



**HAL**  
open science

# Human joint action ontogeny: how does the progression of mindreading and perspective-taking abilities shape capacities for joint action?

Sandra Lasry

## ► To cite this version:

Sandra Lasry. Human joint action ontogeny: how does the progression of mindreading and perspective-taking abilities shape capacities for joint action?. Cognitive Sciences. Université Paris sciences et lettres, 2021. English. NNT : 2021UPSLE010 . tel-03697094

**HAL Id: tel-03697094**

**<https://theses.hal.science/tel-03697094>**

Submitted on 16 Jun 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE DE DOCTORAT  
DE L'UNIVERSITÉ PSL**

Préparée à l'Ecole Normale Supérieure

**Human joint action ontogeny:  
How does the progression of mindreading and  
perspective-taking abilities shape capacities for joint  
action?**

Soutenue par

**Sandra LASRY**

Le 18 janvier 2021

Ecole doctorale n° 158

**Ecole Doctorale Cerveau  
Cognition Comportement**

Spécialité

**Sciences cognitives**

**Composition du jury :**

Frédérique, de VIGNEMONT Directrice de Recherche CNRS, ENS, PSL	<i>Présidente</i>
Cédric, PATERNOTTE Maître de conférence, Sorbonne Université	<i>Rapporteur</i>
Steve, BUTTERFILL Professor, University of Warwick	<i>Rapporteur</i>
Victoria, SOUTHGATE Professor, University of Copenhagen	<i>Examinatrice</i>
Cordula, VESPER Associate Professor, Aarhus University	<i>Examinatrice</i>
Elisabeth, PACHERIE Directrice de Recherche CNRS, ENS, PSL	<i>Directrice de thèse</i>



## **Originality Statement**

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at Ecole Normale Supérieure or any other educational institution, except where due acknowledgment is made in the form of bibliographical reference.

I also declare that the intellectual content of this thesis is the product of my own work, and the claims here reflect my own thinking. However, as I believe scientific research is inherently a social activity, and all work receives support from others in theoretical, methodological, or stylistic matters; I will present this work in first-person plural voice.

## **Abstract**

Many significant achievements of our species are the result of joint actions. Joint action requires that co-agents form shared representations of their environment, their goals and actions; by doing so, they integrate self and other representations within a unitary representation. The purpose of my dissertation is to investigate how a capacity for joint action and shared representations develops in human ontogeny.

In philosophy, maximalists argue that joint action requires co-agents to share intentions (using theory of mind capacities) when minimalists contend that joint action requires only that individuals share goals (understanding others' behaviors as goal-directed). I contend that joint actions come in various forms and make different representational and sharing demands, making room for a minimalist account of at least some types of joint action. In developmental psychology, early theory of mind abilities are still debated. I discuss the implications of this debate for young children's capacity to engage in joint action, concluding that young children may meet representational and sharing demands of minimal joint actions using goal representations and visual perspective-taking capacities (instead of intentions and theory of mind capacities).

In order to shed light on the ontogeny of joint action, I offer a joint action matrix combining: (1) the type of representation involved in joint action based on its complexity and (2) the level of sharedness implied by such representations based on the co-agents' strategy to maximize the joint performance. I argue that, before young children fully develop the mindreading and perspective-taking abilities required for more complex joint actions, they may already successfully engage in minimal forms of joint action – imposing less representational and sharing demands.

## **Acknowledgements**

I would like to thank my supervisor, Elisabeth Pacherie, for taking me as her student and leading me through the domain of philosophy of action through all these years. Thank you for your mentoring and support; it was really an honour working with you during my PhD. Furthermore, I want to thank the entire faculty and staff of the Department of Cognitive Science at the Institut Jean Nicod. Thanks to all the faculty and staff who make this department so welcoming and such a great intellectual hub. Nathalie Abitbol-Evin and Vincent Gaudefroy, thanks for your help and your kindness during these years. They are so many people that I enjoyed discussing philosophy with, so many smart researchers in this department; Alban Bouvier, Roberto Casati, Pierre Jacob, Uriah Kriegel, François Recanati and many others. Many thanks to the Agency team for our regular meetings and discussions – Valérien Chambon, Nura Sidarus – and to the BabyLab for their comments on my work – thanks to Anne Christophe and her team.

Throughout the years, I assisted and participated to research events, whereby I had the chance to meet and receive comments from eminent researchers. I would like to thank Stephen Butterfill, Corrado Sinigaglia, Victoria Southgate and Deborah Tollefsen. More specifically, I would like to thank Kiley Hamlin who welcomed me in her research center (Center for Infant Cognition) several times at the University of British Columbia. It has been a pleasure working in your research center, discussing with you about research and collaborating with exceptional research students who also became good friends; Allana Clague-Morai, Enda Tan and so many other students. Thank you for all the knowledge and experience that I had the opportunity to acquire with you.

I would like to thank my family, my friends and my partner for their unconditional love and support throughout all the years. Thank you for believing in me and encouraging me in my projects. Many thanks to my parents Gérard Lasry and Catherine Peyre; my sister Mathilde Lasry; my grand-parents Paul and Thérèse Peyre, Joseph and Jacqueline Lasry; my uncle and godmother, Jack and Christine

Vandevoorde and my cousins, Nicolas and Jessica Vandevoorde. I would like to thank my friends, Amélie Aubert-Noël, Nicolas Baudin, Louis-Henry Berthet-Garnier, Grayson Clague, Allana Clague-Morai, Antoine Fleury, Marie-Clémentine (Aaren) Germain, Anna Giustina, Jessica Higgins, Ekoé Kangni-Foli, Laura Labarthe, Slawa Loev, Tricia Magalotti, Méryl Marcer, Marlyse Pezzetta, Andrés Soria-Ruiz and Enda Tan. A special thanks to my partner, François Bastard, and his family for their support and interest in my work. Finally, thanks to my cat, Batman, who kept me company while I worked late at night.

I would like to dedicate my PhD thesis to my grandfather, Paul Peyre, who has always been extremely proud of me. I know that you are still watching over me, I love you.

# Table des matières

<b>GENERAL INTRODUCTION .....</b>	<b>11</b>
<b>PART I PHILOSOPHICAL ACCOUNTS OF JOINT ACTION: MAXIMALIST VS. MINIMALIST POSITIONS.....</b>	<b>17</b>
INTRODUCTION TO PART I .....	19
CHAPTER 1 MAXIMALISM OR ENCODING OTHERS IN MENTAL TERMS: THE SHARED INTENTION HYPOTHESIS ...	21
<i>1.1 Sharing intentions as sharing representational content .....</i>	<i>21</i>
1.1.1 Intentions as a sui generis type of mental state .....	21
1.1.2 Shared cooperative activities' requirements.....	23
1.1.3 Participating in shared cooperative activities (SCA): intentions' content, interrelatedness and functioning.....	25
1.1.4 From individuals to groups: planning norms and self-governance.....	27
<i>1.2 Sharing intentions as sharing a specific representational mode .....</i>	<i>29</i>
1.2.1 Joint actions' requirements: we-intentions as "slices" of joint intentions .....	30
1.2.2 Specific modes: "I-mode", "pro-group I-mode" and "we-mode" attitudes .....	33
<i>1.3 Sharing intentions as becoming a plural subject.....</i>	<i>36</i>
1.3.1 Joint actions' requirement: jointly commit as a plural subject .....	37
1.3.2 Joint commitments: mutual obligations, expectations and rights .....	38
CHAPTER 2 MOTIVATING MINIMALISM .....	43
<i>2.1 Motivating minimalism: the cognitive prism .....</i>	<i>44</i>
2.1.1 Young children's social competencies in their first three years of life.....	44
2.1.2 The study of joint action in developmental psychology .....	49
2.1.2.1 Naturalistic games .....	50
2.1.2.2 Joint go/no-go tasks.....	51
2.1.2.3 Puzzle-solving tasks .....	54
<i>2.2 Motivating Minimalism: the contextual prism.....</i>	<i>57</i>
2.2.1 Motor-related contextual cues.....	57
2.2.1.1 Minimal joint actions: concreteness and spontaneity .....	57
2.2.1.2 Coordinating in live interactions .....	58
2.2.2 Emotion-related contextual cues .....	59

<b>PART II DEVELOPMENTAL PSYCHOLOGY CHILDREN’S MINDREADING AND PERSPECTIVE-TAKING CAPACITIES.....</b>	<b>63</b>
INTRODUCTION TO PART II.....	65
CHAPTER 3 PROPOSITION 1: SHARED COOPERATIVE ACTIVITIES REQUIRE SOPHISTICATED COGNITIVE ABILITIES TO REPRESENT AND REASON ABOUT MENTAL STATES.....	71
3.1 <i>Eliminating the mentalistic stance: reasoning about goals</i> .....	71
3.1.1 Behavioral rules.....	73
3.1.2 Mutual responsiveness and common knowledge.....	74
3.2 <i>The existence of a minimal theory of mind?</i> .....	75
3.2.1 Implicit and explicit systems: from minimal to robust theories of mind.....	75
3.2.2 Signature limits, goal-tracking and motor system.....	80
3.2.3 A minimal theory of mind in joint action?.....	83
CHAPTER 4 PROPOSITION 2: YOUNG CHILDREN DO NOT HAVE THE SOPHISTICATED COGNITIVE ABILITIES NEEDED TO ENGAGE IN SHARED COOPERATIVE ABILITIES.....	87
4.1 <i>A mentalistic approach</i> .....	87
4.1.1 Where did the mentalistic approach come from?.....	88
4.1.2 Mentalistic approach as a one-account system.....	90
4.2 <i>A success gap between implicit and explicit measures</i> .....	91
4.3 <i>Challenging the mentalistic approach?</i> .....	94
CHAPTER 5 PROPOSITION 3: CHILDREN ENGAGE IN SHARED COOPERATIVE ACTIVITIES.....	98
5.1 <i>Would ethology shed new light on our problem?</i> .....	99
5.1.1 Burge’s action-sensing attribution scheme.....	99
5.1.2 Are false-belief tasks good markers of belief attribution?.....	100
5.1.3 Collective behaviors in social animals: a case of joint action?.....	104
5.2 <i>Are infants subject to altercentrism?</i> .....	106
5.2.1 Tracking the other’s perspective: switching point of view in false-belief tasks.....	106
5.2.2 The altercentric bias hypothesis favors goal over mental state ascription.....	107
5.2.3 The altercentric bias hypothesis implications for joint action.....	109
<b>PART III A JOINT ACTION MATRIX.....</b>	<b>117</b>
INTRODUCTION TO PART III.....	119
CHAPTER 6 DOES JOINT ACTION’S COMPLEXITY IMPACT THE TYPE OF SHARED REPRESENTATIONS?.....	122

6.1 Tracking and representing goals.....	122
6.1.1 A simple typology of goals .....	123
6.1.2 Goal ascription using motor representations .....	124
6.1.3 Goal ascription using learned statistical correlations between actions and outcomes .....	127
6.1.4 Goal ascription using teleological reasoning.....	129
6.2 Goal tracking in young children: what processes? .....	132
6.2.1 The Motor Theory of Goal Tracking (Butterfill, 2019).....	132
6.2.2 The Simple View (Gergely & Csibra, 2003).....	134
6.2.2.1 Means-goals associations .....	135
6.2.2.2 Teleological reasoning .....	138
6.3 Reasoning about goals in joint action .....	141
6.3.1 Tracking goals in joint action.....	142
6.3.2 Sharing motor representations in joint action.....	143
6.3.3 Categorizing joint actions .....	145
6.3.3.1 Joint action level of complexity: abstractness and advance planning .....	145
6.3.3.2 Joint action level of complexity modulates the nature of the shared representations .....	146
<b>CHAPTER 7 DOES CO-AGENTS' STRATEGY TO MAXIMIZE THEIR PERFORMANCE IMPACT THE SHAREDNESS CHARACTERISTICS OF SHARED REPRESENTATIONS? .....</b>	<b>151</b>
7.1 The co-agents' self- and other-representations in joint action.....	153
7.1.1 Identical and complementary forms of interaction.....	155
7.1.2 Co-agents' performance maximization strategies in joint action .....	158
7.2 The integration of self and other representations in childhood.....	161
7.2.1 Visual Perspective-taking abilities.....	161
7.2.1.1 Level-1 visual perspective-taking abilities .....	162
7.2.1.2 Level-2 perspective-taking abilities .....	165
7.2.2 Forms of interaction and strategies in childhood .....	168
<b>CONCLUSION.....</b>	<b>177</b>
<b>APPENDIX 1 THE JOINT SIMON EFFECT .....</b>	<b>181</b>
<b>APPENDIX 2 KEMPS CARD GAME .....</b>	<b>186</b>
<b>BIBLIOGRAPHY .....</b>	<b>189</b>



## *General Introduction*

In everyday life, we do plenty of things together: we take care of a child, move a heavy object or play team games together. Indeed, it is common for individuals to share goals and act jointly in order to achieve these shared goals. As a first pass, joint action can be defined as a social interaction in which individuals coordinate their actions in time and space in order to create a change in the environment (Sebanz, Bekkering, & Knoblich, 2006).

In recent years, leading philosophical characterizations of joint actions in terms of shared intentions (e.g., Bratman, 2014; Gilbert, 2014; Tuomela, 2007) have come under criticism as ill-suited to account for the joint action abilities exhibited by young children. While diverging on many points, these philosophical accounts of joint action all appear to require that participants in joint action have robust theory of mind skills and high-level reasoning abilities, since in order to have a shared intention they must be able to represent their partners' intentions and other attitudes relevant to the joint activity and reason about these mental states in order to adjust their own intentions and actions to theirs. Leading philosophical accounts of joint action can thus be characterized as cognitively maximalist. Cognitive maximalism nevertheless differs on many points. They may hold that what distinguishes shared from individual intentions is a matter of representational content (Bratman's intention condition), of representational mode (Tuomela's I-mode and we-mode we-intentions) or of subject (Gilbert's notion of plural subject). From maximalist accounts emerges a set of requirements on joint actions, including mutual responsiveness, common knowledge, interdependence, (mutual) belief and (joint) commitment. For agents to form such shared intentions, they need to appeal to high-level cognitive abilities in order to encode and incorporate their co-agent(s) in mental terms. The imposition of such strong cognitive requirements appears to imply that animals and small children, who either lack altogether or have not yet fully developed mentalizing and practical reasoning abilities, cannot form shared intentions and thus cannot engage in truly joint action (Pacherie 2011, 2013; Tollefsen, 2005).

The analysis of developmental psychology literature reveals abundant evidence that young children already engage in some forms of shared cooperative activities before the age of 4. The development of children's socio-cognitive abilities in their first three years of life is marked by three main milestones. During their first year, children develop a capacity for joint attention and at the close of it that can engage in triadic joint attention and by extension, achieve new competencies in communication and interaction (Carpenter, Nagell, & Tomasello 1998; Tomasello, Moore & Dunham 1995, Siposova & Carpenter 2020). In particular, at the start of their second year they start acting together with their partner on objects or events external to the dyad and progressively gain in autonomy, becoming more active partners, able towards the end of their second year to engage in novel goal-directed joint action, across contexts and partners. These advances in social understanding and interaction skills support a remarkable growth in joint action during the children's second year. Their engagement in joint action becomes progressively less dependent on adult guidance and structuring, up to the point, at the close of the second year, where they become able to contribute autonomously to novel goal-directed joint action, across contexts and partners (Brownell, 2011). Several researchers have proposed that this second milestone (autonomy) goes hand in hand with the development of an objective self and an ability to reflect on self in relation with others, which emerges in the second half of the second year (Barresi & Moore 1996; Kagan 2013; Perner, 1991; Brownell et al., 2006). Finally, a third milestone is achieved by the end of the third year, with the emergence of an understanding of the normative dimension of cooperation, the commitments associated with joint action and the concomitant obligations and entitlements of collaborative partners (Hamann et al. 2012; Tomasello & Hamann 2012).

Some researchers try and offer characterizations of joint action less cognitively demanding than those offered by maximalist philosophical accounts and thus compatible with the cognitive and social skills of young children. This has sparked the so-called minimalist program<sup>1</sup>. Minimalist accounts of joint action can take more or less radical forms, either retaining the notion of shared intention as

---

<sup>1</sup> Note that the type of minimalism we are concerned with here is cognitive minimalism in contrast to other kinds of minimalism (e.g., conceptual or ontological). For a discussion of these different kinds of minimalism and of the relations between minimalism and pluralism, see Paternotte (2020).

central to joint action but trying to offer potentially less demanding accounts of what counts as a shared intention (Tollefsen, 2005; Blomberg, 2011; Pacherie, 2011, 2013) or doing away with the notion of shared intention altogether and replacing it with that of a shared or collective goal (Butterfill, 2012, 2017). Minimalist accounts can also in principle be more or less radical in a different way, either embracing pluralism or monism. Monism is the view that all joint actions share a common core and that the proposed minimalist account captures that common core. More often though, minimalists are pluralists, accepting that there are different forms of joint actions, some more sophisticated and involving shared intentions in a strong sense, others more basic and less cognitively demanding. From this pluralist perspective, one important challenge is to understand how the capacities needed for sophisticated joint action relate to the capacities underpinning basic joint actions and what by what developmental steps one becomes capable of engaging in more sophisticated joint actions (see Butterfill & Pacherie, 2020 for an attempt to tackle this challenge). Minimalists are especially interested in joint actions relying on minimal cognitive skills or competencies (cognitive prism) and/or on external features such as the context or the behavioral interaction in itself (contextual prism).

The common ground on which maximalism and minimalism agree is that joint actions require one to be able to track efficiently others' actions and to plan and executes one's own actions towards a common goal in light of what others are doing (Hassin et al., 2005). However, the confrontation of these two main accounts paves the way for a reevaluation of how these main requirements can be met. In particular how they are met may depend on two features of the joint action situation: the demands of the task itself and the degree of uncertainty regarding the information available about the partner (Chackochan & Sanguinetti, 2019).

In Part I, I will introduce the crucial question of the theory of mind abilities needed for participation in joint action via a presentation of the maximalist approach and its variants (Chapter 1). Maximalism implies that joint actions require the use of sophisticated high-level cognitive abilities. I will then motivate a minimalist account of joint action relying on evidence from human ontogeny (Chapter 2). Minimalism contends that some joint actions can rely on proxies for such fully developed cognitive abilities (cognitive minimalism) and on contextual elements (contextual minimalism). In

other words, there are some joint actions which do not require understanding others in mental terms (ascription of intentions) but simply understanding their actions as goal-directed (ascription of goals).

In Part II, I will examine whether observational and experimental data about children are compatible with maximalist accounts of joint action, i.e., I will assess whether the findings from developmental psychology studies about young children's cognitive abilities and their engagement in joint action are consistent with the cognitive requirements – high-level cognitive skills – stemming from maximalist accounts. To do so, I will introduce and discuss a trilemma that captures the gist of the debate between maximalism and minimalism regarding children's ability to engage in joint action (Rakoczy, 2018): (1) Shared cooperative activities require sophisticated cognitive abilities to represent, and reason about, mental states; (2) Young children do not yet have these sophisticated cognitive abilities; (3) They engage in shared cooperative activities. I will discuss each proposition independently, successively evaluating the main positions relatively to young children's theory of mind abilities. Among these positions: (i) nativists, for whom infants have an innate ability to represent complex propositional mental states (Baillargeon et al., 2015); (ii) empiricists, for whom infants use simple teleological representations to understand agentic action (Csibra & Gergely, 2013); and (iii) dual theorists, who argue for the existence of two different systems, an implicit system for tracking belief-like relational states and a later-developing, explicit system supporting the ascription of complex propositional mental states (Butterfill & Apperly, 2012). Consequently, to the question of whether observations in developmental psychology are compatible with the maximalist account of joint actions, the nativist account yields a positive answer and the other two accounts a negative answer. In Chapter 3, I will examine the nativist account which accepts the trilemma's propositions (1) and (3) but rejects the proposition (2). In Chapter 4, I will discuss empiricists and dual theorists accounts which accept the trilemma's propositions (2) and (3) but reject the proposition (1). Finally, in Chapter 5, I will consider the questions whether the activities children engage in with others (and presenting the appearance of shared cooperative activities) could be explained without involving representational abilities or shared representations (Burge, 2018) or can rely on an 'altercentric bias', where the altercentric bias hypothesis holds that young children do not maintain both their own and the other perspectives but rather take the other perspective (Southgate, 2020). Both approaches of this

question dispose of the “sharedness” characteristic of the problem by relying either on an action-sensing attribution scheme or on the elimination of the self-perspective.

In Part III, I will develop a joint action matrix organized around two axes. In Chapter 6, I will focus on the nature of the shared representations deployed in joint action, i.e., the type of representation of their co-agent an agent needs to deploy based on the joint action complexity (representation of their goals vs. representation of their mental states). I will discuss the processes at stake in the encoding of others’ actions in terms of goals, namely the ability to track the goals of others’ actions using either motor simulation (Sinigaglia & Butterfill 2016; Butterfill, 2019), teleological reasoning (Gergely & Csibra 2003; Csibra & Gergely 2007, 2013), learned statistical correlations between actions and outcomes (Elsner & Aschersleben 2003; Leslie 1984) or a combination of these processes, as well as the kind of goals that can be so encoded. I will then propose a categorization of collective goals’ complexity based on two criteria – abstractness and demands for advance planning – and distinguish three main levels of complexity of joint actions. My first axis exploits the idea that the complexity of joint actions determines the nature of the representations shared by co-agents. More specifically, joint actions of low and intermediate complexity only require co-agents to share representations of goals (via motor representations and/or inferential processes) whereas highly complex joint actions require the representation of mental states.

In Chapter 7, I will focus on the level of sharedness involved based on the strategies used by co-agents to maximize their performance (identical or complementary actions). The aim of this chapter is to develop the second axis of the joint action matrix, i.e., the adoption of strategies by co-agents as a way of maximizing co-agents’ performance is relative to perspective-taking abilities. In the case of identical strategies, co-representation effects may facilitate the joint action performance while in the case of complementary strategies, co-agents need to avoid interference created by co-representation effects to maximize their performance – these two types of cases require different levels of perspective-taking abilities (e.g., Moll & Meltzoff, 2011). The level of sharedness depends on the integration of self and other representations in a unitary representation, i.e., as a dual-person (dyadic) plan (for a recent review, see Sacheli, Arcangeli & Paulesu, 2018; della Gatta, Garbarini, Rabuffetti, Viganò, Butterfill and Sinigaglia, 2018). In the case of identical strategies, co-agents would share their

partner's perspective without the need for maintaining a competing representation of the event from their own perspective. In the case of complementary strategies, co-agents would need to maintain or confront their own perspective with their co-agent's perspective in order to complement their partner's action. In human ontogeny, the development of perspective-taking abilities has been linked to changes in the relative representational strength of the encoding of events from a self-perspective and from others' perspectives in relation to the progressive emergence of a self-concept and to the development of executive functions. In particular, the emergence of a self-reflective system has been argued to allow children "to represent explicitly the causal, temporal, and spatial relations between the respective actions of self and a partner...to take the partner's goal related activity into account in concert with their own and to adjust their own behavior by monitoring, timing, and sequencing their behavior together with adults and peers in relation to a common goal" (Brownell, 2011: 2006).

The originality of my thesis resides in the joint action taxonomy I develop to shed light on the ontogeny of joint action. This joint action taxonomy is based on the matrix that will be presented in Chapters 6 & 7. I propose that instead of simply asking the binary question of whether young children are able to participate in joint action or not, one should take into account the fact that joint actions vary in complexity and can impose more or less stringent representational and sharing demands. More specifically, I argue that, before young children fully develop mindreading and perspective-taking abilities, they may successfully engage in minimal forms of joint action that impose less stringent representational and sharing demands.

