: Why and why now? In P. Boyer & J. V. Wertsch (Eds.), *Memory in mind and culture* (pp. 238–251). New York: Cambridge University Press.

Brescó, I., & Wagoner, B. (2016). Context in the cultural psychology of remembering: illustrated with a case study of conflict in national memory. In C. Stone & L. Bietti (Eds.), *Contextualizing human memory: An interdisciplinary approach to understanding how individuals and groups remember the past* (pp. 69–85). London/New York: Routledge.

Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences, 32*(2), 121–138.

Clark, A. (2009). Spreading the joy? Why the machinery of consciousness is (probably) still in the head. *Mind, 118*(472), 963–993.

Harris, C. B., Barnier, A. J., Sutton, J., & Keil, P. G. (2014). Couples as socially distributed cognitive systems: Remembering in everyday social and material contexts. *Memory Studies, 7*(3), 285–297.

Harris, C. B., Barnier, A. J., Sutton, J., Keil, P. G., & Dixon, R. A. (2017). "Going episodic": Collaborative inhibition and facilitation when long-married couples remember together. *Memory, 5*(8), 1148–1159.

Heyes, C. (2016). Who knows? Metacognitive social learning strategies. *Trends in Cognitive Sciences, 20*(3), 204–213.

Hirst, W., & Echterhoff, G. (2008). Creating shared memories in conversation: Toward a psychology of collective memory. *Social Research, 75*(1), 183–216.

Hollingshead, A. B., Yoon, N. G. K., & Brandon, D. P. (2011). Transactive memory theory and teams: Past, present, and future. In E. Salas, S. M. Fiore, & M. Letsky (Eds.), *Theories of team cognition: Cross-disciplinary perspectives* (pp. 421–455). New York: Taylor and Francis.

Huebner, B. (2014). *Macrocognition: A theory of distributed minds and collective intentionality*. New York: Oxford University Press.

Huebner, B. (2016). Transactive memory reconstructed: Rethinking Wegner's research program. *Southern Journal of Philosophy, 54*(1), 48–69.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science, 19*(3), 265–288.

Kirchhoff, M. (2016). Composition and transactive memory systems. *Philosophical Explorations, 19*(1), 59–77.

Klein, S. B. (2015). What memory is. *Wiley Interdisciplinary Reviews: Cognitive Science, 6*(1), 1–38.

Klein, S. B., & Nichols, S. (2012). Memory and the sense of personal identity. *Mind, 121*(483), 677–702.

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). Cambridge, UK: Cambridge University Press.