*Part I*

*Philosophical Accounts of Joint Action:*

*Maximalist vs. Minimalist Positions*



## *Introduction to Part I*

Philosophers agree that in order to coordinate successfully, agents must be able to form shared representations of their environment, their goals and actions. Shared representations are considered as the main element that distinguishes joint actions from parallel or individual actions in which the movements of agents happen to be synchronized (e.g., walking next to each other, waiting in a traffic jam). There is no consensus, however, about the nature of the shared representations required to sustain joint action (e.g., goal representations or representations of mental states) or on what it means exactly to share a representation (e.g., individual representation alignment, creation of a common representation or a softening of subjects' boundaries). While performing joint actions, agents need to take into consideration what their partners do. However, does it mean that agents need to encode their partners in mental terms?

In Part I, I will present two main positions regarding this question. In chapter 1, I will present the maximalist position (e.g., Bratman, 2014; Gilbert, 2014; Tuomela, 2007) which defends that joint action requires co-agents to share intentions – encode their partner in mental terms – and that this requires theory of mind capacities, since agents should be able to attribute mental states (i.e., intentions) to themselves and their partners. In chapter 2, I will motivate an alternative account, namely the minimalist position (e.g., Pacherie 2011, 2013; Tollefsen, 2005) which contends that joint action requires only that individuals share goals – encode their partner in goal terms – and understand other's behavior as goal-directed. Minimalists are concerned by the existence of joint actions relying on minimal cognitive skills or competencies (cognitive minimalism, e.g., Tollefsen, 2005; Butterfill & Pacherie, 2020; Vesper, 2016; Paternotte, 2020) and/or on external features such as the context or the behavioral interaction in itself (contextual minimalism<sup>2</sup>, Vesper, Butterfill, Knoblich & Sebanz, 2010; Tollefsen, 2005).

---

<sup>2</sup> Note that what I call “contextual minimalism” does not correspond to what Paternotte (2020) calls “conceptual minimalism”. Whereas Paternotte refers to situations where contextual factors that could help coordination are reduced to a minimum, whether I use it in to refer to favorable situations where coordination is easy and

---

contextual cues favor cooperative success. Thus, while “contextual minimalism” in Paternotte’s sense is presumably not compatible with cognitive minimalism, “contextual minimalism” in my sense is.

# Chapter 1

## *Maximalism or encoding others in mental terms:*

### *the shared intention hypothesis*

Alonso claims that the specificity of joint action lies in an “internal component...the participants’ having a ‘collective’ or ‘shared’ intention to so act.” (2009: 444–445). Maximalists argue that joint action requires co-agents to share intentions and that this in turn demands theory of mind capacities, since agents should be able to attribute mental states (i.e., intentions, beliefs) to themselves and their partners (for a review, see Butterfill & Sebanz, 2011). Compared to individual or parallel actions, joint actions require sharing intentions (*shared intention hypothesis*).

This chapter aims at giving an overview of the maximalist approach and the shared intention hypothesis, i.e., the hypothesis according to which joint actions require us to encode others in mental terms (using representations of mental states). Within this maximalist approach, different options have been offered regarding how shared intentions differ from individual intentions: the content, mode and subject hypotheses.

#### **1.1 Sharing intentions as sharing representational content**

Bratman (1999, 2006, 2009, 2014) argues that what makes shared intentions special has to do with their contents and the way they are interrelated. Shared intentions are understood as interpersonal structures of interrelated intentions that are used to coordinate and plan actions and as structures around which agents need to adjust.

##### *1.1.1 Intentions as a sui generis type of mental state*

Intentions, according to Bratman are a *sui generis* type of mental state – irreducible to beliefs and desires – distinctive in the fact that an intention “involves a characteristic kind of commitment to act” (Bratman, 1987: 15) and that its role is to control the agent’s conduct by guiding and coordinating her

planning. This kind of commitment to act which characterizes intentions incorporates two closely-related dimensions: volitional and reasoning-centered.

Bratman presents the example of someone who decides in the morning to go to a basketball game later in the day. The intention to go to the basketball game will play its volitional role only if it first plays its role in reasoning, i.e., if the intention is retained, the agent will need to figure out how to get tickets for the game and how to get to the stadium and will not schedule other activities at the same time. As Bratman explains (1987: 16), the volitional dimension of intentions relates to their role as “conduct-*controlling* pro-attitudes”: once an agent has formed an intention and a plan, when the time of action arrives, he or she will normally proceed to act as intended. The “*reasoning-centered* dimension of commitment” such as demands for consistency and means-end coherence as we see in the basketball game example, relates to the role intentions play between their initial formation and their initiation. The specific dispositions associated with these two dimensions of commitments make intentions irreducible to beliefs and desires.

The demand of consistency and means-end coherence which constitutes the specificity of intentions is involved in what Bratman calls the “planning agent’s self-governance” (Bratman, 2007, 2018). The planning agent’s self-governance entails that the agent’s practical thinking is guided by norms of plan rationality. These planning norms track consistency and means-end coherence so that they unify the agent’s standpoint which can play his role in self-governance and provide supplementary support to the norms themselves (their application to specific cases consolidates the use of such norms through the agent’s reflectivity stance on his/her actions). Bratman distinguishes synchronic from diachronic planning norms which are involved in the agent’s self-governance respectively at a given time and over time (through interconnections between forms of synchronic self-governance). In the case of diachronic self-governance, the interrelations between forms of synchronic self-governance include “characteristic forms of continuity of intention, cross-references between intentions at different times, intended and actual mesh between sub-plans at different times, and interdependence between relevant intentions at times along the way” (Bratman, 2018). In other words, intentions and their interconnections play a crucial role in self-governance at the individual level. Bratman compares these intrapersonal interconnections involved in diachronic self-governance to the

interpersonal intention interconnections which characterize shared intention and shared intentional activity. What about shared intentions then? What makes shared intentions different from individual intentions? According to Bratman, what makes shared intentions different from individual intentions is not their type (e.g., mode) but their content<sup>3</sup>. How do these interpersonal intention interconnections function? How do planning norms apply to shared activities and groups?

### ***1.1.2 Shared cooperative activities' requirements***

Bratman's focus is "primarily on the shared intentional activities of small, adult groups in the absence of asymmetric authority relations within those groups, and in which the individuals who are participants remain constant over time" (2014: 7). He considers cases of shared cooperative activities (SCA) such as painting a house together, singing a duet, playing basketball, engaging in conversation, dancing tango, going for a walk or traveling together (Bratman, 1997: 50–51, 1999: 130, 2009a: 150). To perform such activities, individuals form a shared intention which "is not an attitude in the mind or minds of either or both participants. Rather, it is a state of affairs that consists primarily in attitudes (none of which are themselves the shared intention) of the participant and [intersubjective] interrelations between attitudes" (Bratman, 1993: 107-108).

Bratman highlights what he takes to be the three main characteristics of SCA (Bratman, 1999; 2006); a "trio of features" which needs to be present for the activities of two or more agents to qualify as SCA. In this "trio of features", both the attitudes of partners and their interrelations are key elements:

- (i) *Mutual responsiveness*: In SCA each participating agent attempts to be responsive to the intentions and actions of the other, knowing that the other is attempting to be similarly responsive. Each seeks to guide his behavior with an eye to the behavior of the other, knowing that the other seeks to do likewise.

---

<sup>3</sup> Bratman adopts what he calls the continuity thesis, i.e., "the conceptual, metaphysical, and normative structures central to such modest sociality are...continuous with structures of individual planning agency" (2014: 8).

(ii) *Commitment to the joint activity*: In SCA the participants each have an appropriate commitment (though perhaps for different reasons) to the joint activity, and their mutual responsiveness is in the pursuit of this commitment.

(iii) *Commitment to mutual support*: In SCA each agent is committed to supporting the efforts of the other to play her role in the joint activity. If I believe that you need my help to find your note (or your paint brush) I am prepared to provide such help; and you are similarly prepared to support me in my role. These commitments to support each other put us in a position to perform the joint activity successfully even if we each need help in certain ways.

(Bratman, 1992: 328)

SCA are a type of activities “...suitably explainable by a shared intention” (Bratman, 1999: 9). More specifically, SCA require agents to understand the intentions of their co-agent in order to achieve a shared goal, to be committed to the joint activity, to have common knowledge<sup>4</sup> that this is the case and finally to be committed to supporting their co-agent. In other words, this trio of features involves a cognitively demanding bidirectional understanding of mental states and commitments. Let’s imagine two individuals, Grayson and Allana, who decide to go for a hike such as the Garibaldi lake trail in British Columbia where both need to walk about ten kilometers (elevation: 1000 meters) to reach Garibaldi lake. These two individuals need to show (i) mutual responsiveness such as adjusting their speed based on their partner or not bumping into each other, (ii) commitment to the joint activity (possibly for different reasons, as Allana could be willing to relax at the top of the trail when Grayson intends to ask Allana to marry him in front of the beautiful lake), (iii) commitment to mutual support such as giving water to their partner if he/she needs some or giving him/her a hand when the path is difficult.

---

<sup>4</sup> Common knowledge is a phenomenon classically defined, for instance by David Lewis (1969), as a hierarchy of ‘A knows that B knows that ... knows that x’ propositions: “Let us say that it is *common knowledge* in a population P that \_\_\_\_ if and only if some state of affairs A holds such that: 1. Everyone in P has reason to believe that A holds. 2. A indicates to everyone in P that everyone in P has reason to believe that A holds. 3. A indicates to everyone in P that \_\_\_\_” (Lewis, 1969: 56). However, some other researchers do not describe common knowledge in these terms (e.g., Tollefsen, 2005).

Bratman's core "claim...is that shared intention consists primarily of attitudes of individuals and their interrelations" (Bratman, 1999: 129). In fact, agents need to satisfy these conditions mentioned above ("trio of features") with intentions which have a specific content and that are interrelated in a certain way.

### ***1.1.3 Participating in shared cooperative activities (SCA): intentions' content, interrelatedness and functioning***

In order to specify how agents can share intentions and thus participate in SCA, Bratman lists a series of conditions that he claims are together sufficient for agents to share intentions (see Bratman, 1999: 121; 2014: 103). According to Bratman's compressed version of these conditions, we share an intention to *J* if:

**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* (so there is interlocking and reflexivity) 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 favor 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 favor of our *J*-ing.

**C. *Interdependence condition:*** There is interdependence in persistence of the intentions of each in favor 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 the intentions of each in favor of that joint activity.

(Bratman, 2014: 103)

These five conditions provide a model of the “social glue” that ties together the participants (Bratman, 2014: 87). The first condition (A) is a complex intention condition concerning the content of the intention formed by each agent in a SCA and the intentions’ interrelatedness articulated through forms of intentional interconnection and interpersonal support<sup>5</sup>. The second condition (B) specifies the beliefs about success and interdependence the participants have<sup>6</sup>. The third condition (C) ensures the accuracy of the participants’ view concerning their actual interdependence<sup>7</sup>. The fourth condition (D) concerns the common knowledge of the conditions above. Finally, the fifth condition (E) specifies what counts as relevant mutual responsiveness in intentions and beliefs<sup>8</sup>.

Let me illustrate these conditions with an example. Two agents Paul and Thérèse intend to do the dishes together. Condition (A) demands that Paul and Thérèse each intend that they wash the dishes, that they do so by way of each of their intentions and of meshing subplans and actions. Condition (A) would not be satisfied if, for instance, Thérèse intended that they wash the dishes together by forcing Paul to help her against his will or if Thérèse and Paul each insisted that they themselves wash the dishes and the other dry them. Saying that their subplans must mesh us not to say that they must be identical, only that they should not conflict. For instance, Thérèse might plan to use a specific dishwashing detergent but not care whether they wash the plates or the glasses first, while Paul might not care about the dishwashing liquid used but care about the order in which the dishes are washed<sup>9</sup>. Condition (B) demands that both Paul and Thérèse individually believe that they will wash

---

<sup>5</sup> Intentional interconnection and interpersonal support are specified by “(i) intentions on the part of each in favor of the joint activity, (ii) intentions on the part of each in favor of the joint activity by way of the intentions of each in (i) and by way of relevant mutual responsiveness in sub-intention and action, (iii) intentions on the part of each in favor of the joint activity by way of meshing sub-plans of the intentions of each in (i)” (Bratman, 2014: 85).

<sup>6</sup> More specifically, “(iv) beliefs of each that, if the intentions of each in (i) persist, the participants will perform the joint activity by way of those intentions and relevant mutual responsiveness in sub-intention and action, (v) beliefs of each that the intentions of each in (i) are persistence interdependent” (Bratman, 2014: 85-86).

<sup>7</sup> “(vi) the intentions of each in (i) are persistence interdependent” (Bratman, 2014: 86).

<sup>8</sup> “(viii) the connection between the shared intention (as in (i)–(vii)) and the joint action involves public mutual responsiveness in sub-intention and action that tracks the end intended by each of the joint activity by way of the intentions of each (in (i)) in favor of that joint activity. And (viii) specifies what counts as relevant mutual responsiveness in (ii) and (iv)” (Bratman, 2014: 86).

<sup>9</sup> Pre-existing mesh in sub-plans are not necessary, Paul and Thérèse just need to have “plan-like commitment[s] to there being such mesh, where achieving that mesh may require relevant bargaining or shared deliberation” (Bratman, 2018: chapter 2).

the dishes in virtue of the persistence of their individual intentions that they wash the dishes and relevant mutual responsiveness in sub-plan and action. Paul persists in intending that they wash the dishes because Paul believes that Thérèse persists in intending that they wash the dishes, and vice versa for Thérèse. Both agents' persistence in intending that they wash the dishes *depends on* the fact that they *believe* in their co-agent's persistence in intending that they wash the dishes. Condition (C) stipulates indeed that Paul and Thérèse are interdependent in their intentions' persistence. Condition (D) imposes a demand that all those conditions are common knowledge between Paul and Thérèse. Condition (E) requires that Paul and Thérèse show a *public* mutual responsiveness in sub-intentions and actions by observing each other. For instance, Paul washes when Thérèse dries the dishes; both Paul and Thérèse contribute to the joint activity – one washing and the other drying the dishes – in a way that, through the *public* mutual responsiveness, each “tracks the end” of the joint activity from her own role. To complete the joint activity, Paul washes the dishes (he starts with dirty dishes and finishes with clean wet dishes) and then Thérèse dries the dishes (she starts with clean wet dishes and finishes with clean dry dishes). In other words, each has his own role and can track the end of the joint activity through the performance of her intention in favor of the joint activity; the public mutual responsiveness Paul and Thérèse show cementing their respective roles as a block to perform the joint activity together.

To sum up, according to Bratman (2018), three types of elements are crucial for individuals to form shared intentions and thus to perform SCA, namely the *content of the intention* formed by each agent and the *interrelatedness of co-agents' intentions* (intention condition); the coherence of the co-agents' interrelated intentions built onto the content of those intentions (belief and common knowledge conditions); and *the way in which they interact in their functioning* which does not concern the content of intentions but rather a tracking of the intended SCA (interdependence and mutual responsiveness conditions).

#### ***1.1.4 From individuals to groups: planning norms and self-governance***

Bratman's position is a form of reductionism to the extent that it analyses shared intentions in terms of individual's mental states and their interrelations, using conceptual, metaphysical and normative

resources already involved in an analysis of individual planning agency. However, since his account conceives of a shared intention as the complex of attitudes of individuals and their interrelations, it means that individuals on their own cannot have a shared intention. In addition, Bratman's individualistic theory does not accept the group as a primitive notion (2014: 123-124), meaning that group agents are viewed as groups only in terms of "the relevant underlying structure of interrelated individuals" (2014: 123) and that the idea of a group as a causal agent does not need to appeal to a form of causation distinct from the causal processes that are shaped by that underlying structure.

How do the planning norms governing shared intentions function then? Roles and norms of individual self-governance such as "coordinating, structuring, organizing, guiding, and settling roles" and "norms of consistency, agglomeration, means-end coherence, and stability" (Bratman, 2014: 15, 27) give rise to roles and norms of modest sociality. Modest sociality includes roles of "social coordination and organization in relevant thought and action" and "norms of social agglomeration and consistency, social coherence, and social stability" (Bratman, 2014: 27). In other words, according to Bratman, social norms are derived from individual norms in the case of modest sociality<sup>10</sup>. In line with his planning theory of human agency, Bratman argues that planning figures centrally and structurally into the way that humans govern themselves<sup>11</sup> (2007; 2018). However, he does not rule out the possibility of a "thin kind of planning agency in the absence of a concern with self-governance" and considers that "this is how we should think about the planning of young human children, or the great apes, or certain artificial intelligence systems" (2018: 16). Indeed, Bratman's account has also been used to make sense of children's joint activities such as social pretend play, or two children building a block tower together (Carpenter, 2009; Tomasello, Carpenter, Call, Behne, & Moll, 2005; Tomasello & Hamann, 2012). Even though Bratman's focus is on the shared intentional activities of "adult humans in a broadly modern world" (Bratman, 2009: 153) aiming to account for most "robust" cases

---

<sup>10</sup> Bratman and Searle's accounts were accused of not capturing the social normative relations present in collective intentions (Gilbert, 1998) even though Bratman defined strongly shared intentions (as opposed to weakly shared intentions) that involve binding agreement (see Tollefsen, 2004). In his 2018 book, Bratman offers an in-depth discussion of these normative issues and especially of norms of rationality.

<sup>11</sup> "...by studying the ways in which basic norms of plan rationality reflect conditions of a planning agent's self-governance, both at a time [synchronically] and over time [diachronically]" (Bratman, 2018: 4).

of joint activity, some researchers think that even 1-year-old infants satisfy Bratman's conditions on shared intentional activity (see Carpenter, 2009). Others offer instead a less-demanding characterization of intentional activity (see Tollefsen, 2005).

In a nutshell, Bratman defends a constructivist account of joint action in which co-agents share an intention when each has an intention of the form "I intend that we J" accompanied with a belief that their co-agent(s) intend(s) the same, persist(s) in her (their) intention(s) and is (are) mutually responsive during the joint action. Bratman's approach will be interesting to discuss for many reasons, especially the nature of shared representations involved in joint action. Indeed, could his account apply to young children, notably in the context of the debate on false-belief understanding (see Part II)? To what extent do alternative accounts such as less demanding accounts of the shared intention hypothesis (see Tollefsen, 2005), motor representations accounts (see for a recent review, Della Gatta et al., 2017) or accounts that deny the need for shared intentions (see Butterfill, 2010) challenge Bratman's account?

## **1.2 Sharing intentions as sharing a specific representational mode**

Other philosophers argue that what is distinctive of the intentions involved in joint action is their intentional *mode*, a "we-mode" (e.g., Tuomela, 2007; Searle, 1990, 2010). Let us consider members of a group performing a joint action together such as going to the Taj Mahal. According to the representational mode position, the joint action performance is possible in virtue of each agent having the we-intention to go to the Taj Mahal. However, John Searle and Raimo Tuomela defend different versions of the representational mode position. Searle argues that, when two individuals go to the Taj Mahal, each of them has the we-intention "We are going to the Taj Mahal" in his mind from which each of them derive his/her own contribution (as a participatory intention) and assume that the other will perform his/her contribution too (Searle, 2010: 52-53). Searle considers that 'we'-attitudes cannot be reduced to 'I'-attitudes (Searle 2010: 46) i.e., we-intentions are in the individuals' mind. In contrast, Tuomela and Miller (1988) argue that if each individual has an individual intention about his contribution ("doing his part") and certain beliefs about the other individuals involved, he/she can be said to 'we-intend' the joint action to go to the Taj Mahal. From Searle's perspective, Tuomela and

Miller are wrong in claiming that we-intentions can be reduced to I-intentions plus beliefs and their analysis is viciously circular insofar as “the ‘we-intention’ is built into the notion of doing his part” (Searle, 1990: 405). Tuomela (2005) responds to Searle’s criticism by pointing out that the analysis he and Miller proposed was not meant to be reductive but rather “meant to elucidate the irreducible notion of we-intention in a functionally informative way” (2005: 358).

### ***1.2.1 Joint actions’ requirements: we-intentions as “slices” of joint intentions***

In Tuomela’s account, ‘joint intentions’ (also called shared intentions) are attributed to a collective when each of its members has a we-intention. Tuomela argues that, in order to perform joint actions, agents need to form individually ‘we-intentions’:

A member  $A_i$  of a collective  $g$  we-intends to do  $X$  if and only if: (i)  $A_i$  intends to do his part of  $X$  (as his part of  $X$ ); (ii)  $A_i$  has a belief to the effect that the joint action opportunities for an intentional performance of  $X$  will obtain...(iii)  $A_i$  believes that there is (or will be) a mutual belief among the participating members of  $g$ ...to the effect that the joint action opportunities for an intentional performance of  $X$  will obtain...(iv) (i) in part because of (ii) and (iii).

(Tuomela, 2005: 340-341)

First, the agent forms the intention to play his part. For instance, a soccer teammate has the intention of scoring a goal if he/she intends to do her part in scoring a goal such as making a decisive pass. Second, the agent believes that others will play their parts. For instance, he/she believes that each of his/her teammates will play their part as a condition for success in scoring a goal. Third, the agent believes that there is (or will be) a mutual belief among the agents regarding the possibility of achieving the joint activity. For instance, he/she believes that there is a mutual belief among the soccer teammates that conditions for success obtain. Finally, he intends to do his part in scoring the goal at least in part because he believes others will also play their parts and believes that there is a mutual belief among team members that they can together score a goal.

Tuomela draws a distinction between *action-intentions* and *aim-intentions* (Tuomela 2002, 2005, 2007). Action-intentions are intentions that an agent can alone settle and satisfy (Tuomela, 2005) but we-intentions are a kind of aim-intentions, where for an agent to have an aim-intention “it is not required that the agent believes that he with some likelihood can alone bring about or see to it that the action or its result event comes about” (Tuomela, 2005: 329). In a joint action, agents jointly intend to perform a joint action together where the central condition of satisfaction of the we-intention is that each agent should intend to participate by his/her own action<sup>12</sup>. In other words, a we-intention is an aim-intention but the individual’s intention to do his/her part (condition i) is an action-intention.

How then, is reasoning at the group-level (‘group agent’s reasons’, in Tuomela, 2013: 39) connected with reasoning at the individual-level (‘group reasons simpliciter’, in Tuomela, 2013: 39)? Joint actions require joint intentions at the group-level which are formed thanks to the formation of we-intentions at the individual-level. Tuomela argues that agents get engaged in ‘we-mode we-reasoning’ (Tuomela, 2013: 39) and that there is a supervenience relation between joint intentions and we-intentions such that “every change in the group’s intention ...is necessarily accompanied by some change in the we-mode intentions of at least some members of the group” (Tuomela, 2013: 72). In effect, Tuomela defends the view that the intentional structures of joint actions consist in individual attitudes joined together. On the one hand, the notion of a “we-intention” is an irreducible notion but on the other hand, it is grounded in an “ontically individualistic or, better, interrelational” framework (Tuomela, 2005: 342). Although in this ‘we-mode’ approach, the group is viewed as an intentional agent, its members are the “primary ontological agents acting as representatives for the group” and the functional and intentional existence of the group is “extrinsic and basically derives from the joint attitudes, dispositions, and actions of its members, and from the irreducible reference to the group that these attitudes and actions involve and that is here assumed to make groups conceptually irreducible to members’ individual properties and relationships not based on the group” (Tuomela, 2013: 2-3).

---

<sup>12</sup> “A we-intention is not by itself an “action intention” but an “aim-intention” involving that the agent intentionally aims at X and is “aim-committed” to X, while his action commitment is to performing his part of X. The agent’s intention to perform his part of the joint action accordingly is a proper action intention, thus something the agent believes he can, at least with some probability, satisfy by his own action (given, of course, that the others perform their parts)” (Tuomela, 2005: 334).

Formulated differently: "...the intentional subject of a we-intention is 'we', while the ontological subject of a we-intention is a single agent" (Tuomela, 2007: 93). In other words, we-intentions are conceptually irreducible but group agents ontologically depend on their members psychological states and the agency of groups bottoms out in the behavior of their members. Tuomela here invokes the notion of an 'as if' intentional stance and argues that collective attitudes are based on group members' "collective attribution of attitudes to the group" (Tuomela, 2013: 23).

In a joint action context, agents form individually we-intentions which constitute "slices" of a collective or joint intention shared by the group. Tuomela also distinguishes four contexts where variations of normativity (commitment) and publicity (mutual belief) define different forms of joint actions:

(1) *Joint intention in a thick, normative context*, i.e., "(strongly) normatively group-binding on the basis of a joint obligation and collective commitment" (Tuomela, 2005: 346). In this context, joint intention could be based (a) on an agreement or (b) on expressed (normative) acceptance of the plan with conditions of mutual knowledge. For instance, (a) some individuals could make an agreement to paint a house tomorrow or (b) they could sign up for an activity that will happen only if enough people sign up (so that the joint intention is formed).

(2) *Joint intention in a weakly normative context*, i.e., "there are normative participation expectations based on an agent's leading the others to expect normatively that he will participate in the joint action in question" (Tuomela, 2005: 346). In this context, following from a verbal or physical expression of an individual to participate in a joint action, a kind of mutual weak promise is spread among other participants constituting a weakly normative joint intention. In this case, under condition of mutual belief, each participant is committed to the joint task. For instance, a group of individuals cook together and one says that he will help cleaning the kitchen on the next day. The rest of the group expects normatively that he will participate in cleaning the kitchen.

(3) *Joint intention in a non-normative context*; i.e., "there is a publicly shared plan-based joint intention which falls short of being even a weak promise in the sense of case (2)"

(Tuomela, 2005: 347). In this context, the intention is based on a non-normative, thin acceptance (e.g., I intend to participate but I do not promise that I will); participants are jointly committed to the plan due to mutual knowledge that they share the plan and by extension, a joint intention. For instance, a class of students is invited to participate in a charity sale in the next morning, the shared plan being to organize stalls with baked goods, fresh orange juice and other related items. There is a thin acceptance since each student is part of the class but there is no guarantee that all of them will participate.

(4) *Non-standard cases of joint intention not satisfying the publicity requirement.* In this case, there are no public participation expressions but a mutual belief concerning participation creates sufficient intersubjectivity to form joint intention and collective commitment. For instance, let's imagine a case where a woman forgot to take off one of her hair curlers before heading to a party. The hair curler is visible but none of the individuals that she meets and interacts with tells her about her hair curler. Each individual might believe that others did not tell her on purpose to play a trick on her; the mutual belief increases while the woman spends the night interacting with more individuals which results in the formation of a joint intention and collective commitment.

Performing a joint action requires the emergence of a joint intention shared by participants which consists in the formation of we-intentions at the individual level. We-intentions entails three conditions in participants: an intention condition to fulfill their role, a belief condition to the effect that the joint action opportunities for an intentional performance will obtain and a mutual belief condition i.e., a belief that there is or will be a mutual belief among participants to the effect that the joint action opportunities for an intentional performance will obtain. Tuomela also distinguishes different contexts (based on normativity and publicity) in which joint intentions can be formed with weaker or stronger commitment, mutual belief and mutual knowledge. Nevertheless, in a joint action, do participants reason as individuals or as group-members? How do these we-intentions function?

### ***1.2.2 Specific modes: “I-mode”, “pro-group I-mode” and “we-mode” attitudes***

As introduced previously, joint intentions require we-intentions at the individual level. Tuomela categorize different modes of thinking whereby individuals can form either I-mode we-intentions or we-mode we-intentions, where having an attitude in the I-mode means having it ‘as a private person’ and having it in the we-mode means having it ‘as a group member’ (cf. Tuomela 2007: 46). Thinking and acting in the we-mode basically amounts to thinking and acting for a group reason and in accordance with the group goals and, more generally, its ‘ethos’ (2007: 13). In other words, while I-mode thinking leads agents to act as individuals based on private reasons, we-mode thinking leads agents to act as group members based on the group perspective.

We-mode “we-intentions” exhibit three “criterial” features that “cement” group members together in all contexts where they function as group members:

...*group reason* (a unifying reason for group members to participate in group-based activities), a *collectivity condition* for all members (“necessarily being in the same boat”), and *collective commitment* (basically a product of joint intention and the members’ group reason involved).

(Tuomela, 2013: x)

Tuomela argues that members of the group collectively accept a proposition that they consider correct or true for the group in a way that involves “the implicit or explicit agreement to share a we-mode joint attitude” (Tuomela, 2013: 126). Thus, “joint intentions (and hence shared we-intentions) entail collective (or, here equivalently, joint) commitments to action, and this joint commitment also includes that the participants are socially committed to each other to perform their parts of the [joint action]” (Tuomela, 2005: 331). The idea that the agreement to share a we-mode joint attitude can be implicit suggests that the we-mode might be, in some contexts and under certain circumstances, triggered by automatic processes.

This opens up an interesting avenue that I will explore when discussing basic forms of joint actions especially in childhood. Indeed, on the one hand, some studies have shown that children start integrating normative factors, namely ‘joint commitments’ – that could be related to the agreement to share a we-mode joint attitude – by 3 years (Gräfenhain et al., 2009). On the other hand, such implicit

agreements could be resulting from motor alignment (e.g., synchronization, mutual coordination) or from the predominance of norms (e.g., conventions). In the case of adults, some studies have shown that mutual coordination strengthens the sense of joint agency in cooperative joint action (e.g., Bolt et al., 2016); this effect could also result in an implicit agreement even in children. Regarding norms, let's think of someone who just accidentally dropped her pile of documents on the floor and two people happened to be there. If they share the same social conventions, they might jointly bend down to help her gather up the documents. Helping behaviors have been observed in 12-14 month-olds (Liszkowski et al., 2008; Warneken & Tomasello, 2007), which might reveal that some norms are more basic than others or that children have a natural tendency to behave prosocially and might be interpreted as a sign that young children implicitly agree to share a we-mode joint attitude.

Tuomela adds another distinction between we-mode we-intentions and what he calls pro-group I-mode: "Comparing pro-group I-mode thinking with we-mode thinking, the crucial differences are, respectively, the change of agency from individual to collective (or group agency) and the change of I-reasoning to we-reasoning. These differences also account for the claimed irreducibility of the we-mode to the I-mode" (Tuomela, 2013: 7). Tuomela insists here on the 'we-reasoning' by making a clear distinction between 'we-reasoning' i.e., thinking as a group member, for the interest of the group and 'I-reasoning' which could be admittedly pro-group but not as a member of the group (as a form of altruism for instance). For instance, an agent can be altruistic, favoring the interest of a group without thinking of herself as a member of this group; this would be a pro-group I-mode reasoning. In the case of I-mode we-reasoning, the agent will think as a private person about her own interest; in the case of we-mode we-reasoning, the agent will think as a group member about the group's best interest; in the case of I-mode pro-group reasoning, the agent will think as a private person about the group's best interest. Suppose, for instance, that two automobilists are driving in opposite directions and a fallen tree blocks the road. They might form a joint intention to remove the tree together, but their we-intention will probably be I-mode we-intentions, since each driver is motivated by his or her individual interest, i.e., getting to his destination. Suppose instead that two automobilists are driving in the same direction to attend the same wedding and a fallen tree blocks the road. In this case, they might form we-mode we-intentions to remove the tree together, thinking of themselves as members of the same

wedding party. Finally suppose, that a pedestrian walking his dog sees these two people struggling to remove the tree and understands that they are wedding witnesses who can't be late for the ceremony, he might form a pro-group I-mode we-intention to remove the tree together with them. Each of these categories reveals that agents can have different reasons to participate in a joint action based on the mode they use to think as long as they intend to participate in the joint activity.

In a nutshell, on Tuomela's view, (1) joint actions require joint intentions at the group level formed thanks to the formation of we-intentions at the individual level, (2) individuals can reason in I-mode or we-mode based on whether they think individually or as members of the group as long as they intend to participate in the joint activity, (3) there is a supervenience relation between joint intention and we-intentions such as if there is any change in the joint intention, there must be some changes in the we-intentions of at least some members of the group. Tuomela positions himself as a non-reductionist collectivist and the specificity of his approach is twofold. First, rather than offering a content or a subject account of what distinguishes joint intentions from individual intentions, he claims that it is a matter of attitudinal mode, where we-intentions are irreducible to individual intentions. Second the 'we' in question can be more or less thin or thick depending on whether the agent has an I-mode or a we-mode we-intention, that is on whether the agent functions as a group member or not. Tuomela offers an interesting position which will raise the issue of the emergence of the 'we' and of the idea of group-membership in childhood: (i) what cognitive factors would be involved in their apparition? (ii) what can we say about the sense of joint agency in children? Finally, Tuomela presents the group as an intentional agent that ontologically relies on its members as representatives of the group, which interestingly contrasts with the next hypothesis I now turn to, namely the plural subject hypothesis.

### **1.3 Sharing intentions as becoming a plural subject**

Subject accounts of shared intentions claim that the specificity of shared intentions resides in their subjects. According to Margaret Gilbert's influential plural subject theory (Gilbert, 2000, 2009) the subjects of shared intentions are plural subjects. In her account of shared intentions, Gilbert

emphasizes the importance of joint commitment in creating plural subjects (see especially Gilbert, 2014).

### ***1.3.1 Joint actions' requirement: jointly commit as a plural subject***

According to Gilbert, shared intentions centrally involve joint commitments, where joint commitments should not be understood as mere concatenations of personal commitments. In Gilbert's conception, two or more agents are jointly committed to act together if they are "jointly committed to espouse a certain goal as a body" (Gilbert, 2014: 31) but more specifically if each "expresses [his] readiness to be jointly committed to espouse the relevant goal as a body, in conditions of common knowledge<sup>13</sup>" (Gilbert, 2014: 32). This commitment of will originates from individuals who need to express their readiness to be jointly committed in order to "create the whole [which happens] when and only when both expressions are in place [in the case of two individuals]" (Gilbert, 2014: 32).

Individuals' expressions of readiness often involve verbal communication but can also take more implicit forms. Gilbert writes that "the offering of these expressions, in conditions of common knowledge suffices for the joint commitment in question to be established" (Gilbert, 2014: 398). However, explicit agreement is neither necessary to establish a joint commitment nor constitutive of joint commitment in her sense (Gilbert, 2014: 26-28) since "joint commitments are not necessarily brought into being with any clear conscious intent to do so" (Gilbert, 2000: 6). For instance, two individuals might fall into a commitment to have dinner every Thursday after a faculty event; after a few occasions, even though the original intention is not present, the two individuals might feel related to each other, feeling committed as part of a group. In other words, the commitment is set up through a more or less explicit agreement which results from individuals' expressions of readiness. Among individuals, this agreement acts as a kind of normative "cement" that binds together individuals who enter into a joint commitment.

---

<sup>13</sup> Gilbert defines common knowledge between A and B that p: if A truly believes that p, B truly believes that p, A truly believes that B believes that p and so on (see Gilbert, 2014:51).

The expression “to intend as body”, aims at further specifying what is involved in being jointly committed to do something: “a joint commitment to intend as a body to do something is a joint commitment as far as possible to produce, by virtue of the actions of each, a single instance of intending to do that thing (Gilbert, 2009: 181).

### ***1.3.2 Joint commitments: mutual obligations, expectations and rights***

Joint commitments imply normative relationships between the parties to the commitments. Individuals who are jointly committed incur *mutual obligations, expectations and rights* such as reprimanding others for not acting as committed. For instance, “Anne and Ben are jointly committed to doing something as a body, each owes the other appropriate actions by virtue of their commitment...I shall say that a joint commitment *obligates* the parties, one to the other” (Gilbert, 1999a: 151).

Gilbert (2009) supports her claims regarding the centrality of joint commitments to shared intention and the normative structure of joint commitments, on the basis of observations concerning the way people talk and think about shared intentions in everyday settings. From these observations she derives three criteria of adequacy for accounts of shared intentions: the obligation criterion, the concurrence criterion and the disjunction criterion. She illustrates the first two criteria by means of the following example:

Queenie and Rom share an intention to do some shopping in a nearby town. In order to get there in time they must walk some miles along a dusty road at a certain pace. They are now half way along the road. Queenie’s pace begins to slow. In a tone of mild rebuke Rom says “Can you hurry up a bit? We won’t be able to get any shopping done at this rate!” Queenie says “Sorry!” and starts to move more quickly. Later she stops and for some reason announces: “That’s it! I’m not going any further!”<sup>21</sup> Rom is likely to be taken aback. Whatever he says, his thoughts may well run along these lines: “You can’t just decide to stop here, not just like that!”

(Gilbert, 2009: 173)

The first part of the interaction, with Rom asking Queenie to hurry up, illustrates the obligation criterion. Agents that are parties to a joint activity understand that they have a standing to demand of one another that they act in a manner appropriate to their joint activity, and to rebuke the other party should they act in a manner inappropriate to it. They understand, moreover, that this standing has its source in the joint activity itself. Gilbert takes this as evidence that participants in a joint activity have obligations towards each other to act in conformity with their shared intentions and correlative entitlements or rights to others so acting. The second part of the interaction with Queenie abruptly announcing that she won't go further and Rom's reaction to the announcement is evidence, according to Gilbert, that joint commitments cannot be rescinded or altered unilaterally but that the concurrence of all the parties involved is required for shared intentions to be rescinded or changed (concurrence criterion). Finally, Gilbert illustrates the disjunction criterion by means of another example. Ned and Olive had formed a shared intention to hike to the top of a hill together; at some point, Ned and Olive individually realized that they would only make it half way up the hill but refrained from informing each other. For theorists such as Michael Bratman, Ned and Olive stopped being engaged in joint action since they do not individually intend to get to the top of the hill and there is no longer a collective end. According to Gilbert, however, the Ned and Olive example shows that shared intentions cannot be reduced to individual intentions and also that they can persist even if individual intentions have ceased to exist since the subject of the shared intention is a plural subject. The concurrence of all parties is thus required in order for their shared intention to be rescinded. However, in the example there is no concurrence in the sense intended by Gilbert since neither Ned nor Olive neither of us have said anything about their change of mind to the other.

Gilbert's plural subject theory of shared intentions emphasizes the constitutive link between plural subjects and commitments. It is by being jointly committed that agents bind themselves into a plural subject. A web of obligations and entitlements forms the normative structure of plural subjects. Finally, the way the plural subject, once constituted, manifests its unity is by emulating as far as possible a single body that intends to do the thing in question.

Gilbert claims that all shared intentions involve joint commitments. Even if one has reservations with this strong claim, it is reasonable to consider that some (or perhaps a majority) of shared intentions and joint actions involve joint commitments. From a developmental perspective, this raises the question of when and how children start forming joint commitments. Gilbert allows that joint commitments can be formed through either explicit or tacit agreements. This suggests that plural subjects can have their basis in different types of coordination (notably, planned and emergent coordination; Knoblich, Butterfill & Sebanz, 2011; Tollefsen & Dale, 2012<sup>14</sup>). In fact, some joint actions might require more planification than others. As young children have limited planning skills (e.g., Paulus, 2016), they might need a higher degree of explicitness with obvious clues such as words like “with” or “together” (Tomasello, 2009: 67) compared to adults to feel committed to a joint action performance. In contrast, other joint actions might be more spontaneous and can arise from emergent coordination mechanisms (e.g., imitation, synchronization or alignment; see chapter 2 for more details). As young children tend to imitate adults and repeat over and over sequences of action notably with their caregiver (e.g., Rochat & Passos-Ferreira, 2009), it could be that mutual coordination in itself is at the origin of the plural subject foundation. Indeed, some researchers argue that these infant-caregiver daily plays support the infant’s participation in developing a joint sense and cooperative qualities (e.g., Fantasia, Fasulo et al., 2014<sup>15</sup>). Thus, shall we observe the ability to engage in spontaneous joint action earlier in development? Also, does joint sense-making require that children have a fully developed cognitive sense of self?

---

<sup>14</sup> Both Knoblich and colleagues (2011) and Tollefsen & Dale (2012) introduce the idea that the mechanisms involved in emergent coordination (e.g., imitation, synchronization, and alignment) can jointly engage participants. Tollefsen & Dale (2012) say that emergent coordination in itself can be sufficient for joint actions and can result in the formation of we-intentions. For instance, a joint action can arise from low-level mechanisms (emergent coordination) and might, in a second phase, become supplemented by the deeper commitments of planned coordination and collective intentionality. In other words, the initiation and the maintenance of a joint action can be based on different mechanisms and low-level mechanisms (from emergent coordination) can play a role in the constitution of we-intentions and planned coordination.

<sup>15</sup> Fantasia and colleagues (2014) explored early social games, i.e., vocal-kinetic play routines that mothers use to interact with infants from very early on with 3-month-old infants. Researchers tested infants’ participation and expectations in the game by measuring their level of engagement (facial expressions, gaze and body movement) while asking their mothers to perform two violation conditions (without gestures and without sound) compared to the usual play. Infants showed a significant decrease in body movements and expressions of positive affect and an increase in gazing away and stunned expression when the game structure was violated, indicating that the violated game conditions were experienced as less engaging.

Maximalists share the view that joint actions stem from shared intentions but disagree on how best to construe the notion of a shared intention. In this chapter, I discussed three versions of the shared intention hypothesis. What distinguish shared intentions from individual intentions may be their contents and interrelatedness (Bratman's intention condition), their representational mode (Tuomela's I-mode and we-mode we-intentions) or their subject (Gilbert's notion of plural subject). Each of these three competing accounts comes with its own set of requirements on shared intentions, but all appear to require high-level cognitive abilities, i.e., fully developed mentalizing and meta-representational abilities (corresponding to a robust theory of mind, see Tollefsen, 2005), since co-agents need to be able to represent the intentions of their partners and to incorporate them into their own intentions (in their content, in their mode of thinking or as a plural subject).



## Chapter 2

### *Motivating Minimalism*

As presented in chapter 1, maximalist theoretical models of joint action all require high-level cognitive abilities and “the imposition of such strong cognitive requirements would imply that animals and small children who lack altogether or have not yet fully developed mentalizing and meta-representational abilities as well as adequate mastery of the norms of practical rationality associated with planning cannot share intentions and engage in joint action” (Pacherie, 2011: p. 182<sup>16</sup>; see also Tollefsen, 2005; Pacherie & Dokic, 2006). In fact, maximalism accounts tend to limit joint action to human adults, excluding children and animals that do not fit the *shared intention hypothesis*. What would then be an alternative to the *shared intention hypothesis*? In other words, are shared intentions necessary for all types of joint actions?

Some researchers consider that while some sorts of joint actions require that co-agents share intentions (maximalism), others might be characterized without appealing to high-level states and sophisticated mindreading abilities. These researchers have developed a position called minimalism (see for instance, Tollefsen, 2005; Butterfill, & Apperly, 2012; Vesper et al., 2010; Pacherie & Dokic, 2006; Pacherie, 2011)<sup>17</sup>. In this chapter, I consider two types of resources minimalists can appeal to in

---

<sup>16</sup> “As pointed out by Tollefsen (2005) and Pacherie and Dokic (2006), [maximalists] analysis requires that the participating agents have concepts of mental states, since each participant should represent that the other participants have intentions and other attitudes relevant to the joint activity. It requires full-fledged meta-representational abilities insofar as the contents of the intentions of each participant make reference to both their own intentions and the intentions of the other participants. The imposition of such strong cognitive requirements would imply that animals and small children who lack altogether or have not yet fully developed mentalizing and meta-representational abilities as well as adequate mastery of the norms of practical rationality associated with planning cannot share intentions and engage in joint action. Although it is currently debated whether non-human animals ever engage in true joint action, existing evidence indicates that children clearly do and do so before they acquire the kind of cognitive sophistication [maximalists’] analysis suggests is required (Rakoczy 2006, 2007; Tomasello and Carpenter 2007; Tomasello et al. 2005; Warneken et al. 2006, Warneken and Tomasello 2007)” (Pacherie, 2011: 182).

<sup>17</sup> Paternotte (2020) distinguishes different kinds of minimalism: scale minimalism (joint actions that involve a minimal number of interchangeable agents), conceptual minimalism (that aims at characterizing joint actions in terms of small core of concepts), ontological minimalism (that attempts to keep to a minimum the number of potentially problematic entities appealed to in a characterization of joint action), cognitive minimalism (focussed

order to provide such characterizations. First, although young children do not yet possess all the cognitive abilities that human adults can deploy when acting together, they develop early on a set of social cognitive skills and abilities that enable them to engage in some forms of cooperative activities (cognitive prism). Second, contextual features such as the situation of action or the nature of the behavioral interaction itself can also help shape cooperative activities in a way that bypasses the need for sophisticated cognitive abilities (contextual prism). In this chapter, I review relevant empirical evidence and discuss which minimal cognitive capacities and minimal contextual configurations might be sufficient for some basic forms of joint action.

## **2.1 Motivating minimalism: the cognitive prism**

In the recent literature, there has been a trend towards “cognitive minimalism” (Paternotte, 2020). Several researchers have attempted to characterize a minimal core of (social) cognitive skills and abilities sufficient to enable some forms of joint action (e.g., Tollefsen, 2005; Butterfill & Pacherie, 2019; Vesper, 2016). For instance, Butterfill & Pacherie (2020) introduce a thought experiment in which they describe different creatures with limited cognitive and social skills. They argue that although these 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 that allow them to coordinate in a limited but useful range of ordinary circumstances. Cognitive minimalism assumes that there might be some kinds of joint action that do not require sophisticated high-level cognitive abilities and that proxies for, or precursors of, such cognitive abilities might be sufficient to perform them. One important motivation behind this trend towards cognitive minimalism is to offer characterizations that fit the cognitive abilities of less sophisticated agents (e.g., young children) that are able to act jointly.

### ***2.1.1 Young children’s social competencies in their first three years of life***

A body of research investigates young children’s competencies before they acquire the kind of sophistication required by maximalists in order to evaluate potential good candidates to support the

---

on identifying the minimal set of cognitive abilities needed for joint action and contextual minimalism (that explores contextually minimal joint action). My focus in this chapter is on cognitive minimalism.

existence of some minimal forms of joint action (for a review, see Brownell, 2011; see also Rakoczy 2006, 2007; Tomasello and Carpenter, 2007; Tomasello et al., 2005; Warneken et al., 2006; Warneken and Tomasello, 2007).

During their first months of life, infants participate in dyadic interactions with adults, engaging in dyadic joint attention and in proto-conversations where they progressively become tuned to the structure of the exchange, showing turn-taking abilities, exchanging looks, noises or mouth movements (2-6 months, see Goswami 2008; Brownell et al. 2009). Infants show interest and behavioral reactions (notably emotional ones) when the caregiver's attention is directed toward themselves or an object (Reddy, 2003). Around six months of age, they show anticipatory looking, revealing that they can predict an agent's actions and show surprise when something different from what they predicted happens (6 months, see Kovacs et al., 2007; Southgate & Verneti, 2013).

During the second half of their first year, infants start engaging in triadic interactions that include objects, events or persons outside the dyad. For instance, around 9 months of age, infants engage in 'sharing looks' with the adult, alternating between looking at the object of attention and engaging in communicative eye contact with each other (Hobson, 2005; Carpenter & Liebal, 2011; Venezia, Messinger, Thorp & Mundy, 2004). These 'sharing looks' are a marker of triadic joint attention (highest level of joint attention) which implies that two individuals look at the same object, understanding that the other is looking at the same object as an element of shared attention (Oates & Grayson, 2004; Reddy, 2005; Hobson, 2005). They also start engaging in play routines with toys together with their adult partner. These triadic activities are at first highly routinized, initiated and heavily scaffolded by the adult, but they progressively develop in complexity and at the close of their first year infants exhibit more control over the structure and unfolding of the joint action, initiating it or responding in more varied ways to the behavior of the adult (Brownell, 2011).