Krueger, J. (2014). Varieties of extended emotions. *Phenomenology and the Cognitive Sciences, 13*(4), 533–555.
List, C. (2018). What is it like to be a group agent? *Noûs, 52*(2), 295–319.
Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12*(4), 361–366.
Mace, J. H. (2007). Involuntary memory: Concept and theory. In J. H. Mace (Ed.), *Involuntary memory* (pp. 1–19). Malden: Blackwell.
Mahr, J. B., & Csibra, G. (2018). Why do we remember? The communicative function of episodic memory. *Behavioral and Brain Sciences, 41*, e1.
Malanowski, S. (2016). Is episodic memory uniquely human? Evaluating the episodic-like memory research program. *Synthese, 193*(5), 1433–1455.
Manier, D., & Hirst, W. (2008). A cognitive taxonomy of collective memories. In A. Erll, A. Nünning, & S. B. Young (Eds.), *Cultural memory studies: An international and interdisciplinary handbook* (pp. 253–262). Berlin: Walter de Gruyter.
Meade, M., Nokes, T. J., & Morrow, D. G. (2009). Expertise promotes facilitation on a collaborative memory task. *Memory, 17*(1), 39–48.
Michaelian, K. (2014). Review essay on Anastasio et al., Individual and collective memory consolidation: Analogous processes on different levels. *Memory Studies, 7*(2), 254–264.
Michaelian, K. (2016). *Mental time travel: Episodic memory and our knowledge of the personal past*. Cambridge, MA: MIT Press.
Michaelian, K., & Arango-Muñoz, S. (2018). Collaborative memory knowledge: A distributed reliabilist perspective. In M. Meade, C. B. Harris, P. Van Bergen, J. Sutton, & A. J. Barnier (Eds.), *Collaborative remembering: Theories, research, applications* (pp. 231–247). Oxford: Oxford University Press.
Michaelian, K., & Robins, S. (2018). Beyond the causal theory? Fifty years after Martin and Deutscher. In K. Michaelian, D. Debus, & D. Perrin (Eds.), *New directions in the philosophy of memory* (pp. 13–32). London/New York: Routledge.
Michaelian, K. & Sutton, J. (2017a). Memory. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. http://plato.stanford.edu/entries/memory/
Michaelian, K., & Sutton, J. (2017b). Collective mental time travel: Remembering the past and imagining the future together. *Synthese*. https://doi.org/10.1007/s11229-017-1449-1
Michaelian, K., & Sutton, J. (2018). Collective memory. In M. Jankovic & K. Ludwig (Eds.), *The Routledge handbook of collective intentionality* (pp. 140–151). London/New York: Routledge.
Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation, 26*, 125–173.
O'Connor, T. & Wong, H. Y. (2015). Emergent properties. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/entries/properties-emergent/
Olick, J. K., Vinitzky-Seroussi, V., & Levy, D. (2011). *The collective memory reader*. New York: Oxford University Press.
Perrin, D., & Rousset, S. (2014). The episodicity of memory. *Review of Philosophy and Psychology, 5*(3), 291–312.
Pettit, P. (2003). Groups with minds of their own. In F. Schmitt (Ed.), *Socializing metaphysics* (pp. 167–193). Lanham: Rowman & Littlefield.
Proust, J. (2013). *The philosophy of metacognition: Mental agency and self-awareness*. New York: Oxford University Press.
Rajaram, S., & Pereira-Pasarin, L. P. (2010). Collaborative memory: Cognitive research and theory. *Perspectives on Psychological Science, 5*(6), 649–663.
Reese, E., Haden, C. A., & Fivush, R. (1993). Mother-child conversations about the past: Relationships of style and memory over time. *Cognitive Development, 8*(4), 403–430.
Ren, Y., & Argote, L. (2011). Transactive memory systems 1985–2010: An integrative framework of key dimensions, antecedents, and consequences. *The Academy of Management Annals, 5*(1), 189–229.
Robins, S. K. (2016). Misremembering. *Philosophical Psychology, 29*(3), 432–447.

Roediger, H. L., & Abel, M. (2015). Collective memory: A new arena of cognitive study. *Trends in Cognitive Sciences, 19*(7), 359–361.

Schwitzgebel, E. (2015). If materialism is true, the United States is probably conscious. *Philosophical Studies, 172*(7), 1697–1721.

Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences, 18*(4), 186–193.

Stone, C. B., Coman, A., Brown, A. D., Koppel, J., & Hirst, W. (2012). Toward a science of silence: The consequences of leaving a memory unsaid. *Perspectives on Psychological Science, 7*(1), 39–53.

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences, 30*(3), 299–313.

Sutton, J. (2008). Between individual and collective memory: Coordination, interaction, distribution. *Social Research, 75*(1), 23–48.

Szanto, T. (2014). How to share a mind: Reconsidering the group mind thesis. *Phenomenology and the Cognitive Sciences, 13*(1), 99–120.

Theiner, G. (2013). Transactive memory systems: A mechanistic analysis of emergent group memory. *Review of Philosophy and Psychology, 4*(1), 65–89.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–402). New York: Academic Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne, 26*(1), 1–12.

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–208). New York: Springer.

Wegner, D. M., Giuliano, T., & Hertel, P. T. (1985). Cognitive interdependence in close relationships. In W. Ickes (Ed.), *Compatible and incompatible relationships*. New York: Springer.

Wegner, D. M., Erber, R., & Raymond, P. (1991). Transactive memory in close relationships. *Journal of Personality and Social Psychology, 61*(6), 923–929.

Weldon, M. S. (2000). Remembering as a social process. *Psychology of Learning and Motivation, 40*, 67–120.

Werning, M., & Cheng, S. (2017). Taxonomy and unity of memory. In S. Bernecker & K. Michaelian (Eds.), *The Routledge handbook of philosophy of memory* (pp. 7–20). London/New York: Routledge.

Wilson, R. A. (2005). Collective memory, group minds, and the extended mind thesis. *Cognitive Processing, 6*(4), 227–236.

Wimsatt, W. (1986). Forms of aggregativity. In A. Donagan, J. Perovich, N. Anthony, & M. V. Wedin (Eds.), *Human nature and natural knowledge* (pp. 259–291). Dordrecht: Springer.

Zahle, J., & Kaidesoja, T. (2019). Emergence in the social sciences. In S. Gibb, R. Findlay, & T. Lancaster (Eds.), *The Routledge handbook of emergence* (pp. 400–409). London/New York: Routledge.