When infants start engaging in activities with others (such as proto-conversations, sharing looks or various kinds of infants' games) where there seems to be a common goal (the social engagement), forms of common knowledge, (mutual) belief (in joint attention) and mutual responsiveness (in turn-taking abilities and action prediction), it is legitimate to wonder whether these activities do not constitute basic joint actions. For instance, Tollefsen (2005) focuses on mutual responsiveness and

common knowledge which are key elements in Bratman's maximalist position. Common knowledge is a phenomenon classically defined, for instance by David Lewis (1969), as a hierarchy of 'A knows that B knows that ... knows that  $p$ ' propositions. This hierarchy presupposes that agents can form higher-order attitudes (for more details, see footnote 4 on p.24). However, some other researchers do not describe common knowledge in terms of such higher-order attitudes. Tollefsen (2005) argues that joint attention can be substituted for common knowledge both in infant and adult joint action, as joint attention would allow partners to know that they are focused on the same task, in mutual awareness of this. In addition, Carpenter points out that "1-year-olds understand something basic about common knowledge, in the sense of what is known or has been experienced together" (2009: 383) and that while this is not yet full-blown common knowledge, it takes us beyond joint attention in the sense that it is not tied to the perceptual present. Tollefsen also suggests that children's capacities of intentional prediction enable them to be mutually reactive. She specifies that children are capable of an intentional reading of intentions-in-action but not of prior intentions<sup>18</sup>, meaning that young children would only focus and rely on live interactions, hence their precision in following others' gazes and pointing (Woodward, 2005) and, in contrast, their poor abilities in action planning (Paulus, 2016). In other words, children would not use this complex cognitive hierarchy of higher-order attitudes but they would rather take into account that the other agent can see the scene as they do and use this information in their behavior (for discussions of common knowledge and physical constraints, see Liebal et al., 2009, 2010; Woodward, 2005).

In their second year of life, children show a predisposition to help others reach their goals (12-14 months, see Liszkowski et al., 2008; Buttelman et al., 2009) such as pointing in the direction of keys when an agent is looking for them. Young children are able to use common knowledge established during a previous interaction in a current interaction (14 months, see Liebal, Behne,

---

<sup>18</sup> Intentions-in-action have been introduced by John Searle (1983). For Searle, an action is composed of an intention and a movement; intentions-in-action are the cause and are contemporaneous of the agent's bodily movement that is its condition of satisfaction. Searle distinguishes intentions-in-action from prior intentions which are formed in advance of the action itself. In the case of intentions-in-action, Searle specifies that: (i) for the intention to be successful, the world must conform to the specified intentions' conditions; (ii) its causal domain extends only as far as the bodily movement of an action (Searle, 1983: 95), i.e., not covering the overall goal(s) or condition(s) that the movement presumably brings. Moreover, Searle considers that the intention-in-action "presents" its conditions of satisfaction which implies a more direct access to what is represented; still his position is a representationalist one as presentations are representations (Searle, 1983: 46).

Carpenter & Tomasello, 2009; Liebal, Carpenter & Tomasello, 2010; Ganea & Saylor, 2007). For instance, from one year of age, infants initiate gestural communication such as pointing to the most relevant objects based on the experience they had just shared with their communicative partner (Liebal, Carpenter & Tomasello, 2010) and respond to others' gestural and verbal communication (Ganea & Saylor, 2007; Liebal, Behne, Carpenter & Tomasello, 2009; Saylor & Ganea, 2007) differently based on the object or event they had previously attended to with their partner (compared to a random other partner). By 15 months of age, young children understand the link between eyes and seeing since they understand that physical objects can block sight (Woodward, 2005)<sup>19</sup>. By 18 months of age, young children grasp the intentional, referential nature of looking and they are skilled at following both gaze and pointing with precision (Woodward, 2005)<sup>20</sup>. In fact, around 18 months of age, typical children start demonstrating a conceptual or objectified sense of self (i.e., they can recognize and evaluate themselves, e.g., "that's me in the mirror, look how tall I am"; see Rochat & Zahavi, 2011) which emerges in conjunction with a general symbolic competence (i.e., a linguistic ability to refer to things differentiating signifier and signified across domains). In other words, it could be that such a double ability (better cognitive sense of self and general symbolic competence) enables young children to understand that the others' looks are "about" something (instead of just interpreting the gaze direction as a cue for action prediction). They also tend to have greater coordination abilities due to better general motor skills (18-24 months of age, see Southgate et al., 2008). For instance, in

---

<sup>19</sup> From an early age, infants show an understanding of the function of eyes and head movements, using a gaze following paradigm (Brooks and Meltzoff, 2002; Caron, Kiel, Dayton & Butler, 2002; Dunphy-Lelii & Wellman, 2004). For instance, Brooks and Meltzoff (2002) found that 12-month-old infants followed an experimenter's gaze direction and head turns more when the experimenter's eyes were open than when they were closed. When the experimenter's eyes were covered with a blindfold (rather than just closed), only 14- and 18-month-olds (not 12-month-olds) followed gaze less than when the eyes were open.

<sup>20</sup> Some studies tested children's understanding of the referential nature of looking by using a variation of the gaze following paradigm whereby researchers obstructed the experimenter's line of sight with a barrier (Butler, Caron and Brooks, 2000; Caron, Kiel, Dayton & Butler, 2002; Dunphy-Lelii and Wellman, 2004). In this kind of task, the child sits opposite to the experimenter who looks towards a target on the left or right. In the barrier condition, the experimenter's line of sight was obstructed by an opaque barrier. In the clear barrier condition, the barrier was a window enabling a clear line of sight to the experimenter and in the no-barrier condition, there was no barrier to obstruct the experimenter's line of sight. Based on the complexity of the barrier paradigm version, the understanding of the referential nature of looking emerges between 14 and 18 months of age. For instance, Dunphy-Lelii and Wellman (2004) found that 14-month-olds looked to the target more when the experimenter's line of sight was clear (clear barrier or no-barrier conditions), rather than when it was blocked by an opaque barrier.

non-verbal tasks where two children need to coordinate in order to retrieve some toys, only a few successful coordinations were reported in eighteen-month-old peers but they were largely coincidental but children succeed at above-chance at around 24 months (Brownell & Carriger, 1990; Brownell, Ramani, & Zerwas, 2006).

In a nutshell, during their second year of life young children develop better motor skills (including better coordination), they start using gestural communication (e.g., pointing) and respond to others incorporating a more fine-grained understanding of others in the environment (e.g., taking into account physical constraints regarding the ability to see, their precision in following gaze and pointing of others, the use of common knowledge from previous interactions). In other words, in their second year of life, children seem to acquire the ability to interpret and provide cues (e.g., communication and motor skills) which will be useful in a joint action context.

In addition, some researchers point out a socio-cognitive revolution at the end of the second year (Brownell, Nichols & Svetlova, 2013). They consider that there is a first socio-cognitive revolution by the end of the first year of life when children participate in joint attention and by extension, achieve new competencies in communication and interaction (Carpenter, Nagell, & Tomasello, 1998; Moore & Dunham, 1995). The second socio-cognitive revolution they identify occurs at the end of the second year of life when young children start being able to generate accessible representations of self and others in relation to one another and to the world (Barresi & Moore, 1996; Kagan, 1981; Perner, 1991; Zelazo, 2004), which enlarges children's prosocial behaviors in terms of context and partners. In fact, a body of research shows that children with more advanced self-other understanding – e.g., talking about their own and others' actions and internal states, taking adults' perspective or using personal pronouns (Brownell et al., 2006) – display more skillful cooperation with peers and represent their own and others' actions together in symbolic play at more advanced levels (Brownell & Carriger, 1990).

In their third year, young children progressively come to understand the normative dimension of collaboration and engage in forms of cooperation with others. From 21 to 27 months of age, young children clearly form common goals with other agents not because they need others to reach their own goals (*social tool hypothesis*) but for the sake of the interaction itself (*collaborative partner*

*hypothesis*); this suggests “that young children do not just view their collaborative partners as mindless social tools, but rather as intentional, cooperative agents with whom they must coordinate intentional states” (Warneken et al., 2012: 54). By 24 to 27 months of age, toddlers readily achieve common goals with peers, they are better able to coordinate their behavior both spatially and temporally. In addition, young children try to re-engage their partner who stopped playing (especially when they were unable rather than unwilling to continue) and they pay attention to games’ rules (25-27 months, Warneken and Tomasello, 2007; Gräfenhain et al., 2009, Warneken et al., 2012). Within games, they use gestures and verbal indicators to coordinate more efficiently with their partner (30 months, see Hamann et al., 2012). In their third year of life, children seem to form common goals with others (and are sensitive to the difference between a partner’s unwillingness or inability interact with them), use gestural and verbal communication to coordinate with others and show a better understanding of normative factors during their interaction i.e., they commit and expect commitment from their partner in games for instance.

Social competencies in children’s first three years of life seem to provide at least proxies for what are, according to maximalists, pre-requisites for joint action, such as common knowledge, mutual responsiveness, communication and motor skills, action prediction, commitment or (mutual) belief. In other words, children’s early social competencies (lower-level cognitive mechanisms) are good candidates to support some kinds of joint action. Is the development of these lower-level social cognitive abilities correlated with children’s performance in joint action?

### ***2.1.2 The study of joint action in developmental psychology***

In developmental psychology, joint action is studied with different experimental paradigms such as naturalistic games (Hay, 1979), joint go/no-go tasks (Milward, Kita & Apperly, 2014; Saby, Bouquet & Marshall, 2014; Meyer et al., 2015) or puzzle-solving tasks (Paulus, 2016; Warneken et al., 2006, 2012). Naturalistic games emphasize the role of motor and coordination skills coupled with the interaction parent/child in joint action performance. Joint go/no-go and puzzle-solving tasks emphasize the partner’s impact on the agent’s performance during a joint action.

### *2.1.2.1 Naturalistic games*

Naturalistic games focus on the parent-child interaction from an early age, infants being inclined to imitate and repeat over and over sequences of action, notably with their caregiver through daily plays (e.g., Rochat & Passos-Ferreira, 2009). In fact, some researchers consider that daily plays support the infant's participation in developing a joint sense and cooperative qualities (e.g., Fantasia, Fasulo et al., 2014). Fantasia and colleagues (2014) explored early social games (i.e., vocal-kinetic play routines) that mothers use to interact with infants from very early on (3-month-old infants). Researchers tested infants' participation and expectations in the game by measuring their level of engagement (facial expressions, gaze and body movement) while asking their mothers to perform two violated conditions (without gestures and without sound) compared to the usual play (baseline condition). Infants showed a significant decrease in 'body movements' and 'expressions of positive affect' and an increase in 'gazing away' and 'stunned expression' when the game structure was violated, indicating that the violated game conditions were experienced as less engaging. In continuity with infants' studies, toddlers' abilities can be explained with their development of motor and coordination skills along with a more advanced self-other understanding. For instance, Hay (1979) recorded early forms of cooperation and sharing displayed by 12-, 18-, and 24-month-old children while interacting with their parents in a play setting (e.g., coordinated operations on a mutual array of toys). The recordings revealed that cooperative interchanges and sharing (i.e., by showing and giving objects to the parent) appeared to increase in frequency in the course of the second year. As cooperative exchanges, Hay (1979) measured that, around 18 months of age, toddlers are able to participate in games such as throwing a ball back and forth between themselves and a partner (Hay, 1979) which seems coherent with the evolution of motor and coordination skills between 18-24 months of age (see Southgate et al., 2008). Other studies have shown that self-other understanding (e.g., talking about their own and others' actions and internal states, taking adults' perspective or using personal pronouns, see Brownell

et al., 2006) is correlated with a more skillful cooperation with peers<sup>21</sup> notably during symbolic play (i.e., children use objects to represent or symbolize other objects, e.g., a banana becomes a cell-phone or play tea parties). Furthermore, there is an impact of the type of goal at stake in children's interactions; when the joint action entailed more concrete goals (e.g., making a ball bounce on a trampoline) rather than abstract goals (e.g., maintaining interaction with a partner, as in ball throwing), children were less able to coordinate their actions with a partner at 18 months, but even so, they could succeed at above-chance at around 24 months (Warneken, Chen & Tomasello, 2006). In their third year of life, children start integrating normative factors and engaging in more complex forms of cooperation with others (Warneken et al., 2012; Hamann et al., 2012). Such games with their caregiver and later with peers would accentuate young children's abilities in developing a joint sense and cooperative qualities by creating more experiences of social engagement (as a common goal at first), forms of common knowledge, (mutual) belief (in joint attention) and mutual responsiveness (in turn-taking abilities and action prediction).

#### 2.1.2.2 Joint go/no-go tasks

Joint go/no-go tasks<sup>22</sup> involve a continuously presented series of stimuli with the display of "go" cues to which subjects are instructed to respond as rapidly as possible and of "no-go" cues to which

---

<sup>21</sup> In fact, Bakeman and Adamson (1984) found that 18-month-olds' play consists of 25% joint activity with their mothers, whereas only 7% was with a peer.

<sup>22</sup> In *individual configurations*, no/no-go tasks give rise to a *spatial stimulus-response effect* referred as the Simon effect (Simon & Wolf, 1963). In the *standard Simon task*, for instance, participants are placed in front of a screen and instructed to press down on the left key (with their left hand) when a green circle is displayed and to press down on the right key (with their right hand) when a red circle is displayed. The green and red circles appear successively either on the left or on the right side of the screen. The Simon effect corresponds to the fact that participants respond faster (shorter reaction time) and more accurately (lower error rate) to stimuli displayed on the side of the screen which is congruent with the limb they use to answer (stimulus-response congruency; although the stimulus location is irrelevant to the task). Note that the Simon effect does not emerge when participants only perform a *simple go/no-go task*, i.e., when they are instructed to press down on a single key only when one of the two stimuli is displayed. In *joint configurations*, two participants are partnered and instructed to respond as in the go/no-go task, i.e., one stimulus is assigned per participant who has his/her own key to press down on when his/her stimulus is displayed on the screen (the side of the screen where stimuli are displayed is irrelevant to the task); stimuli are displayed successively (Sebanz, Knoblich & Prinz, 2003). The goal of this paradigm is to see whether participants would act as in a *simple go/no-go task* (no Simon effect) or if they would act as in a *standard Simon task* i.e., represent their partner "as their other limb" (Simon effect). The results reveal the *joint Simon effect* i.e., as when participants are confronted with the standard Simon task and although each participant only needs to focus on one stimulus, participants' response time is shorter and responses are more accurate when there is a stimulus-response congruency. In other words, the joint Simon effect suggests that agents do not merely represent their own part of the task, but they rather represent the goal of the task as a joint goal. This paradigm referred to as the 'social' or joint Simon task has been extensively used to

subjects should refrain from responding. For instance, some studies adapted the Bear Dragon task (Kochanska, Murray, Jacques, Koenig, & Vandegest, 1996) in which a pair of children were instructed by puppets to point to a picture; they were required either to point when one puppet told them or to inhibit pointing when instructed to by the other puppet. In the same condition, both children were required to respond to the same puppet (same “go” and “no-go” cues) and in the different condition, children were required to respond to different puppets (different/reversed “go” and “no-go” cues; see Milward, Kita & Apperly, 2014). In joint go/no-go tasks, researchers aim at measuring a co-representation effect (referred as the joint Simon effect, see footnote n°22) i.e., to measure whether adults or children do co-represent their partner while performing the joint task. The presence of a co-representation effect in adults is considered as evidence that participants are not only representing their own task but also aspects of their partner’s intentions, goals or actions (Atmaca, Sebanz & Knoblich, 2011). Milward, Kita & Apperly (2014) found a co-representation effect only in 4- and 5-year-olds (not in 2- and 3-year-olds), meaning that only from four years of age, children perform significantly better in the same condition compared with the different condition (the presence of another child doing a different task interferes with their performance)<sup>23</sup>. Following these results, Milward, Kita & Apperly (2016) devised a follow-up study assessing individual differences which revealed that theory of mind, helping children to clearly differentiate and separate task representations of self and other<sup>24</sup>, and inhibitory control skills were predictors of the ability to avoid interference from co-representation effects<sup>25</sup>. The fact that theory of mind helps children to clearly differentiate and separate task

---

study joint action (e.g., Dittrich, Rothe, & Klauer, 2012; Iani, Anelli, Nicoletti, Arcuri, & Rubichi, 2011; Klemnova & Liepelt, 2016; Kuhbandner, Pekrun, & Maier, 2010; Liepelt, Wenke, & Fischer, 2013; Liepelt, Wenke, Fischer, & Prinz, 2011; Stenzel et al., 2012, 2014; Tsai & Brass, 2007; Vlainic, Liepelt, Colzato, Prinz, & Hommel, 2010; Welsh, 2009; Welsh et al., 2013) and is considered as evidence of co-representation of the partner in a joint configuration.

<sup>23</sup> “Although children’s performance was clearly above-chance in all conditions, they performed significantly better in the Same Task condition than in the Different Task condition, suggesting that the presence of another child doing a different task interferes with performance” (Milward et al., 2014: 276).

<sup>24</sup> Theory of mind (ToM) was tested using Wellman and Liu’s (2004) ToM scale (specifically using knowledge access, contents false belief, and real apparent emotion tasks). ToM ability was positively correlated with children’s performance during the task (i.e., ability to avoid interference from a partner and thereby perform well), meaning that ToM helps children to separate representations of self and other and thus avoid interference.

<sup>25</sup> Inhibitory control skills (IC) were measured using Day–Night (Gerstadt, Hong, & Diamond, 1994) and Pictures tasks (Burns, Riggs, & Beck, 2012). Participants with stronger inhibition abilities performed better on the different task once controlling for the same task. ToM and IC each had an independent positive relationship

representations of self and other suggests that before four years of age, children might perform better in joint tasks not requiring these separate task representations (thus not leading to the emergence of interference in self-performance due to a co-representation effect; e.g., tasks in which co-agents imitate each other, such as lifting up an object together). In fact, Meyer and colleagues (2015) instructed 2 ½- and 3-year-olds to perform complementary (taking turns to press down on their buttons) actions with an experimenter in order to get a character climb on a ladder (after children played individually). Meyer and colleagues (2015) found that toddlers who showed more inhibitory control in an individual context were more accurate in their turn-taking performance during the joint action and that toddlers with more precision in their action predictions were less variable in their action timing during the joint play. Meyer and colleagues (2015) found no relation between timing variability and inhibitory control or between turn-taking accuracy and action prediction. In other words, this study reveals that action prediction and inhibitory control have distinct roles in coordination.

In joint go/no-go tasks, it seems that (1) the co-representation effect is a mechanism whereby participants represent what their partner is doing, (2) inhibitory control skills play a crucial role in turn-taking performance – disentangling whose turn it is to act (Wenke et al., 2011) – and (3) action prediction (presumably associated with quite good motor skills) improves coordination by making action timing more accurate during a joint action. However, before 4 years of age, children do not present a co-representation effect during joint go/no-go tasks<sup>26</sup>. These results appear to have two implications. First, they cast doubt on the idea that activities, performed before the apparition of a co-representation effect, are truly “joint” actions rather than coordinated individual actions (Brownell, 2011; Milward et al., 2014). Second, they suggest that theory of mind is necessary to represent the partner’s task, notably since theory of mind helps children in the clear differentiation and separation of task representations of self and other (Milward et al., 2016). Indeed, the ability to complement each other’s actions requires one to maintain and confront self and other task representations; if children

---

with the ability to avoid interference from a partner in a joint task (i.e., less conflict in the different task condition).

<sup>26</sup> Children are not significantly better in the same condition compared with the different condition, meaning that the presence of another child doing a different task does not interfere with their performance.

succeed in complementing their partner's action, it means that they represent what she is doing (other task representation) and that they are able to identify a way to complement her action (self task representation) in the realization of the common goal. However, an alternative explanation could be that representing what the partner is doing can be more or less challenging due to the task itself and, more specifically, to the skills needed to successfully complement the partner's action (e.g., problem-solving, flexibility, action planning skills, motor skills). If the alternative explanation is correct, (1) children's ability to perform complementary joint tasks (i.e., ability to complement their partner's action) should be related to the skills required by the task itself to successfully complement the partner's action; (2) young children's ability to maintain and confront self and other representations in simple complementary tasks might involve a lower-level mechanism which develops earlier than theory of mind, such as visual perspective-taking skills (Moll & Meltzoff, 2011a,b), in children's proactive planning to accommodate others, instead of developed mentalizing and meta-representational abilities (theory of mind).

### *2.1.2.3 Puzzle-solving tasks*

In puzzle-solving tasks, children's action planning is tested while they interact with another person to operate an apparatus; the idea is to measure whether children are able to accommodate their partner's action to perform the joint action. Note that puzzle-solving tasks constitute the perfect case to put to the test the alternative explanation presented above.

During their second year of life, children begin to successfully solve collaborative puzzle-solving tasks that require behavioral coordination (Brownell, Ramani, & Zerwas, 2006; Warneken, Chen, & Tomasello, 2006). For instance, Brownell and Carriger (1990) presented 12- to 30-month-old peers with a very simple collaboration task in which one child had to push a lever to align a toy beneath a hole in a Plexiglas table so that their partner could then retrieve it. In these tasks, each individual action is quite simple but requires that peers coordinate in time. The results showed that 18-month-old peer pairs were coordinating infrequently and accidentally but that 24- and 30-month-old peer pairs were coordinating quickly and efficiently. Meyer and colleagues (2016) assessed 2.5-, 3-, 3.5- and 5-year-old children's proactive planning (without prior experience) and flexible adjustment of

action plans over time in a joint cup-stacking task. Children were instructed to pass cups to their partner (the experimenter) who had only one hand available (alternating over time) to build a tower; researchers measured whether children were able to accommodate their partner's action by handing the cup to her free hand (rather than her occupied hand). Meyer and colleagues (2016) revealed that, at all ages, children proactively planned their actions in a way that accommodated their partner's actions<sup>27</sup> but that there is a developmental increase in children's ability to flexibly adjust their action plans to accommodate their joint-action partner<sup>28</sup>. In fact, it is only by 3.5 years that children start to flexibly integrate their partner into their action plans and even at age 5, children only showed minimal adjustments to their action partner (5-year-olds only did so in less than 60% of the trials).

In contrast, Warneken, Steinwender, Hamann & Tomasello (2014) designed a study whose two conditions required close attention to the actions of the partner but only one of them additionally required the anticipation of the partner's constrained tool choice. Researchers tested 3- and 5-year-old peer pairs who were presented with two different tool choice situations – a bidirectional condition and an unidirectional condition – in which they had to choose complementary tools with which to subsequently work on a collaborative problem-solving apparatus. In the bidirectional condition, exemplars of the two necessary tools appeared in front of each child. In the unidirectional condition – which is the condition which additionally required the anticipation of the actions of the partner – one child had to choose between two different tools first, while the other child had only one tool available. Five-year-olds were proficient planners in both conditions. Three-year-olds did not consistently make the correct choice (only 3-year-olds who had first experienced the unidirectional condition chose the correct tool at an above-chance level). These results provide evidence that between 3 and 5 years of age children develop the ability to plan the division of labor in a collaborative task. Finally, in Paulus (2016), 3-, 5-, and 7-year-old children and adults were instructed to hand the experimenter a tool with

---

<sup>27</sup> Children's proactive action planning is assessed at the very start of the joint-action game, when children decide which side to pass the first cup. In fact, 70% of all children passed the cup to the joint-action partner on her free-hand side.

<sup>28</sup> The flexible adjustment to their partner is assessed along the task, while the experimenter's free hand changes and children need to switch side in order to accommodate her. In this case, only by age 3.5 children show flexible adjustment to the partner's constraints as their performance differed from chance.

which she could activate 1 of 2 different effects on the apparatus (based on the orientation of the tool during its insertion). In order to elicit a specific effect, participants should plan their grasping and reaching actions in such a way as to enable the partner to efficiently handle the tool, i.e., anticipating the final end state of the joint activity. Adults demonstrate joint action planning from the beginning, 5- and 7-year-olds improve their joint action performance over time but 3-year-olds children do not adjust their behavior to accommodate the other's action (they also do not improve over time and first-hand experience with the task before acting with a partner does not facilitate performance in the joint action task).

Puzzle-solving tasks reveal that children's ability to perform complementary joint task vary based on the skills required to successfully complement the partner's action (e.g., problem-solving, flexibility, action planning skills, motor skills). In fact, children are better able to complement their partner's task using motor skills for flexible adjustment during the task before they can use action planning or problem-solving skills.

Observations from developmental psychology reveal that young children use certain abilities such as turn-taking, action prediction, common knowledge, coordination skills or normative factors to perform activity which look like joint action before the age of four i.e., before they have acquired a robust theory of mind<sup>29</sup>. Relatively to the cognitive minimalist position, children's social competencies in their first three years of life seem to constitute at least proxies for the main maximalist prerequisites for joint action (e.g., common knowledge, mutual responsiveness, communication and motor skills, action prediction, commitment or mutual belief). The only cloud on the horizon is the absence of co-representation effect before 4 years of age, which could suggest that children need to have a fully developed theory of mind to perform truly 'joint' actions (and not coordinated individual actions). If children do not co-represent their partner during a joint activity, they should not be able to complement her action in any type of joint activity before the apparition of a robust theory of mind. However, children's ability to perform complementary joint tasks (i.e., ability to complement their

---

<sup>29</sup> From a developmental point of view, it is crucial to discuss the maximalist account to the extent that maximalism requires some high-level cognitive capacities such as a robust theory of mind which, as I will discuss in detail in chapter 3, many developmental psychologists argue children before the age of four have not yet fully developed (Wellman, Cross & Watson, 2001).

partner's action) vary based on the skills – problem-solving, flexibility, action planning skills, motor skills – required by the task itself to successfully complement the partner's action (these skills develop progressively during childhood), meaning that theory of mind is not necessary to perform at least some forms of joint actions. In other words, children do co-represent their partner during a joint activity when they possess the skills required to perform the task (e.g., motor skills develop earlier than problem-solving skills).

How then do children co-represent their partner's task if they do not use theory of mind? My hypothesis is that, before four years of age, children use elements from the interaction itself to inform their self and other task representations. These elements are accessed thanks to visual perspective-taking abilities (Moll and Meltzoff, 2011) whose level depend on the strategy co-agents adopt to maximize the joint action performance (based on the trade-off between the task requirements and the uncertainty of the information available about the partner; for a recent review, see Chackochan & Sanguineti, 2019). I will discuss the involvement of visual perspective-taking abilities in joint action in chapter 7.

## **2.2 Motivating Minimalism: the contextual prism**

In the literature, other minimalists defend an approach we can call “contextual minimalism”, considering that some contextual features could be sufficient to trigger coordination without the need for any sophisticated cognitive abilities (Vesper, Butterfill, Knoblich & Sebanz, 2010; Tollefsen, 2005).

### ***2.2.1 Motor-related contextual cues***

The context in which an action is performed is relevant to the agents' motor planification, through agents' reading of motor-related information.

#### ***2.2.1.1 Minimal joint actions: concreteness and spontaneity***

Joint actions may be split in two categories based on their cognitive abilities requirements – high-versus low-level cognitive abilities – minimalists being interested in the second category. Some joint actions which require high-level cognitive abilities can be categorized by their abstractness and the

planning skills their performance requires (e.g., playing the Kemps card game; see Appendix 2). Other joint actions which only require low-level cognitive abilities can be categorized by their concreteness (motor component) and their spontaneity, where goals and ways of coordinating emerge from salient features of the situation (e.g., picking up someone who has fallen over or lifting up a tree fallen together). For instance, playing Kemps requires high-level cognitive abilities as it involves bluffing and secret communication with a partner (abstract and planned) whereas lifting up together a tree fallen on the road because it is heavy and blocks the way is more concrete and seems to require a better motor understanding of the co-agent (concrete and spontaneous). In the Kemps card game, players need to read their partner who is both supposed to trick their opponents and get their co-player to understand when it is the right moment to say “Kemps!” to win the game. In other words, this is a game context with rules (which might involve an explicit agreement or commitment) and mindreading abilities are required to read the partner (and opponents) during the game. In the example of lifting a tree fallen because it blocks the way, two individuals might jointly attend to the tree and spontaneously act together (which might involve an implicit agreement or commitment). In between, there might be intermediary joint actions such as moving an object from one room to another in which a strategy of planning ahead could be associated with basic motor understanding (concreteness) throughout the process. In this case, there might be an explicit agreement (or commitment) between agents and an online adjustment during their joint action. What elements from the interaction itself enable participants to successfully coordinate?

#### *2.2.1.2 Coordinating in live interactions*

Vesper and colleagues (2010) discuss “coordination smoothers” which could be either a modulation of one’s own behavior in order to facilitate the coordination or specific elements of an object such as size, form, weight which afford a particular task distribution.

In the case of agents’ behaviors modulations regarding specific elements of an object, if two agents decide to lift up an earthenware vase or a shopping bag with two handles, their actions will be influenced by the structure of the object, namely the handles. Some studies suggest that infants from 5 to 15 months of age incorporate visible information about an object’s structure into their action on the

object (e.g., Barrett et al., 2008). Agents could also modulate their behaviors by delimiting and structuring their action space in order to be more predictable. In the case of the earthenware vase, they could place themselves on one specific side; in a joint building task, agents could try not to enter the other action space as much as possible, enhancing turn-taking (Vesper, Soutschek, & Schubö, 2009).

In the case of agents' behavior modulations aimed at facilitating coordination, other coordination smoothers considered as coordination signals such as gestures (Goebel and Palmer, 2009) and synchronization (Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007) might be taken into account. During joint actions, co-agents use sensorimotor communication<sup>30</sup> to make their intentions and action goals more salient for their co-agent, notably via specific kinematic cues in order to be more predictable for each other (for a recent review, see Pezzulo et al., 2019; see also, Vesper et al., 2010; Pezzulo, 2011a,b; Pezzulo and Dindo, 2013; Sacheli et al., 2013; Candidi et al., 2015; McEllin, Knoblich and Sebanz, 2018). In fact, these coordination smoothers are crucial elements of the interaction itself allowing agents to better predict and adjust their behaviors notably regarding the incorporation of concrete factors (compared to abstract factors). In addition, contextual elements (e.g., playing a game or be willing to get around an object that blocks the way) could trigger a commitment between agents and emphasize the salience of their implicit communication (e.g., eye gaze, gestures).

### ***2.2.2 Emotion-related contextual cues***

Even with motor-related contextual cues, the persistence of intentions is important to the proper performance of the joint action. In this case, adults and young children can focus their attention on emotion-related contextual cues from others around them. For instance, Tollefsen (2005) considers that joint attention and social referencing – infants use adults' emotional expressions (positive or negative) to decide whether to approach or avoid objects and people (Campos & Stenberg, 1981) – are sufficient to allow a persistence of intentions-in-action. In the case of helping behaviors where children see an adult in difficulty and spontaneously offer their help, it could be that the social

---

<sup>30</sup> Sensorimotor communication is described as a “natural form of communication that does not require any prior convention or any specific code...[which] amounts to the continuous and flexible exchange of bodily signals, with or without awareness, to enhance coordination success; and it is versatile, as sensorimotor signals can be embedded within every action” (Pezzulo et al., 2019).

environment favors a process of rule internalization leading to the emergence of such behaviors (e.g., Hammond & Carpendale, 2012) or that such helping behaviors originate from the prosocial nature of humans who have an intrinsic motivation of helping others<sup>31</sup> (Warneken & Tomasello, 2013). More precisely, one-year-olds behave prosocially only in obvious goal-oriented helping situations with an adult where the needs of another person and what to do to help are clear (Brownell, Nichols & Svetlova, 2013). For instance, from one year of age, young children will spontaneously do things like opening a cabinet door when they see an adult whose hands are full or returning a pen or a clothespin that an adult has accidentally dropped (Warneken & Tomasello, 2006, 2007, 2013). Toddlers also seem to be more likely to help a familiar person compared to an unfamiliar person (Allen, Perry and Kaufman, 2018). Furthermore, previous studies have shown that from 18 months of age, infants also use previously observed affects in order to represent the motivation state of others (Repacholi and Gopnik, 1997; Reschke, Walle and Dukes, 2020). As described by Repacholi & Gopnik:

Children observed an experimenter expressing disgust as she tasted one type of food and happiness as she tasted another type of food. They were then required to predict which food the experimenter would subsequently desire. The 14-month-olds responded egocentrically, offering whichever food they themselves preferred. However, 18-month-olds correctly inferred that the experimenter wanted the food associated with her prior positive affect. They were able to make this inference even when the experimenter's desires differed from their own.

(Repacholi & Gopnik, 1997: 12)

These studies show that for young children, the context associated with their ability to read others' emotional expressions plays a crucial role in identifying the motivation state of others and acting on it.

---

<sup>31</sup> Children's motivation to help an adult in difficulty is thus independent from socialization practices (e.g., social norms) or the encouragement of children's caregivers.

In other words, by exploiting contextual features, using elements from the interaction itself or considering the other agent's emotional cues as markers of their motivational state to perform joint actions, agents may be able to successfully engage in joint activities using limited cognitive resources.

The type of joint actions categorized by their concreteness and spontaneity might form a category of minimal joint actions; as proposed by Tollefsen (2005), some proxies for more sophisticated mindreading abilities might be sufficient for adults to perform this category of joint actions. Besides, concrete and spontaneous joint actions seem to match with children's early social competencies. While by 18 months of age, successful coordination with peers is largely coincidental (Brownell & Carriger, 1990; Brownell, Ramani, & Zerwas, 2006), between eighteen and twenty-four months of age, young children's coordination abilities improve along with better general motor skills (18-24 months of age, see Southgate et al., 2008). In other words, young children's coordination seems to rely on live interactions and to depend on general motor skills. In sum, minimalists interested in the cognitive prism have been studying mechanisms at the origins of action prediction, control and monitoring in joint action, assuming that some kinds of joint action do not require sophisticated high-level cognitive abilities and/or that proxies for, or precursors of, such cognitive abilities might be sufficient to perform some kinds of joint action. In a complementary way, minimalists interested in the contextual prism have shown that the exploitation of various types of contextual cues can also help children successfully engage in joint action.

\*\*\*

Maximalists defend variants of the *shared intention hypothesis* i.e., the hypothesis according to which joint actions require shared intentions. What distinguishes shared from individual intentions may be a matter of representational content (Bratman's intention condition), representational mode (Tuomela's I-mode and we-mode we-intentions) or subject (Gilbert's notion of plural subject). From maximalism variants, a set of requirements describing joint actions emerge, including mutual responsiveness, common knowledge, interdependence, (mutual) belief and (joint) commitment. For agents to form such shared intentions, they need to appeal to high-level cognitive abilities in order to encode and incorporate their co-agent(s) in mental terms.

However, a body of research called minimalism argues that some kinds of joint action do not require fully developed mentalizing and meta-representational abilities (i.e., a robust theory of mind). Observations from developmental psychology reveal that young children use certain abilities such as turn-taking, action prediction, joint attention, coordination skills or normative factors to perform actions which look like joint action before the age of four i.e., without a robust theory of mind. Minimalists are concerned by the existence of joint actions relying on minimal cognitive skills or competencies (cognitive prism) and/or on external features such as the context or the behavioral interaction in itself (contextual prism). Minimalists propose that some joint actions might require only goal representations (instead of intentions). In fact, joint actions where several agents share a common goal and which involves spontaneous and motor factors might form a good category of minimal joint actions. Minimal cognitive abilities such as proxies for or precursors of, fully developed cognitive abilities (which could be found in children's early social competencies) and contextual features such as intentions-in-action reading, the use "coordination smoothers" or the ability to understand the motivational state of others might be sufficient to perform such joint actions.

*Part II*

*Developmental psychology*

*Children's mindreading and perspective-taking capacities*



## *Introduction to Part II*

Maximalist theoretical models of joint action all demand high-level cognitive abilities, requiring that participating co-agents have concepts of mental states enabling them to represent other participants' intentions and other attitudes relevant to the joint activity, can reason about the mental states of their co-agent and incorporate them in the content of their own intentions as described in Chapter 1. As Pacherie puts it: “the imposition of such strong cognitive requirements would imply that animals and small children who lack altogether or have not yet fully developed mentalizing and meta-representational abilities as well as adequate mastery of the norms of practical rationality associated with planning cannot share intentions and engage in joint action” (Pacherie, 2011: 182; see also Tollefsen, 2005; Pacherie & Dokic, 2006). Consequently, maximalist accounts tend to limit joint action to human adults, excluding children and animals. Yet, what about certain forms of collective behaviors observed in many animal species? And what about children who play together with adults and peers from a very young age? Are they engaging in some minimal form of shared cooperative activity?

Ethologists have documented the cooperative abilities and behaviors of social carnivores such as hyenas and lions (Drea & Carter, 2009; Stander, 1992), organized communities of leafcutter ants or bees in which each member has its own specific role, elephants' protecting their young from predators by creating a circle around them, fire and army ants' strategies to overcome an obstacle using their bodies to create bridges or rafts to cross a stretch of water, or again bottlenose dolphins' food-retrieval behaviors (see for review, Kuczaj, Winship & Eskelinen, 2015; see also King et al., 2016). Similarly, primatologists have investigated the cooperative abilities and behaviors of non-human primates (see Suchak et al., 2016; for a recent review and see also Schmidt & Tomasello, 2016). In addition, existing evidence suggests that children can engage in some forms of joint action even before they acquire the kind of cognitive sophistication required by maximalists (for a review, see Brownell, 2011; see also Rakoczy 2006, 2007; Tomasello and Carpenter 2007; Tomasello et al. 2005; Warneken et al. 2006, Warneken and Tomasello 2007). Should this lead us to take a pluralist stance about joint actions and accept that there are different types of joint action? In other words, do those observed behaviors

constitute types of joint actions that only require lower-level cognitive abilities when more complex types of joint actions would require higher-level cognitive abilities?

In order to properly discuss the implications of developmental psychology observations for existing accounts of joint action, it will be crucial to review the different current stances regarding theory of mind abilities in childhood. The high-level cognitive abilities maximalist accounts appear to require relate to theory of mind abilities, i.e., the ability to understand that other agents have different beliefs, desires, and knowledge than oneself (Premack & Woodruff, 1978). Theory of mind abilities have been widely tested in experiments using false-belief tasks, i.e., tasks where children are asked to predict an agent's action based on that agent's (mis)representation of the state of affairs (Wellman et al., 2001) such as an object's location (see for instance, Wimmer and Perner, 1983). Wimmer and Perner's findings were widely replicated with a meta-analysis of 178 studies confirming that around 4 years of age, children switch systematically from acting and talking presumably without an understanding of false beliefs to being able to recognize false beliefs (Wellman, Cross and Watson 2001). In classic false-belief tasks, children were asked to answer a question verbally, so explicitly. Other studies, in contrast, started using implicit measures such as anticipatory looking (Clements and Perner, 1994; Southgate, Senju and Csibra 2007), violation-of-expectation paradigms (Onishi & Baillargeon, 2005; He, Bolz and Baillargeon, 2011; Träuble, Marinović and Pauen 2010) or interactive tasks (Buttelman et al, 2009; Knudsen & Liszkowski, 2012; Southgate et al, 2010). If explicit measures revealed that under four years of age, children were unable to pass false-belief tasks, implicit measures showed that two- and three-years-olds were able to pass the test when non-verbal measures were used. Researchers have offered different interpretations of these results<sup>32</sup>. Some researchers, namely nativists, consider that implicit measures are proof that children do possess a robust theory of

---

<sup>32</sup> Note that the robustness of these paradigms has been called into question based on a certain number of partial or failed replications (Burnside et al., 2018; Dörrenberg et al., 2018; Grosse Wiesmann et al., 2017; Kulke & Rakoczy, 2017, 2018; Kulke et al., 2018, Powell et al., 2018; Priewasser et al., 2018; Schuwerk et al., 2018), leading researchers to debate the replicability of infant ToM findings (Baillargeon et al., 2018; Poulin-Dubois et al., 2018). However, two large-scale projects have emerged, namely ManyBabies and ManyBabies2, which gathered a group of authors from the original studies in a multi-lab initiative aimed at assessing the replicability and robustness of key findings in development, including findings on infant ToM. ManyBabies2 which was launched in 2019, focuses on a large-scale conceptual replication of anticipatory looking-based false belief tasks in 18-36-month-olds and adults (based on Southgate et al., 2007; Surian & Geraci, 2012). In this chapter, our aim is not to discuss extensively this debate, but rather to evaluate the implications of each interpretation for joint action.

mind from birth (e.g., Scott & Baillargeon, 2010, 2017), when other researchers respectively, empiricists and dual theorists, consider that children rely on teleological reasoning (e.g., Gergely & Csibra, 2003; Perner & Roessler, 2012) or use a minimal theory of mind (e.g., Apperly & Butterfill, 2009; Surtees et al., 2016) under four years of age.

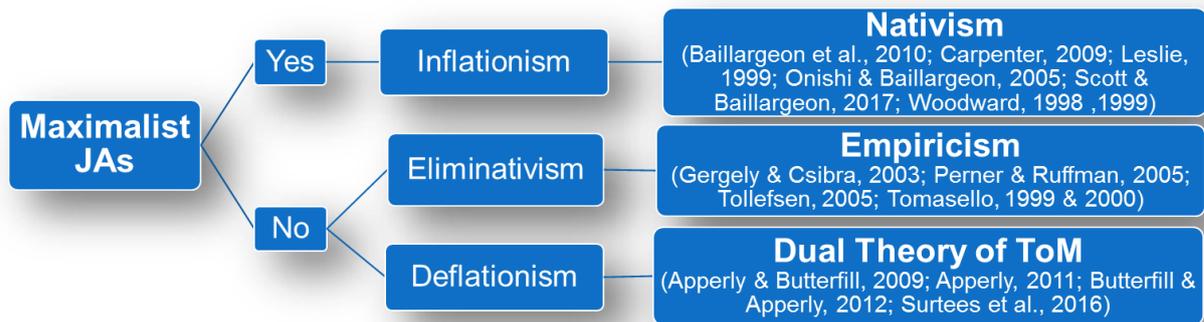
Developmental psychology data raise difficulties for maximalism since young children seem to be able to successfully engage in at least some forms of joint actions. From a developmental point of view, these observations are crucial to the extent that maximalism requires high-level cognitive capacities such as a robust theory of mind which, according to standard psychological accounts of the development of theory of mind, children before the age of four have not yet developed (Wellman et al., 2001). Advocates of minimalism (see for instance, Tollefsen, 2005; Butterfill, & Apperly, 2012; Vesper et al., 2010; Pacherie & Dokic, 2006; Pacherie, 2011) consider that some kinds of joint action could rely, as discussed in Chapter 2, on internal features such as minimal cognitive skills or competencies and/or on external features such as the context or the behavioral interaction in itself.

The debate between maximalism and minimalism regarding children's ability to engage in joint action can be represented in the form of a trilemma (Rakoczy, 2018):

- (1) Shared cooperative activities require sophisticated cognitive abilities to represent, and reason about, mental states;
- (2) Young children do not yet have those sophisticated cognitive abilities;
- (3) They engage in shared cooperative activities.

In order to solve this trilemma, each proposition needs to be evaluated separately in order to assess its plausibility.

A robust theory of mind entails an understanding of others in terms of thoughts, intentions and beliefs, but also the ability to dissociate these thoughts, intentions and beliefs from one's own and thus, to understand that others can form false beliefs that do not represent correctly the reality (Tollefsen, 2005). There are three main positions regarding children's theory of mind abilities: nativism (inflationists), empiricism (eliminativists) and the dual model of theory of mind (deflationists). Each of these positions takes a different stance regarding the cognitive abilities of children and whether they are consistent with the demands made by maximalists accounts (see Fig. 1).



**Fig. 1:** Trilemma answers about the maximalist account of joint actions. Nativism considered as an inflationist position would say that observations fit the maximalist account of joint actions. Empiricism and ToM dualism respectively considered as eliminativist and deflationist positions would say that observations do not accord with the maximalist account of joint actions.

Nativists consider that children have an innate ability to represent mental states, meaning that they are able to engage in joint actions in the maximalist sense. Nativism evacuates the sophisticated cognitive ability problem because if young children are able to represent and ascribe mental states to others, then they should be able to form shared intentions. However, even assuming that young children are capable of forming shared intentions, the debate between maximalists and minimalists would not yet be settled. What would remain at issue is whether all sorts of joint actions require shared intentions or whether there are minimal joint actions that do not necessitate shared intentions. Empiricists think that young children do not have these sophisticated abilities before the emergence of meta-representational cognitive resources, i.e., not before 4 years of age. They draw a distinction between a mentalistic and a teleological stance when interpreting others' actions. Whereas taking a mentalistic stance involves attributing beliefs, desires and intentions to the observed agent, taking a teleological stance involves inferring the goal of an action from the action itself together with situational constraints, through a principle of rational or efficient action (Gergely & Csibra, 2003) or the application of behavioral rules (Perner & Ruffman, 2005). Empiricists consider that young children use a teleological understanding of others, meaning that to perform joint action they rely on less complex cognitive abilities than those demanded by maximalist accounts. Finally, the dual model introduces a different approach by arguing in favor of the existence of two distinct systems: one implicit from which originates an understanding of relational mental states (unconscious, automatic,

inflexible and quick) and the other explicit from which originates an understanding of propositional mental states (conscious, costly in terms of cognitive resources, flexible and slow). The dual theory of mind theorists focus on whether there is something else than a robust theory of mind that would allow young children to achieve at least some kinds of joint actions.

Consequently, to the question of whether observations in developmental psychology confirm the maximalist account of joint actions, a positive answer is obtained based on the nativist account and a negative answer based on the eliminativist and the deflationist accounts (see Fig. 1). In other words, the nativist account accepts the trilemma's propositions (1) and (3) but rejects the proposition (2) whereas the eliminativist and deflationist accounts accept the trilemma's propositions (2) and (3) but rejects the proposition (1).

My aim in Part II will be to examine whether observations in childhood are compatible with maximalist accounts of joint action, i.e., to assess whether the findings from developmental psychology studies about young children's cognitive abilities and their engagement in joint action are consistent with the cognitive requirements – high-level cognitive skills – stemming from maximalists accounts.

In order to tackle this question and solve the trilemma introduced by Rakoczy (2018), I will assess each proposition independently in the next three chapters.



## *Chapter 3*

### *Proposition 1: Shared cooperative activities require sophisticated cognitive abilities to represent and reason about mental states*

The aim of this chapter is to assess the first proposition of the trilemma by considering whether shared cooperative activities require sophisticated cognitive abilities. To challenge this proposition, I will adopt two different strategies. First, I will discuss whether (some forms of) joint actions could be interpreted without appealing to the mentalistic stance adopted by maximalists (empiricism) and second, I will discuss whether there could be some minimal theory of mind (dual theory of mind).

#### **3.1 Eliminating the mentalistic stance: reasoning about goals**

The first option consists in considering that young children can perform joint actions without taking a mentalistic stance (Gergely & Csibra, 2003; Perner & Ruffman, 2005; Tollefsen, 2005). Empiricists consider that shared intentions are not needed for joint action and propose instead that a capacity for goal ascription could be sufficient (see Fig. 1).

Gergeley & Csibra (2003) suggest that preverbal infants use a teleological, rather than a mentalistic, interpretation of the observed action. In other words, within an action context, instead of ascribing mental states such as desires, intentions, or beliefs, young children ascribe goals. In ascribing goals, they predict what others will do based on objective reason-giving facts, i.e., on the basis of what would make objective sense for them to do (Perner & Roessler, 2010; Roessler & Perner, 2013). The teleological posture would allow children to establish “[an] explanatory relation among three relevant aspects of current and future reality: the action, the (future) goal state, and the current situational constraints” (Gergely & Csibra, 2003: 289). These three aspects are connected by the rationality principle which provides explanations and predictions about the observed action (Gergely & Csibra,

2003: 289). In fact, Csibra and Gergely (1998) place themselves in the tradition of rationality theory suggesting that infants from around nine months of age understand animate action within a ‘teleological’ scheme of movements (adjusting to given circumstances) in order to reach a goal. Within the rational theory context, their ‘teleological scheme’ is based on the efficiency principle (intended as a rational requirement) i.e., that one ought to employ the optimal means to reach one’s goals<sup>33</sup>.

Almost in the same vein, Perner & Ruffman (2005) argue that infants create three-way associations between actor, object and location, and use behavioral rules to predict how an observed agent will act. For instance, children notice that individuals search for an object based on the location where they have seen it for the last time and that this location is not necessarily congruent with the real location of the object. According to Perner and colleagues (Perner & Esken, 2015; Perner & Roessler, 2010), by the age of 9–18 months, children use teleology since they are able to derive an agent’s reason for an action without concern for the subjective views provided by mental states. In fact, they consider that children’s social competency is based more on inferences about what is likely to happen in a particular situation and on objective reasons for action<sup>34</sup>, compared with inferences about agents’ mental states (Priewasser, Rafetseder, Gargitter & Perner, 2018). However, Perner and colleagues go further compared to the rationality theory, considering that children are basic teleologists until about four years of age and then use teleology-in-perspective, i.e., they use counterfactual reasoning (for a recent review, notably comparing theory-theory, theory of simulation, rationality theory and Perner and colleagues’ conception of teleology, see Perner et al., 2018).

In the case of shared cooperative activities, it may be the case that children would use teleological reasoning in order to ascribe goals and thus predict their co-agents through the use of this “three-way association” together with behavioral rules.

---

<sup>33</sup> “...reference to a future state is accepted as a teleological explanation (reason) for a behaviour in case it justifies it, i.e., when, given the constraints of reality, the behaviour can be seen as a rational way to bring about the goal state” (Csibra & Gergely 1998: 256)

<sup>34</sup> In Perner and Roessler’s account, teleological reasoning is based on objective facts, perceiving the possibility of a desirable state should give a reason for the agent to make this the goal of their action. However, it could be that a process of goal contagion might be involved by which children take over the other’s goal and consequently act as if it was their own (e.g., Kärtner, Keller, & Chaudhary, 2010; Kenward & Gredebäck, 2013).

### ***3.1.1 Behavioral rules***

Some researchers who defend the nativist position consider that an appeal to behavioral rules is not sufficient to explain why five- and six-month-olds are more surprised when an agent grasps a novel object at the habitual grasp location compared to when an agent grasps the same object from the familiarization trials but in a novel location (Woodward, 1998; Luo & Baillargeon, 2005). Concretely, these researchers consider that this surprise effect could only be explained by attributing to the agent a preference for the object. They suggest that a mentalistic stance is necessary to enable children to attribute this preference for a specific object to the agent.

Nevertheless, several elements suggest that the behavioral rule account defended by empiricists is sufficient in this example. First, it seems that the behavioral rule could explain this surprise effect since during the familiarization trials, children could associate a specific object with the grasp performed by the agent and note that the agent grasps more frequently this object compared to any other object in the same location. In other words, children could use statistical regularities to evaluate the most probable upcoming event. Second, objects may be more cognitively salient for children, and hence more robustly encoded than locations. Third, the notion of emotional valence implied by the concept of preference may not be relevant since, in another context, studies revealed that eighteen-month-olds do not need to perceive an emotional reaction from a victim to understand the teleological structure of the action performed by the assailant (Vaish et al., 2010). It may be that children only focus on the teleological structure of the action. However, note that Perner and colleagues adopt an Aristotelian view of teleology in which value judgments are essential (which is not the case in the rationality theory; see Perner et al., 2018). For instance, in the example mentioned above, children use information collected in the public sphere (in opposition to the private sphere) and use value judgment in their goal ascriptions, in the sense that “the goal is something ‘good’ in some minimal sense of being attractive, desirable, needed, worthwhile having, etc ...: the goal is ‘the good to be achieved’” (Perner et al., 2018: 99). The use of behavioral rules might be sufficient in the case of (some) shared cooperative activities since children can base their predictions on the most probable upcoming events, observing elements in the public sphere and ascribing value judgments.

### ***3.1.2 Mutual responsiveness and common knowledge***

Mutual responsiveness and common knowledge are two main criteria involved in joint action as described by maximalists. In a joint action, agents are responsive to the actions and intentions of their co-agents—mutually responsiveness – and base their predictions on a common knowledge that they share with their co-agent. Mutual responsiveness and common knowledge are typically thought to require higher-order beliefs and more specifically recursive mentalizing<sup>35</sup>. If so, how then could they possibly be accommodated within a teleological framework?

Some researchers get around the problem of recursive mentalizing, by resorting to a less demanding notion of common knowledge or to proxies for common knowledge. Carpenter discusses “*common knowledge, in the sense of what is known or has been experienced together*” (Carpenter, 2009: 383). Tollefsen (2005) considers that joint attention might be an adequate substitute for common knowledge in many simple joint activities, where it would provide a shared perceptual space in which cooperative actions can take place. Within such interpretations, common knowledge could refer to the public sphere (see the sub-section above), combining observations from a common environment with value judgments both based on elements publicly observed. In fact, from early on, children begin to interact and play with others. For instance, in their first months of life, they interact through proto-conversations notably with their caregivers; in their first year of life, they coordinate to build block towers, flip through a book together and they can engage in pretend-play (e.g., pretend that a banana is a cellphone) from eighteen months of age. In these situations, children tend to show mutual responsiveness and the ability to use basic common knowledge even though it does not amount to common knowledge in the full sense demanded by maximalists. Some studies showed that fourteen-month-olds use common knowledge established in a previous interaction more than their own individual experience (Liebal et al., 2009) and that they can even integrate others’ ignorance (Liszkowski, Carpenter & Tomasello, 2008) such as elements observed during the interaction to understand an ambiguous discourse (Moll et al., 2008). Regarding mutual responsiveness, young

---

<sup>35</sup> For instance, common knowledge is classically described as a hierarchy of ‘A knows that B knows that ... knows that x’ propositions (Lewis, 1969). For more details, see chapter 1 and footnote 4 on p. 24.

children are able during proto-conversations to synchronize or to take turns with adults (Trevarthen, 1980).

Some researchers consider that children start acting together from two years of age (Brownell & Carriger, 1990, 1991; Brownell, 2011). For instance Brownell notes that in her studies with Carriger:

24- and 30-month olds were able to predict their partner's actions, movements, locations and to adjust their own behavior in anticipation, such as positioning themselves at the opposite end of the task just as the partner moved to one end. Thirty-month-old children were also more likely to gesturally direct or vocally command their partner to engage in a relevant action ahead time, thus recognizing and acknowledging the partner's role and behavior in relation to their own and coordinating both vocalizations and behavior with the peer...

(Brownell, 2011: 202)

These observations highlight the fact that motor and verbal abilities (Warneken & Tomasello, 2007) associated with a better understanding of norms from twenty-seven months of age (Gräfenhain et al., 2009; Hamann, Warneken & Tomasello, 2012) could bring children to another level in terms of the kinds of joint action they can engage in. In other words, in their first years of life, children experiment and learn, through 'shared experiences', how to interpret and react upon contact with others (Tomasello & Rakoczy, 2003). Such experiences may lead to the emergence of some kinds of joint actions.

### **3.2 The existence of a minimal theory of mind?**

The second option consists in considering a deflationist version of shared intentionality. Inspired by other dual models in other domains (see Carey, 2009) and by measures done with adults, researchers defending the dual model of theory of mind posit the existence of two independent cognitive systems: an *implicit system that is* efficient, unconscious and inflexible and an *explicit system* that is slow, costly in terms of cognitive resources, conscious and flexible (Surtees, Butterfill & Apperly, 2012).

#### ***3.2.1 Implicit and explicit systems: from minimal to robust theories of mind***

The *explicit system* relies on a robust theory of mind that involves the representation of mental states understood as propositional attitudes and develops with language, executive functions and knowledge. The *implicit system* relies on a *minimal* theory of mind which involves tracking others' mental states (perceptions, beliefs, knowledge states) understood as relational states rather than in terms of propositional attitudes. These relational states, tracked by the implicit system, reflect the agent's sensitivity to others' engagement regarding objects or events. In studies, such sensitivity is revealed by spontaneous or indirect measures such as anticipatory eye gaze or looking time measures. According to Butterfill and Apperly (2013), several theoretical principles govern the tracking of these relational states. The first principle concerns goal-directed actions and states that bodily movements form units which are directed to goals. It rests on the idea that a component of the minimal theory of mind is the ability to track the functions of things (enabling an observer to link some bodily movements to the goals to which they are directed) and that "representing goal-directed action as we have characterized it does not require representing representations. It only requires representing outcomes as functions of bodily movements" (ibid, 614). For instance, learning by imitation could be considered as an attempt at reproducing an action necessary to achieve a goal. The second principle states that an agent cannot goal-directedly act on an object unless he/she has encountered<sup>36</sup> it. If an outcome involves a particular object and the agent has not encountered that object, then that outcome cannot be a goal of her actions. For instance, "the ability to track perceptions in the ways scrub-jays do could involve representing encounterings only" (ibid, 615) rather than the perceptual states per se of the observed agent. In the same manner, in the presence of a dominant competitor, chimpanzees only retrieve food when the dominant is not informed about its location (Hare et al., 2001). The third principle concerns the conditions of success of the goal-directed actions', where *correct registration* is a condition for the success of goal-directed action (with a goal that specifies an object). Butterfill and Apperly (2013) define registration as "a relation between an individual, an object and a location" such "that an individual registers an object at a location if and only if she most recently encountered it at that

---

<sup>36</sup> "... representing encounterings may differ from representing perceptions in that it does not require any kind of sensitivity to constitutive links with reasons, knowledge or informational states and nor does it require understanding anything about appearances, modalities or the possibility of illusion" (ibid, 617).

location” (ibid, 617). A registration is in turn correct when the object is in that location. Thus, to successfully perform a goal-directed action – with a goal that specifies an object – the agent must correctly register that object in a particular way. The notion of registration can be applied bi-directionally by individuals, i.e., in terms of prediction and of inference. In the case of prediction, an observer can predict whether an agent will successfully perform a goal-directed action based on his/her ability to detect the correctness of the agent’s registration. In the case of inference, an observer can infer whether an agent has a correct registration of the object’s location based on observing the agent successfully performing a goal-directed action. For instance, twelve-month-olds point to inform others based on their goals and ignorance (Liszkowski, Carpenter & Tomasello, 2008). This example shows that the gesture could be the sign that infants understand that a correct registration is a condition of successful action and that pointing is a way to ensure that correct registration. Then, Butterfill and Apperly (2013) offer to think of registration not as a success condition but rather as a causal factor for goal-directed actions, yielding a fourth principle that states that “when an agent performs a goal-directed action with a goal that specifies a particular object, the agent will act as if the object were in the location she registers it in” (ibid, 619). They consider that if registration is a causal factor influencing action, it could serve as a proxy for false belief. For instance, when an agent acts while having an incorrect registration of the object’s location (object toward which the agent’s goal-directed action is directed), the observer could predict that she will not be successful in her goal-directed action. As Butterfill and Apperly (2013) put it:

[A] minimal theory of mind involves representing goals to which actions are directed, encounterings and registrations, none of which are representations...[T]he construction describes a minimally elaborate form of theory of mind cognition, one that does not entail cognitive and conceptual demands associated with representing perceptions, knowledge states and beliefs as such but is capable of grounding theory of mind abilities that generalize across goal-directed actions and are sufficient for systematic success on some false belief tasks.

(Butterfill and Apperly, 2013: 622)

Butterfill and Apperly (2013) compare the minimal theory of mind supported by the implicit system with level-one perspective-taking abilities i.e., abilities to understand that an object visible for a person is not necessarily visible for another person (Flavell, Everett et al., 1981). As a minimal theory of mind involves representing some actions' goals, encounterings and registrations, it implies that one understands that an object visible for one person is not necessarily visible for another person. For instance, one must understand that an agent who does not encounter an object cannot form a goal-directed action involving that object. Let us consider two individuals, François and Gérard, facing each other with two balloons placed in-between them and one extra balloon placed behind Gérard; François must understand that Gérard does not encounter the balloon placed behind his back. Researchers also compare the robust theory of mind supported by the explicit system with level-two perspective-taking abilities, i.e., abilities to understand that an object visible for two distinct persons can lead to different visual experiences if they see the object from different viewpoints. Consider again François and Gérard, facing each other, this time with the digit '6' on the ground between them; if François sees a '6', he must understand that Thomas sees a '9'.

In fact, these two systems can be characterized and tested based on the *aspectual* character of the representations they yield. As Oktay-Gür, Schulz and Rakoczy (2018) put it: “mental (and linguistic) representations are aspectual in the sense that agents represent objects (e.g., Clark Kent, who in fact is Superman) and situations always only under some aspects (e.g., “Clark Kent”) and not under others (e.g., “Superman”)” (ibid, 2). The *implicit system*, restricted to the representation of relational attitudes such as registering an event, does not enable agents to make such fine-grained distinctions regarding the aspects under which an agent has encountered an object. The *explicit system* represents perceptions, beliefs, and other propositional attitudes as such and these representations are inherently aspectual. In fact, ascribing a belief about an object to an agent would depend on the aspects under which the agent subjectively represents the object in question. In a series of studies, Low and colleagues tested adults and children (three and four years of age) in implicit and explicit false belief tasks, some of them involving aspectual characteristics (Low, Drummond, Walmsley, & Wang, 2014; Low & Watts, 2013; Wang, Hadi, & Low, 2015). For instance, in Low and Watts (2013), subjects

participated in two tests – a location and an identity tests – each test consisting in four familiarization trials and one belief-induction trial. In both tests, researchers used an indirect measure, i.e., gaze direction (interpreted as anticipatory looking), at the end of the trials and a direct measure, i.e., verbal prediction (after asking “Which box will the woman look in?”). In the location test, participants were familiarized to a task in which they viewed a female actor behind a screen with two windows, each window with a box underneath. A puppet monkey placed a slice of watermelon in the left-side box, closed the lid, and left. The actor then reached through the left-side window, retrieved the watermelon from the box, and smiled. In the belief-induction trial, the puppet reappeared, placed a ball in the left-side box, and left. The actor turned away from the scene to answer a phone call; while her back was to the participants, the puppet reappeared, transferred the ball to the other box, and exited. The actor then turned around, and the windows illuminated as a beep sounded. In the identity test, participants were familiarized to a task in which they saw two boxes, which were lifted to reveal a red boat under the left-side box and a blue boat under the right-side box. The blue boat traveled to the left-side box, and then the red boat traveled to the right-side box. After two windows above the boxes lit up and a beep sounded, the actor reached through the left-side window, retrieved the blue boat to show he had a preference for blue, and smiled. In the belief-induction trial, the experimenters used a dog-robot toy with one red side and one blue side. The boxes were lifted to reveal a dog-robot toy, with its red side facing participants and left side facing the actor, underneath the left-side box. There was nothing underneath the right-side box. The actor watched the dog robot travel to the right-side box, where, in a recessed chamber visible only to the participants, the dog robot spun to reveal its blue side, spun again to reveal its red side, and repeated the spins. Finally, with its blue side facing participants and red side facing the actor, the dog robot traveled to the left-side box, and the windows were illuminated as a beep sounded. In the location test, the indirect measure revealed that all of the 4-year-olds and the majority of 3-year-olds (94%) and adults (90%) showed anticipatory looks toward the correct location whereas the direct measure revealed an increase in verbal prediction accuracy as a function of age (3-year-olds: 31%; 4-year-olds: 75%; adults: 100%). In the identity test, the indirect measure revealed that the majority of participants first looked to the wrong location (only 6% of 3-year-olds, 6% of 4-year-olds, and 25% of adults showed anticipatory looks toward the correct location) whereas the direct

measure also revealed an increase in verbal prediction accuracy as a function of age (3-year-olds: 13%; 4-year-olds: 56%; adults: 95%).

This study and other studies mentioned above showed that under four years of age, children consistently failed both in non-aspectual (location test) and aspectual (identity test) versions of *explicit* false-belief tasks (direct measure: verbal-predictions) whereas 4-year-olds and adults consistently mastered both tests. 3-year-olds failed at verbal-predictions in both location and identity tests whereas 4-year-olds and adults mastered verbal-predictions in both tests. However, in *implicit* false-belief tasks (indirect measure: anticipatory look), all age groups showed signs of tracking the agent's belief in the non-aspectual change-of-location false-belief task (location test), but not in the aspectual task version (identity test). All age groups both mastered the anticipatory look in the location test and failed in the identity test.

These results suggest a clear “signature limit” in the early developing implicit system (Apperly & Butterfill, 2009) since the highest significant dissociation between implicit and explicit measures show reversed patterns in the location and identity tests. All age groups mastered anticipatory look toward the correct location in the location test (with a very significant dissociation between measures for 3-year-olds) but failed in the identity test (with a very significant dissociation between measures for 4-year-olds and adults).

### ***3.2.2 Signature limits, goal-tracking and motor system***

Apperly and Butterfill (2009) introduced the notion of “*signature limits*” to refer to the distinctive limitations on what a given efficient system can achieve compared to a more flexible system and proposed that it was possible to tell which of a flexible or an efficient system is being used in a given task by examining whether or not performance is subject to signature limits. Several studies have tested the existence of two systems for processing beliefs with adults and three- to eleven-year-olds (e.g., Surtees et al., 2012; Low & Watts, 2013; Milward et al., 2014).

In a series of three visual perspective-taking experiments, adult participants were asked to judge their own or someone else's visual perspective in situations where both perspectives were either the same or different (Samson et al., 2010). Concretely, either both could see the same objects (consistent

condition) or the other person could see a smaller number of objects (inconsistent condition) compared with the participants themselves. This series of experiments revealed that participants could not easily ignore the other person's perspective when making self-perspective judgments which suggests that the computation of what someone else sees is done very rapidly and effortlessly. In other words, the implicit system automatically computes the other person's perspective (Samson et al., 2010) revealing interference effects (participants were slower and less accurate in making self-perspective judgments in the inconsistent than in the consistent condition; see chapter 2 for more details about interference effects). Samson and colleagues (2010) experiments support the existence of an implicit system and explain the absence of an explicit system in three-year-old children by a lack of cognitive resources necessary for this system and by the inflexibility of the minimal theory of mind (as a signature limit). The implicit system would be efficient in infants, children, and adults, using a minimal theory of mind, which enables belief-like states to be tracked. The explicit system would develop later and use a canonical model of theory of mind, incorporating propositional attitudes (Low, Apperly, Butterfill & Rakoczy, 2016).

Following these results and the dual system of theory of mind theory, a body of research focused on these belief-like states studying the relationship between motor and mindreading processes in goal ascription. Previous research indicates that the motor system and its planning-like processes<sup>37</sup> can generate expectations about agents' goals based on various facts about the agent's actual environment (such as object visibility) but also based on non-actual environments (e.g., in motor imagery). Against this background, Zani, Butterfill and Low (2020) reasoned that when observing an agent, the observer's motor system may generate expectations about the agent's goal based not only on facts about the actual environment but also on facts about the environment as specified by the agent's belief or belief-like state and, thus, that goal ascriptions achieved by the motor system in the form of motor representations may track a goal-relevant false belief. In their study, Zani, Butterfill and Low

---

<sup>37</sup> A body of research suggests that the neural representation of motor imagery and action observation is similar to that of motor planning and execution (Szameitat, Shen and Sterr, 2007; Filimon, Nelson, Hagler and Sereno, 2007), which leads researchers to focus on the relation between motor representations and planning-like processes.

(2020) used an interactive helping scenario<sup>38</sup> and showed that participants' early mediolateral motor activity (evidenced by leftwards–rightwards leaning on a balance board) and anticipatory looking foreshadowed the agent's belief-based action preparation. Their results confirmed this conjecture, showing that participants formed spontaneous motor representations of the belief-based action of the observed agent, leaning towards the empty box in the false belief condition and towards the full box in the true belief condition. In other words, their results showed for the first time that an implicit, motorically based, goal ascription system can accommodate cases where belief-tracking informs goal ascription, thus showing the existence of a relation between two implicit systems: a goal-tracking and a belief-tracking system.

Interestingly, the measures of participants' leaning fit with models suggesting that motor representations may be able to remap the agent's allocentric frame of reference into subjects' own egocentric frame of reference (Oh, Braun, Reggia and Gentili, 2019). I will consider in more detail the relationship between motor processes and perspective-taking abilities in chapter 7.

A recent study by Wiesmann, Friederici, Singer and Steinbeis (2020) tracked the existence of two distinctive neural networks, one network including the supramarginal gyrus and dorsal prefrontal cortex and the other the temporoparietal junction, middle temporal gyrus, precuneus, and medial prefrontal, respectively involved in implicit and in explicit false-belief tasks. The results revealed:

[A] double dissociation in that all of the reported effects correlated only with one task type but not the other with no overlap between the two networks. Furthermore, the effects found for the implicit ToM task were independent of performance on the explicit ToM tasks and vice versa. These findings clearly support independent processes involved in implicit and explicit ToM tasks—mature ToM is measured by the explicit tasks and only develops late in

---

<sup>38</sup> In the classical version of the helping task (Buttelmann, Carpenter and Tomasello, 2009), participants (18-month-olds and 2.5-year-olds) observed an agent store an object inside box A. In the false belief (FB) condition, while the agent was absent, an experimenter transferred the object into box B and then shut both boxes. In the true belief (TB) condition, the agent saw the experimenter move the object to box B. In both conditions, the agent approached the now-empty box (box A) and unsuccessfully tried to open it. Participants' final helping action suggested some sensitivity to the agent's belief about the content of the boxes: children opened the now-full box (box B) for the agent in the FB condition, and they opened the now-empty box for the agent in the TB condition (see also Priewasser, Rafetseder, Gargitter and Perner, 2018: their findings tended to indicate that children's performance may be driven by tracking another's ignorance rather another's FB).

preschool age, whereas looking behavior in the implicit ToM tasks is driven by independent earlier-developing processes.

(Weismann, Friederici, Singer and Steinbeis, 2020: 6931)

Researchers found no correlation with other developmental skills such as language, executive function, working memory, and general intelligence. In addition, while in implicit false-belief tasks young (three-year-old) children only recruited their implicit theory of mind neural network, adults also recruited their right temporo-parietal junction, which is part of the explicit theory of mind neural network. In other words, as theoretically introduced by Apperly (2013: 14), “adults automatis[e] some of their effortful mindreading abilities”, meaning that the two networks are not mutually exclusive.

To sum up, the dual model of theory of mind posits the existence of two independent cognitive systems –implicit and explicit – which have specific characteristics. The implicit system (efficient, unconscious and inflexible) would be present at an early age (as a minimal theory of mind) and would enable individuals to track belief-like states. More specifically, the implicit system would combine implicit goal- and belief-tracking processes since adults’ motorically grounded expectations of an observed agent’s action seem to be modulated by a belief-like tracking system (Zani, Butterfill & Low, 2020). The explicit system (slow, costly in terms of cognitive resources, conscious and flexible) would develop later (as a canonical theory of mind) and would enable individuals to track propositional attitudes. The existence of two independent systems tends to be confirmed by recent studies which found that two distinctive neural networks are involved in implicit versus explicit false-belief tasks (Weismann, Friederici, Singer and Steinbeis, 2020). Now, how can be this dual model of theory of mind integrated in the joint action framework?

### ***3.2.3 A minimal theory of mind in joint action?***

Considering these implicit and explicit systems, one hypothesis could be that an ‘implicit’ understanding of goals and beliefs would enable adults and children to perform joint action without the use or the presence of high-level cognitive abilities. Within this framework, researchers consider that the element necessary to enable the performance of a joint action is a common goal rather than a shared intention (Butterfill, 2012). A common goal is described as (i) a distributive goal (each agent

individually formed this goal), (ii) recruiting action coordination (each agent has a motor representation of the goal, is inclined to modify her motor plans to reach the goal and think that she will contribute to its success), (iii) with a coordination suitable for the realization of the expected outcome.

According to the simple account proposed by Butterfill, for some joint actions it is sufficient that agents share goals, where sharing goals require only an understanding of goal-directed actions and their common effects, and coordinate their activities (but not their plans). In other words, co-agents would not need to form any strong form of intention, or to incorporate their co-agent's intention into their own but only to motorically coordinate in order to favor the success of the action. A common goal associated with the use of their implicit system would allow the performance of some sorts of joint actions through the use of distributed motor representations. Moreover, based on Zani, Butterfill and Low (2020), it could be that motor representations can be involved both in goal ascription (modulated by belief-tracking) and in the facilitation of the coordination between agents, making joint actions more spontaneous and interactive. For instance, two adults could spontaneously form a common goal and coordinate without the need of representing each other's intentions; they would rather use their implicit system. In the case of children, the lack of strong cognitive resources – necessary for advanced planning and for explicit representations of their co-agents' mental attitudes – would set a limit to the kind of joint actions they can successfully engage in (Brownell et al., 2006). Nevertheless, Butterfill's simple account gives us a sufficient basis to consider that children who start playing together around one year of age perform at least minimal forms of joint actions.

To recap, the main differences between the dual theory of mind and the teleological approach is that while the teleological approach proposes that children (and sometimes adults) reason about the about agents' relations to factual situations as a proxy for reasoning about mental states, dual theory of mind proponents consider that they use a minimal theory of mind that tracks mental states (though without representing them as such) and has distinctive signature limits (e.g, insensitivity to aspectual characteristics of mental representations). In particular, while both theories take it that children are capable of goal ascription, the dual theory considers that motor representations and motor planning are used to generate expectations about goals (and that motor representations might play a role in

understanding others' mental states), whereas the teleological account proposes that expectations about goals can be generated by reasoning about facts, using behavioral rules or a principle of efficiency.

These two theories offer alternatives to the strong high-level cognitive requirements (such as recursive mentalizing) posited by maximalists, namely the existence of a minimal theory of mind and the use of teleological reasoning in the interpretation of others' actions. More specifically, in the case of joint action, agents might use teleological reasoning or a minimal theory of mind in order to ascribe goals to their co-agents to predict their behaviors and adjust their actions accordingly.



## *Chapter 4*

### *Proposition 2: Young children do not have the sophisticated cognitive abilities needed to engage in shared cooperative abilities*

The second proposition in the trilemma is classical in the sense that it relies on developmental studies which suggest that children lack a (robust) theory of mind before four years of age (Wellman et al., 2001). To assess this proposition, we must discuss a set of studies using implicit measures of theory of mind abilities that have been interpreted as showing that very young children are already endowed with robust theory of mind abilities (Leslie, 1999; Onishi & Baillargeon, 2005; Carpenter, 2009). Is this nativist interpretation of experimental data warranted? If so, could younger children avail themselves of their mindreading skills to satisfy two of the main requirements on joint actions highlighted by maximalists – common knowledge and mutual responsiveness?

#### **4.1 A mentalistic approach**

To the classic false-belief tasks involving explicit measures, a body of research added implicit measures such as anticipatory looking (Clements and Perner, 1994; Southgate, Senju and Csibra 2007), violation-of-expectation paradigms (Onishi & Baillargeon, 2005; He, Bolz and Baillargeon, 2011; Träuble, Marinović and Pauen 2010) or interactive tasks (Buttelman et al, 2009; Knudsen & Liszkowski, 2012; Southgate et al, 2010). As we saw earlier, these implicit measures shed light on the fact that children under four years of age do appear to have at least some understanding of others' mental states; but how mentalistic is this understanding and what evidence would support the nativist position?

#### ***4.1.1 Where did the mentalistic approach come from?***

Three main positions about false-belief tasks, which can get organized around their interpretation of implicit and explicit measures, can be distinguished. First, some researchers argue that implicit measures can provide reliable evidence of false-belief attribution in childhood (e.g., Baillargeon et al. 2010; Scott and Baillargeon, 2017). Second, others argue that only explicit measures provide reliable evidence of false-belief attribution in childhood (e.g., Perner and Ruffman 2005; Perner and Roessler 2012). Third, some researchers do not consider that explicit measures provide sufficient evidence of false-belief attribution in childhood (e.g., Oktay-Gür and Rakoczy, 2017). So, what arguments can advocates of the mentalistic approach adduce? In 2005, Onishi and Baillargeon tested fourteen-month-olds in a violation-of-expectation version of the false belief task. As in a classic false-belief task, children observed a number of events involving an agent who manifestly has a true or a false belief about the location of an object. The children were not asked to provide a verbal answer to the question where the agent would look, rather they simply observed the agent reach for the correct or the wrong location and the experimenters measured the duration of looking time, Onishi and Baillargeon expected that children would look longer when there was a contradiction between the agent's belief and her action (for instance, when the agent manifestly had a false belief but nevertheless acted as if she had a true belief or vice versa). Fourteen-month-olds did indeed look longer when the agent manifested a contradiction between her belief and subsequent action of reaching. Based on their findings, several variants of this experiment were created, focusing on the agent's belief about the contents of a container rather than about the location of an object (He, Bolz and Baillargeon 2011), on expressions of emotion rather than on instrumental actions (Scott 2017), or on the need to infer belief from a story (Scott et al. 2012) or from action (Träuble, Marinović and Pauen 2010). In these different versions of the task, the same set of results was observed, meaning that before four years of age, children's expectations, as measured by preferential looking, are sensitive to what others believe.

Buttelmann and colleagues (2009) complemented their violation-of-expectation paradigm with an interactive task. i.e., children were given the possibility to help the agent. In order to help the agent, children needed to use the agent's epistemic state (true or false belief) to infer a goal based on the

agent's behavior. In this study, two opaque boxes with an object placed in one of them were set on a table and two conditions were run: the agent's belief about the object location was congruent (true belief) or incongruent (false belief) with the actual position of the object. In each condition, the agent tried to open one of the two boxes (the empty box) and researchers recorded which box the children would help the agent open. Researchers predicted that in the false belief condition, children would help the agent find the object and so direct their helping behavior toward the box containing the object whereas in the true belief condition, if the agent tries to open the empty box, it means that the agent does not intend to retrieve the object, so the children should direct their helping behaviors toward the empty box. Results revealed that in the false belief condition, children reliably helped the agent find the toy by opening the non-empty box (which the agent was not trying to open) whereas in the true belief condition, the children reliably helped the agent open the empty box<sup>39</sup>.

In a more recent study, Buttelmann and Kovács (2019) tested fourteen-month-olds in a novel eye-tracking-based unexpected-identity task to see whether young children could also predict others' behavior based on their beliefs about an object's identity. As previous studies had shown that, by fourteen months of age, infants seem to be able to represent the duality of a single object (e.g., Cacchione, Schaub, & Rakoczy, 2013), Buttelmann and Kovacs combined a violation of expectation and an anticipatory looking paradigm. In this study, children observed an agent who either knew about the two identities of a deceptive object (true belief condition, e.g., a spoon that looks like a frog or a pen that looks like a cat) or knew only the appearance of the object (false belief condition, e.g., the object being a frog or a cat). Once familiarized with the scenario, children were assigned to the true or the false belief condition; their anticipatory looking was tested in the first four test trials and their duration of looking time was tested in the fourth and fifth test trials (in the last two test trials, researchers used the same deceptive object and presented in a counterbalanced order the agent reaching for the object resembling the apparent identity or the real identity of the deceptive object). Researchers found that children's anticipatory looks were coherent with an understanding of the agent's belief about the deceptive object and that children looked longer at an actor's reach that was

---

<sup>39</sup> Note that the replicability of this study is debated (see for a recent review, Crivello & Poulin-Dubois, 2018).

incongruent with her false belief about the identity of an object compared to a congruent reach. In other words, when the agent represented the deceptive object based on its apparent identity (false belief condition), children predicted that the agent would reach for the object resembling the apparent identity of the deceptive object and in the last two test trials, they looked longer when the agent reached for the object resembling the real identity of the deceptive object. These results reveal that when young children represent dual identities, they can integrate them in false belief attributions (results were not significant in the true belief condition) and use them to make behavioral predictions. Researchers consider these results as a support for young children's rich theory-of-mind abilities, a flexible single mindreading system shared by infants and adults (in the same vein as e.g., Jacob, 2020).

As highlighted previously, the results of these studies (among others) motivated several researchers to create more designs and discuss whether children have strong cognitive abilities before four years of age. Now, what are the main characteristics of early false-belief understanding according to nativists? How can they explain the success gap between implicit and explicit measures?

#### ***4.1.2 Mentalistic approach as a one-account system***

Nativists' results suggest that early false-belief understanding is flexible, context-sensitive, and integrated with other information available in cognitive processes such as memory and language comprehension (for a recent review, see Scott et al., 2017). According to nativists, this early false-belief understanding is guided by a single mentalistic system (Leslie, 2005) emerging early in infancy (e.g., Baillargeon et al., 2016; Carruthers, 2016, 2018; Scott and Baillargeon, 2017) but success on false belief tasks also depends on the development of executive function skills including working memory and inhibitory control (Carlson and Moses, 2001; Devine and Hughes, 2014; Duh et al., 2016). In other words, nativists argue that mentalizing capacities are present from early on but are not fully expressed due to the immaturity of executive functions. Executive function skills might be involved notably in the young children's ability to refresh, update their representations of agent's beliefs based on situational constraints and observations; this ability is part of executive functions.

In sum, nativists defend a one-system account because they consider that the integration of many elements – such as physical reasoning, inferences based on general knowledge, language

understanding skills – into psychological reasoning evolves following the acquisition of new knowledge and the emergence of new cognitive abilities (Carruthers, 2018). As this body of research sheds light on the implication of executive functions in false-belief understanding, it tends to challenge the view that traditional and non-traditional false-belief tasks (tasks using implicit versus explicit measures) tap two distinct mindreading systems, and thus the need for a dual theory of mind.

#### **4.2 A success gap between implicit and explicit measures**

As discussed previously, different paradigms have been developed using different measures in order to test theory of mind in childhood. In fact, many researchers analyze these paradigms trying to identify specific features and to see the extent to which these features affect individuals' performance. In other words, researchers try to elucidate why there is success gap between implicit and explicit measures.

The main strategy used to explain the gap between implicit and explicit measures consists in considering that classic false-belief tasks involve extraneous demands, notably response selection and inhibitory control skills (Leslie and Polizzi, 1998; Leslie, 2000; Leslie, German and Polizzi, 2005; Scott, Baillargeon, Song and Leslie, 2010). For instance, Leslie, German and Polizzi (2005: 49) consider that children have conflicting perspectives as they represent both a true and a false belief at the same time. The process of selecting – ‘selection processor’ – between these perspectives involves inhibitory control skills which would depend on the ‘salience’ of the representational content. In other words, in a classic false belief task, adults (under some pressure) and young children would act as if they did not represent the false belief because they end up ascribing a true belief to the agent – this true belief being more salient. These researchers consider that false-belief tasks involving implicit measures do not require response selection (Scott and Baillargeon 2009; Baillargeon, Scott and He 2010; Scott et al. 2010). More precisely, the existence of many potentially confounding factors in the analysis of the false-belief tasks has been stressed. In particular, it has been suggested that the presence of a verbal question (explicit measure) could confuse children about which perspective to adopt depending on whether the experimenter is addressing them or just thinking out loud (Helming, Strickland, & Jacob, 2014) or place them in a “justification mode” leading them to say where the protagonist should search (Perner & Roessler, 2012). A recent study by Setoh, Scott and colleagues

(2016) revealed that two-and-an-half year-olds can succeed at a traditional false-belief task if the inhibitory control and response-generation demands are reduced. Concretely, reducing the inhibitory control demand involves removing the test object from the scene and reducing the response-generation demands involves asking beforehand two practice questions similar in structure to the test question. In the same direction, several studies have tested the implication of various executive function skills, such as working memory or inhibitory control skills, revealing that their improvement contributes to the children's success in traditional tasks (e.g., Carlson and Moses, 2001; Duh et al., 2016). These skills might be involved notably in the young children's ability to refresh, update the agent's beliefs based on situational constraints and observations; this ability is part of executive functions. According to nativists, the mentalistic understanding of children is then masked by the extraneous demands from the task itself (Carruthers, 2013; Baillargeon et al., 2016).

The hypothesis according to which false-belief tasks involving implicit measures do not require any response selection among perspectives can be explained in different ways. First, it could be that false-belief tasks do not involve any form of belief ascription; they could instead rely on action prediction for instance. Action prediction could explain that children distinguish between situations involving true and false beliefs (e.g., Kovács, Téglás and Endress 2010; Southgate and Vernetti 2014). However, this interpretation seems to be limited in explaining cases in which infants appear to have specific expectations concerning how someone with a false belief will act (e.g., Southgate, Senju and Csibra 2007; Knudsen and Liszkowski 2012a).

Second, it could be that young children or adults in specific contexts (such as under time pressure or with cognitive load) only represent the content of a false belief but not the content of the corresponding true belief (Scott and Baillargeon 2009; Baillargeon, Scott and He 2010; Scott et al. 2010). In fact, previous studies have showed that selection processes only get implicated when several potential targets are represented or attended to (Buetti, Lleras and Moore 2014). In false-belief tasks, there is only one target to attend to and selection processes might not be involved. However, this interpretation does not explain why the content of the false belief would be represented rather than the content of the corresponding true belief. One potential explanation is that children show a form of "negativity bias" – in which negative information is preferentially attended to – oriented toward the

agent's choice (Di Gangi, Farroni & Southgate, 2014; see Vaish, Grossmann, & Woodward, 2008, for a review of the developmental literature). But how can a form of negativity bias explain the exclusive representation of a false belief? Previous studies have shown that children preferentially encode approach over avoidance goals. Indeed, Feiman, Carey & Cushman (2015) showed that while observing another's goals (e.g., approaching vs. avoiding an object over another), seven- and fourteen-month-olds do represent approach goals over avoidance goals, revealing a significant asymmetry between approach versus avoidance and that infants looked significantly longer at violation trials than at expected trials only in the approach condition (where they had been habituated to a hand consistently approaching an object) compared to the avoidance condition (where they had been habituated to a hand consistently avoiding an object), meaning that infants expect that the hand will continue reaching for the same object, again only in the approach condition. In other words, children show a significant reaction when a violation of expectation appears in the approach condition. Relatively to the response selection, false-belief tasks only set an approach context (no avoidance context) and the agent either has a true or a false belief about an object (e.g., location, identity). In the true-belief condition, the agent is aware of all the information about the object. In the false-belief condition, the agent is not aware of all the information about the object and furthermore, she has a false belief about it. The fact that children preferentially encode approach goals and that, in false-belief tasks the agent is hindered in her approach goal (as she has a false-belief about the object she previously showed a preference for), can trigger a form of negativity bias in observers. In other words, this form of negativity bias would stem from the perception of the hindrance created by the agent's false-belief in her approach goal. It could be that the false-belief condition, in which children and adults (in some specific context mentioned above) anticipate a violation of expectation, while triggering a form of negativity bias, lead them to represent only the false belief and not the corresponding true belief. This interpretation could also explain why in the study of Buttelmann and Kovacs (2019), results were significant only in the false belief condition.

Third, it could be that young children exhibit an altercentric bias (Southgate, 2020), i.e., infants take the agent's perspective and by extension represent her false belief without having a competing self-perspective. As Southgate (2020) puts it:

[T]his bias results from a combination of the value that human cognition places on others' attention, and an absence of a competing self-perspective, which would, in older children, create a conflict requiring resolution by Executive Functions. A self-perspective emerges with the development of cognitive self-awareness, sometime in the second year of life, at which point it leads to competition between perspectives.

(Southgate, 2020: 505)

The altercentric hypothesis could also explain the lack of response selection in false-belief tasks with implicit measures but it would also rule out the nativist account. In fact, according to this hypothesis, while children do take the agent's perspective, they do not ascribe any belief to the agent; this is the reason why I will discuss this account separately in chapter 5. I will also consider this hypothesis and its importance for joint action in more detail in chapter 7.

Nativists explain the success gap between false belief tasks using implicit and explicit measures by adducing the presence of extraneous demands for a response selection in the case of explicit measures and, conversely, the absence of response selection demands in implicit measures. In other words, false-belief tasks involving implicit measures do not mask children's mentalistic understanding of false belief as they do not involve extraneous demands. The aim of this subsection was to highlight the different interpretations of this lack of response selection, presenting the point of view of nativists but also introducing the altercentric bias hypothesis, which, as I will argue later, could be a good candidate for explaining the success gap.

### **4.3 Challenging the mentalistic approach?**

Recently, some researchers (Powell, Hobbs, Bardis, & Carey, 2017; Priewasser, Rafetseder, Gargitter, and Perner, 2018) have raised a challenge to this mentalistic approach using a modified version of the classic interactive task from Buttelmann and colleagues (2009). For instance, Priewasser and colleagues (2018) propose three possible factors that could provide teleological reasons for children to show distinctive behaviors in the two conditions without relying on belief reasoning: (i) playing a trick, (ii) ownership and (iii) interest. In the false belief scenario, the experimenter 'sneakily' moves

the toy from one box to the other while the agent is away; this element suggests that (i) the experimenter is playing a trick to the agent (ii) about an object which belongs to the agent (iii) who will be strongly motivated to get it. First, Priewasser and colleagues (2018) tried to replicate the original study. Although they did not replicate the original results in the true belief condition, they replicated them in the false belief condition. Then, they designed a second study introducing a third box in order to test the teleological interpretation i.e., to see whether children show a distinct helping pattern in the two conditions (not based on belief reasoning). Their new false belief condition is identical to the original except that the experimenter tries to open the third box (empty) instead of the box where the toy was located before the transfer.

What is the purpose of this new false belief condition? Researchers expected different patterns of helping behaviors based on the fact that the experimenter tries to open the third box. According to the teleological approach, when the experimenter tries to open the third box, young children should presumably display a tendency to help and open the box where the toy is currently located (without any inference about the experimenter's false belief). According to the mentalistic approach, when the experimenter tries to open the third box, young children should presumably display a tendency to help and open this third box where the toy is not located (the experimenter might want to open the third box for an unknown reason). The results revealed that young children helped the experimenter by directing her to the box with the toy in the new false belief task. Priewasser and colleagues (2018) argue that their account “provide[s] teleological reasons for children to show a distinct helping pattern in the two conditions that are not based on belief reasoning” (71).

In addition, as presented previously, Grosse Wiesmann, Friederici, Singer and Steinbeis (2020) used false-belief tasks with implicit and explicit measures and found a “neural dissociation [which] suggests two systems for reasoning about others' minds—mature verbal ToM that emerges around 4 years of age, whereas nonverbal ToM tasks rely on different earlier-developing possibly social-cognitive processes” (ibid, 6928). The existence of these two distinctive neural patterns is difficult to reconcile with the nativist position since they seem to reveal the existence of two mindreading systems, one involved in the performance of false-belief tasks with implicit measures and accounting

for their success and the other involved in the successful performance of false-belief tasks with explicit measures. This study constitutes yet another challenge for nativists.

Nativism offers a mentalistic interpretation of children's false-belief tasks understanding which is comparable to adults but where the competence of children is masked by task demands (verbal or executive function skills). These researchers defend a one-system account, considering that the integration of many elements into reasoning evolves following the acquisition of new knowledge and the emergence of new cognitive abilities. On this nativist position, children have the mindreading abilities needed to perform joint actions in the maximalist sense. However, this mentalistic approach is challenged by some recent studies using a modified version of the classic interactive task and by studies investigating the neural networks involved in the success of the tasks.



## *Chapter 5*

### *Proposition 3: Children engage in shared cooperative activities*

Challenging the third proposition consists in questioning the “*sharedness*” characteristic of the activities children engage in with others. So far, when discussing the *shared intention hypothesis*, I have mainly focused on the *intention* part, but the *sharedness* characteristic also warrants discussion.

According to the *shared intention hypothesis*, in order to “share” an intention, agents need to attribute certain mental states (intentions, beliefs) to their co-agents and to incorporate the attitudes attributed to co-agents in their own intentions. In chapters 3 and 4, I discussed the kind of representations young children must be able to form to perform joint actions: representing the mental states of their partners for nativists or simply representing their partners as pursuing goals for advocates of teleological reasoning and dual theory of mind. The main question here will be about the way co-agents *share* representations so that they can perform *shared* cooperative activities. In this section, the introduction of two different bodies of research will be key in order to fully embrace this question of children’s engagement in shared cooperative activities. On the one hand, it could be that the activities children engage in with others (and presenting the appearance of shared cooperative activities) could be explained without involving representational abilities or shared representations. To discuss this possibility, I will consider a recent paper by Tyler Burge (2018) in which Burge argues that young children’s success at false belief should not be taken as evidence of belief attribution and can be explained in terms of the same kind of generic action-sensing attribution scheme that biologists use to explain the behavior of snails or ticks. On the other hand, in order to share representations, children should be able to manage several perspectives at the same time which seems to be difficult as presented previously; they might instead only take the other perspective. To discuss this possibility, I will present in more detail the altercentric bias hypothesis (Southgate, 2020) which holds that young children do not maintain both their own and the other perspectives but rather take the other

perspective. Both approaches of this question dispose of the “sharedness” characteristic of the problem by relying either on an action-sensing attribution scheme or on the elimination of the self-perspective.

## **5.1 Would ethology shed new light on our problem?**

Ethologists have documented a wide range of cooperative activities in the animal kingdom, from social carnivores to ants, including of course the study of non-human primates (Drea & Carter, 2009; Stander, 1992; Kuczaj, Winship & Eskelinen, 2015; King et al., 2016; Suchak et al., 2016; Schmidt & Tomasello, 2016). Does this mean that the mindreading abilities issue also applies to animals? Should researchers adapt false-belief tasks to animals?

### ***5.1.1 Burge’s action-sensing attribution scheme***

In a recent paper, Burge (2018) criticizes false-belief tasks and argues that success at these tasks does not constitute evidence of belief attribution<sup>40</sup>. Burge even compares the structure of false-belief tasks with the structure of a tick’s failed action (Burge, 2018: 420). Instead, Burge offers a deflationary account, defending an action-sensing attribution scheme. In this scheme, Burge distinguishes generic from specific internal states that could be attributed to an agent. Generic internal states comprise mental (conscious or representational) and non-mental sub-species. To take two opposite examples: intentional actions performed by human adults, but also young children or animals (notably apes) would be caused by first-order mental representations whereas actions performed by ticks, bacteria or snails would be non-intentional and guided by generic internal non-mental states (lacking aspectuality).

The mental and non-mental states can be either conative (i.e., motivational) or sensory (i.e., informational). In other words, mental conative states such as intentions or desires would have a representational content when non-mental conative states would lack representational content. To illustrate, a non-mental conative state would have a biological function i.e., to provide energy to the agent in order for her to reach her action target. The same applies to sensory states, where Burge draws

---

<sup>40</sup> “False belief tests not only show nothing about attribution of belief. They do not in themselves show attribution of any mental states. All the experiments bear on subjects’ assessing whether an action will meet its target” (Burge, 2018: 418).

a distinction between perception as a mental sensory state and sensing as a non-mental sensory state. Again, to illustrate, a non-mental sensory state would have the biological function to carry information about, or to statistically covary with, environmental features. In other words, non-mental conative and sensory states would have a biological function and as Burge puts it: “Neither action with targets nor sensing the environment entail mentality” (2018: 409). Burge exemplifies non-mental states: a snail crawling toward and eating a leaf caused by a generic internal non-mental conative state; bacteria or mollusks sensing light (e.g., for predators) and ticks sensing heat (e.g., to track a human arm) thanks to generic internal non-mental sensory states carrying information about light and heat. This last example could be applied to a human sensing heat from the stove and automatically distancing her hand from the hot surface. In fact, Burge’s aim is to show that it is possible to explain our observations regarding children’s behavior in false belief tasks without involving appealing to mental state attributions.

### ***5.1.2 Are false-belief tasks good markers of belief attribution?***

Burge considers that false-belief tasks do not show that young children attribute beliefs defined as *propositional attitudes*. Regarding the implicit measures, Burges adds:

The fact that children anticipate where an individual will look, even when the individual will not look where the children would look, *could not* show anything about whether the children attribute the attitude *belief*...False belief tests not only show nothing about attribution of belief. They do not in themselves show attribution of any mental states.

(Burge, 2018: 418)

For Burge, the results of these tasks can be explained in terms of a generic action-sensing attribution scheme, where what children attribute to others are simply generic states of sensory registration.

In his paper, Burge criticizes all the positions presented in the chapter 3 and 4 –empiricism, dual theory of mind, nativism – about our trilemma and offers as an alternative his action-sensing attribution scheme. For instance, he discusses a “false perception” study by Song and Baillargeon (2008) in which young children were given evidence that an agent had a preference for a doll with blue hair over a skunk. Young children saw that the skunk was placed in an opaque box with an

apparent tuft of blue hair while the doll was placed in a plain opaque box in the presence (true belief) or absence (false belief) of the agent. Finally, young children saw the agent reach for one box. The results revealed that 14.5-month-olds looked reliably longer when the agent reached for the box with the tuft of blue hair than when the agent reached for the plain box in the true belief condition and conversely, showed the opposite pattern in the false belief condition. Burge (2018) argues that young children only need to track the agent's target (the doll with blue hair) thanks to their ability to sense one of its significant properties<sup>41</sup>.

However, it seems that Burge's position confronts several difficulties. In a recent paper, Jacob (2020) discusses three challenges to Burge's position: the true-belief, the helping false-belief tests and the dissociation between false-belief and ignorance challenges. In the true-belief challenge, Jacob (2020) highlights the fact that young children show opposite patterns in the true and the false belief conditions. As Jacob puts it:

...on Burge's interpretation, infants in the true-belief condition should also presumably attribute to the agent the generic capacity to track the doll with blue hair by sensing the property of blue (hair) instantiated by the tuft of blue hair attached to the box containing the skunk. As a result, in the true-belief condition, infants should be torn between attributing to the agent the knowledge (or, as Burge would say, the sensory retention of the information) that the doll with blue hair is hidden in the plain box and attributing to her the non-mental capacity to track the doll with blue hair by sensing the currently visible tuft of blue hair attached to the other box.

(Jacob, 2020: 10)

In the dissociation between false-belief and ignorance challenge, since Burge's account does only presuppose that an agent is either provided with the opportunity for sensing a property or not, it is hard for him to explain results from experimental studies which show different patterns in true belief, false belief and ignorance conditions (e.g., Scott and Baillargeon, 2009).

---

<sup>41</sup> "...the evidence is explained by taking the infant to attribute a generic capacity to chain sensing of the doll with sensing of blueness" (Burge, 2018: 420-421).

In the case of helping false-belief tests (see footnote n°7 for a description of Buttelmann et al., 2009), Burge argues that children's behaviors can be explained by his non-mentalistic action-sensing scheme "together with the assumption that the child wants to help the actor reach the actor's target" (Burge, 2018: 421). But what would motivate children to help the agent? The spontaneous helping behaviors children display could result from an altruistic motivation (e.g., Warneken & Tomasello, 2006), a motivation intrinsic to participating in a joint action, an aversion to others' distress (e.g., if others are getting nervous or upset about not achieving a goal) or a form of 'goal-alignment'<sup>42</sup> (that is, when children identify an agent's goal, they take up that goal as their own; see Paulus, 2014). These alternative explanations of children's motivation in spontaneous helping situations involve either an understanding or an anticipation of a subsequent affective state in the child (e.g., goal fulfillment in itself or helping the agent to reach his/her target could give rise to positive feelings, see Aknin, Van de Vondervoort & Hamlin, 2018) or in the agent (based on the satisfaction or frustration of their goal, e.g., finding a toy).

Regarding the agent's affective states, Burge argues that the response to behaviors associated with others' degree of satisfaction or frustration results in the attribution of generic affect (not emotion) which leads to an issue. As mentioned by Jacob (2020), such a non-mental emotional state could be ascribed to the agent on the basis on the execution (*not in anticipation*) of the agent's action. In paradigms where an experimenter repeatedly attempts but fails to produce a target action on an object, it could be that young children ascribe a goal frustration to the agent based on the repetition of the target action (e.g., Reschke et al., 2017). However, it is not the case in paradigms where an experimenter (within a neutral condition, i.e., no affect, no verbal help-seeking or eye contact) accidentally drops an object out-of-reach on the floor; in this case, from eighteen months of age, infants still spontaneously help the experimenter reach his/her target (Warneken & Tomasello, 2007).

---

<sup>42</sup> More precisely, this goal-alignment could result from different treatments of the goal represented. First, it could result from a 'goal-contagion', meaning that children's representation of the observed goal may "prime" the performance of an action with the same goal (Kenward and Gredebäck, 2013). This interpretation generates a false prediction i.e., that infants end up in competition with the agents whose goals they take up. Second, it could result from a 'goal-slippage', meaning that the identification of the agent's goal elicits a motivation to complete the action and to achieve the goal whereby the goal slips from perception into action (Michael et al., 2016). In this 'goal slippage' treatment, goals are represented in an agent-neutral manner, i.e., irrespectively of *who* desires them or attempt to bring them about (for a recent review, see Michael & Székely, 2017).

Regarding the child affective states, in order to help the agent (e.g., by altruism or goal-alignment), children need to re-enact or complement the agent's goal which, first, rests on their ability to identify and represent the agent's goal (as an outcome event toward which the agent's movements are directed) and second, depends on the effectiveness of the child's action guidance by this representation to bring about the event outcome. For instance, if an experimenter accidentally drops a clothespin out-of-reach on the floor, children must identify and represent the experimenter's goal (reach for the clothespin) when bending in the direction of the object and reach for it themselves before giving it back to the experimenter. According to Burge (2018), infants would ascribe a non-mental conative state of achieving a goal to the observed agent. Remember that, in Burge's conception, non-mental conative states lack representational content (e.g., they would have a biological function i.e., to provide energy to the agent in order for her to reach her action target). However, there does not seem to be a sufficient structure for the goal to 'slip' from observation of the agent (the experimenter's goal is identified) to the child's own motivation (the child has the motivational state to help the experimenter's goal performance) so that the child helps the agent reach his/her target.

Another comment that could be raised up with Burge's account is that it seems a bit off to compare a tick, a snail or a mollusk facing their environment or their source of food with paradigms such as false-belief tasks which involve observing a peer's action. It is really important to observe other species and to see whether they can shed light on some unanswered questions; at the same time, it is key to compare analog situations. For instance, it could be interesting to discuss bird flocks, fish schools or cycling pelotons (packs of bicycle racers). In a recent study, Belden and colleagues (2019) analyzed the structure of cycling pelotons as a network where individuals sense and react to stimuli within close spatio-temporal limits. The structure of the network is diamond shaped which works thanks to the visual sensory system with cyclists aligning in patterns within an arc (more or less thirty degrees) corresponding to the human near-peripheral visual field. These behaviors could be related to generic internal non-mental sensory states as described by Burge (2018). Interestingly, researchers noticed that, at the end of the race, the group structure changes (lengthening the diamond shape) suggesting a narrowing of the used field of vision modulated by arousal-dependent neurological

effects. Belden and colleagues (2019) draw an analogy between cycling pelotons and elastic solid mechanics models; they hope that their mechanistic description will help dissect collective behaviors in animals. In fact, Burge's scheme seems to apply and be more adequate to the cycling pelotons example in comparison with false-belief tasks analysis. So now, could Burge's deflationary approach be applied to several animals interacting together and what could it bring to the definition of joint action?

### ***5.1.3 Collective behaviors in social animals: a case of joint action?***

In the animal kingdom, many social animals engage in collective behaviors. Here, I will focus on collective behaviors in ants' communities. For instance, army ants construct bivouacs (Franks, 1989) and build bridges out of their bodies to cross gaps along a foraging trail (Reid et al., 2015) and fire ants gather together to form rafts and towers when their habitat floods (Mlot et al., 2011, 2012; Phonekeo et al., 2017). So does Burge's account applies to ants and can we consider these collective behaviors as joint actions?

A growing body of research studies ants' behaviors and, more specifically, in the case of army ants, their ability to build bridges. In two recent studies, researchers discovered that army ants adjust living bridges according to a cost–benefit trade-off (Reid et al., 2015) in an unplanned manner but apparently following two main rules (Graham et al., 2017). Researchers observed bridge building by putting gaps in the colony trail: when the first line ants come to a gap, they slow down while the rest of the colony keeps going (12 centimeters per second). At this point, the second line ants come trampling over the first line ants' backs. As soon as the first line ants feel other ants on their back, they freeze (first rule); this process repeats in the next ants until the gap is filled in so that the rest of the colony walks over it. Conversely, when the traffic over their back dips below a threshold, the ants unfreeze (second rule) and follow the colony. Researchers thus hypothesize the existence of a “bridging” algorithm. The two rules highlighted by researchers seem to accord with Burge's sensing-action attribution scheme since ants *sense* their environment and their back and this sensing determines whether they freeze or unfreeze (which is related to other ants' behaviors). Going further in this direction, fire ants such as ants of the genus *Formica selysi* form rafts in flooding situations and

researchers study the structure of such rafts (e.g., position of workers, brood). While workers in large rafts tend to be more mobile in fire ants colonies (Adams et al., 2011), changing from linear to diffusive motion when the raft size increased (Mlot et al. 2012), workers in smaller rafts position themselves according to their individual specializations and based on previous experience (Avril et al., 2016). Ants seem to show a form of memory and position themselves depending on their specialization.

Based on these observations, it seems that ants show (i) a form of mutual responsiveness since they react to each other based on their sensing abilities, (ii) a form of social learning (e.g., the use of previous experience in building rafts based on the preservation of the brood), (iii) a form of commitment (e.g., social or behavioral rules: freezing and unfreezing). Mutual responsiveness and commitment are evoked by maximalists as components required for the existence of a joint action (see chapter 1). However, in the description of the fire ants' behaviors – freezing and unfreezing – it seems that there is no common or even distributive goal between ants. It seems that ants follow some behavioral rules that researchers study in order to generate a model (e.g., based on individual decision-making, see Sakiyama, 2017). So, if Burge's sensing-action attribution scheme could be used to describe ants' behaviors, it is not sufficient to qualify these collective behaviors as minimal joint actions. It could be that we tend ascribe intentions to ants as a form of anthropomorphism or that, as in Searle's dancers example (Searle 1990), from an external point of view, nothing could tell us whether ants' behavior relies on a collective intention; however, the mechanistic analysis of their behaviors seems to confirm the idea that ants act individually, not collectively.

Burge's sensing-action attribution scheme offers an alternative to nativism, empiricism and dual theory of mind theory, excluding the mental aspect of nativists and dual theory of mind theorists and, arguing against empiricists (notably about goal-attribution and behavioral rules). However, this scheme faces several obstacles notably with false-belief tasks (even though Burge denies the relevance of these tasks in measuring any sort of mentality). I also considered that Burge's comparisons were not on the same level – comparing snails or ticks interacting with their environment with humans analyzing others – and introduced a discussion about collective behaviors in social insects. Ants are social animals that show sophisticated collective behaviors (e.g., building bridges, rafts). I proposed to

apply Burge's scheme to fire ants. From this analysis, I concluded that, even though this scheme seems relevant to fire ants, it is not sufficient to qualify fire ants' collective behaviors as a sort of minimal joint action. In addition, adding the cycling peloton example to the fire ants case, Burge's scheme even seems to apply only to parallel actions. In other words, Burge's scheme does not apply to shared cooperative activities but rather to parallel activities.

## **5.2 Are infants subject to altercentrism?**

Previously, I discussed the success gap between false-belief tasks involving implicit versus explicit measures, presented the nativist position and introduced the altercentric bias hypothesis. If young children have an altercentric bias, they take the other's perspective and do not need any executive function to solve a potential conflict with their self-perspective since, on this view, "self-perspective emerges with the development of cognitive self-awareness, sometime in the second year of life, at which point it leads to competition between perspectives" (Southgate, 2020: 505). What would this bias imply regarding the sharedness characteristic of joint action ontogeny?

### ***5.2.1 Tracking the other's perspective: switching point of view in false-belief tasks***

Adults and young children *track* the locus of others' attention thanks to gaze or head following (e.g., D'Entremont, Hains, & Muir, 1997) and *draw inferences* about others' behaviors based on their perspective (e.g., Luo & Baillargeon, 2007).

In order to reason and draw correct inferences about others' behaviors, especially when a false belief is involved, it seems that one must disregard one's own perspective, which would have to rely on the use of executive functions (which include inhibitory control for instance). However, young children show immature executive functions before four years of age (Diamond, 2013; Devine & Hughes, 2014) and there is a success gap between implicit and explicit false-belief tasks (Chapter 4.2.). Taking this as her starting point, Southgate (2020) questions classical interpretations of success and failure at false-belief tasks. Instead, she proposes that young children display an 'altercentric' bias, that is, a bias to attend to, register and encode events from the visuo-spatial perspective of others. For instance, in the traditional false-belief task, this bias would make an object's location more salient in

the agent's presence (i.e., co-witnessing the event) than in her absence (i.e., not co-witnessing the location switch to the other box). According to this hypothesis, young children, whose self-representation only develops in their second year of life, would pass implicit false-belief tasks because they do not need to solve a self-other representation conflict. This hypothesis is a plausible alternative to the nativists' interpretation of this success gap (false-belief tasks with explicit measures would make extraneous demands). Southgate argues that tracking the other perspective would then allow children to "generate appropriate content and expectations" (Southgate, 2020: 506) because it would reinforce the strength of the others' targets of attention (which would not compete with any self-perspective). What would these expectations rely on? How could infants combine what they perceive (with a lack of self-perspective) with what the other perceives?

### ***5.2.2 The altercentric bias hypothesis favors goal over mental state ascription***

According to Southgate (2020: 517), "self and other representations rely on shared cognitive and neural mechanisms, and the other's perspective is represented in a format akin to a first-person representation". In order to pass false-belief tasks, infants should generate correct content and expectations via predicting what the agent will do based on her knowledge, her previous observations and by inferring unknown reasons when the agent behaves differently than expected. Is Southgate's altercentric bias hypothesis compatible with any of the accounts presented previously?

The altercentric bias hypothesis is compatible with all three accounts previously discussed, namely nativism, empiricism and dual theory of mind. Regarding the nativist position, the altercentric bias hypothesis is compatible with the response load account but this account faces other issues mentioned previously (Chapter 4, 2; Southgate, 2020: 509). Regarding the dual theory of mind, it is also compatible with a two systems account of mindreading (Chapter 3, 2). As Southgate points out, on this view:

...perspective-taking without explicit judgment (i.e., without an explicit question) could be entirely modular, and by virtue of this modularity and encapsulation, be immune from the influence of one's own knowledge of the true location of the object stored elsewhere in the system. However, as soon as a question is posed and an explicit judgment required, the

perspective-taking challenge can no longer be dealt with by an encapsulated module, and at this point is no longer protected from the influence of information – such as the observer’s own perspective – stored elsewhere.

(Southgate, 2020: 508)

Finally, the altercentric bias hypothesis could also be compatible with empiricism since young children might adopt a teleological posture and use behavioral rules, applying them from the other’s perspective.

At the same time, Southgate insists that, while an altercentric bias leads infants to preferentially encode events in the other’s perspective space, what is encoded need not be attributed to that individual as their representation. As she puts it:

Rather, the cowitnessed event, by virtue of being preferentially encoded and better remembered, imposes itself as the infant’s reality. This is possible because self and other representations rely on shared cognitive and neural mechanisms, and the other’s perspective is represented in a format akin to a first-person representation”

(Southgate, 2020: 517)

Going further, the altercentric bias is also compatible with the idea that observed action-events are encoded in terms of goals, without the need to attribute to the observed agent an intention to pursue that goal. For instance, it could be that young children use motor representations to encode goals, mapping what they observe onto their own motor repertoire using motor simulation (Becchio, Sartori & Castiello, 2010) or that they supplement motor simulation with teleological reasoning, as a basis for predicting, and reasoning about, the other behavior.

As mentioned previously, the neural dissociation found by Grosse Wiesmann and colleagues (2020) between implicit and explicit false-belief tasks does not sit well with the nativist position<sup>43</sup>. In

---

<sup>43</sup> “We showed that verbal ToM reasoning was supported by cortical surface area and thickness of the precuneus and temporoparietal junction, classically involved in ToM in adults. Nonverbal ToM reasoning, in contrast, was supported by the cortical structure of a distinct and independent neural network including the supramarginal gyrus also involved in emotional and visual perspective taking, action observation, and social attention or

addition, that study found that the neural network involved in nonverbal false-belief tasks includes the supramarginal gyrus also involved in emotional and visual perspective-taking, action observation and social attention. This result thus tends to support the altercentric bias hypothesis, especially since Grosse Weismann and colleagues (2020) suggest that “success in the implicit ToM task might be related to a stronger focus on other agents, possibly resulting in a social modulation of encoded events” (ibid, 6932). Furthermore, Grosse Wiesmann and colleagues (2020) note that in implicit false-belief tasks, “adults recruit those processes underlying implicit ToM success in early childhood, and, in addition, engage in explicit ToM” (ibid, 6932). In other words, these results tend to confirm the existence of two distinct systems respectively useful for implicit and explicit false-belief tasks. In addition, previous studies have found a similar dissociation between developmental trajectories and brain networks (SMG and dorsal PC versus TPJ and ventral PC), notably “when comparing visual perspective taking with ToM in tasks where participants had to overcome social biases between their own and another’s incongruent visual perspective” (ibid, 6932; see previous studies, e.g., Schurz, Aichhorn, Martin and Perner, 2013). In other words, some social-cognitive processes such as visual perspective-taking or social biases between incongruent perspectives have been shown to develop earlier than explicit ToM reasoning. These elements suggest that visual perspective-taking might support a *belief-like* tracking system efficient in implicit false-belief tasks, before the development of explicit ToM reasoning.

Now, the questions that arise are: (i) how do infants generate correct expectations on that basis?  
(ii) What light does this hypothesis shed on the development of joint action?

### ***5.2.3 The altercentric bias hypothesis implications for joint action***

The altercentric bias hypothesis implies that young children may not need to disentangle their self-perspective from the other-perspective, which would explain how they manage to pass implicit false-belief tasks while lacking mature executive functions. As mentioned above, it could be that young children track the other’s perspective and use motor representations to generate appropriate content

---

encoding biases” (Wiesmann et al., 2020: 1). Another recent study compared whole-brain activation for implicit and explicit false-beliefs tasks; researchers found distinct brain regions for implicit false-belief tasks but an overlap with explicit false-belief tasks only in the right temporo-parietal junction (Boccardoro et al., 2019).

and expectations. For instance, if young children see an agent always reaching for the same object (even if it changes location), they could predict that the agent will keep reaching for the same object and even that she will be looking for this object (e.g., in a false-belief task). Previously, I argued that the altercentric bias hypothesis favors goal over mental state ascription. In addition, while Weismann and colleagues (2020) revealed two distinctive neural systems respectively involved in implicit and explicit false-belief tasks, they also indicated that a similar dissociation was present in tasks comparing visual perspective taking with theory of mind. In other words, it seems that the altercentric bias hypothesis, taken together with motor simulation and teleological reasoning, could account for infants' predictive abilities, without the need to appeal to mentalistic attributions, while at the same explaining the success gap of mentioned above.

What implications could this hypothesis have for the ontogeny of joint action? Over their first three years of life, children's social abilities tend to extend with the development of different competencies such as action prediction, turn-taking, basic forms of common knowledge, motor abilities and coordination skills, normativity understanding and commitment. More specifically, during their first year of life, children interact a lot with their caregivers and they start engaging in activities such as building a block tower, flipping through a book together (around 12 months, see Brownell, 2011). It seems that they create their own motor repertoire by testing reactions they can create during their interactions with others and they also tend to imitate others as part of a learning process. This could be a good use of the altercentric bias. During their second year of life, children notably show a predisposition to help other meet their goals (12-14 months, see Liszkowski et al., 2008; Buttelman et al., 2009), their motor abilities develop and they interact more and more with others. According to Southgate (2020), self-perspective develops with self-representation during their second year of life; it might be that children need to have a good motor repertoire basis and a stronger self-perspective in order to achieve some sorts of joint actions.

A major issue arises from this position with children in their second year of life, i.e., if young children first display an altercentric bias and then develop their self-perspective during their second year of life, one should expect a perspective conflict to emerge with the development of their self-perspective. However, a body of research investigating the development of visual perspective-taking

abilities in childhood found that children seem to develop such abilities from fourteen months of age (Luo & Beck, 2010; Sodian, Thoermer and Metz-Moll, 2007; Moll and Tomasello, 2006<sup>44</sup>). These results seem to show a pattern opposite to the pattern expected from an altercentric bias, i.e., before a certain age, children were unable to take the other visual perspective, meaning that they seem to use first their self-perspective and only later become able to use the other's perspective. One possible explanation could be that there is a difference between being able to *adopt* the other's perspective and being able to *use* it in reasoning. This explanation is plausible since studies found evidence of perspective-taking abilities at different ages depending on the measures used (implicit measures such as looking time, or explicit measures such as a behavior following a question directed to the child). As in false-belief tasks in which successful performance was found in younger children using implicit measures compared to explicit measures, the same reasoning applies to perspective-taking tasks. For instance, Luo and Beck (2010) used a kind of implicit false perception task, where the adult had a different perception of an object compared to the child. In this study, 16-month-olds were first shown an agent always pointing to a red object over other (black or yellow) objects, suggesting that the agent had a preference for red over other colors. In a second orientation phase, while the agent was absent, two screens were introduced and then rotated so that the children could see both of their sides. The screens were either (1) red or green on both sides; (2) red on the front (infants' side) but green on the back (the agent's side) or vice versa; or (3) only colored red or green on the front. In the test phase, the agent, who could only see the back of the screens, pointed to one of the two screens and infants' looking time was measured. The study revealed two main elements. First, infants expected the agent to keep acting based on her color preference and to point to the red rather than the green screen during test. Second, infants' expectations were in accord with the agent's perception of the screens (not their own perceptions), that is infants expected the agent to point to the red screen in (1)<sup>45</sup>, but to the green-

---

<sup>44</sup> In Moll and Tomasello study (2006), toddlers were placed in a room with two objects in front of them (one out in the open and one placed in front of an occluder such that the child see it but someone coming in front of the child would not see it). An adult enters the room searching for an object; the object out in the open was visible but the occluded object was not visible from the adult's perspective. When toddlers were directly asked by the adult to help find the object, only twenty-four-month-olds handed him the occluded object (whereas in a control condition they showed no preference for the occluded toy).

<sup>45</sup> Infants looked reliably longer when the agent pointed at the green screen in (1).

front screen in (2)<sup>46</sup>, and they had no prediction as to which screen the agent would point to in (3). The study by Luo and Beck (2010), in which implicit measures were used (looking time), revealed that from 16 months of age, children were able to form expectations about an observed agent's behavior based on that agent's perspective (rather than their own perspective). In contrast, the study by Moll and Tomasello (2006) study, in which explicit measures were used (see footnote n°15), revealed that only twenty-four-month-olds were able to use the adult's perspective (who could only see one of the two objects) to hand him the object he was looking for (i.e., the object which only the child could see from his/her perspective). In other words, it could be either that the measures used and the nature of the task show results that are not representative of children's perspective-taking abilities or that children are able to adopt the other perspective but they cannot use the other perspective in reasoning.

In addition, another body of research investigating the ability to co-represent the partner in a joint task concluded that the "co-representation effect" (which creates interferences in self-performance) does not appear before four years of age (see Chapter 2 for more details; Milward, Kita & Apperly, 2014<sup>47</sup>, 2016<sup>48</sup>). In that case then, it seems that young children are not able to take into account the other perspective before four years of age. Again, it could be that the nature of the task does not represent properly the conflict. To my knowledge, the experimental paradigm created by Milward and colleagues (2014, 2016) is currently the only one available to test the conflict of perspectives but it involves complex tasks. Indeed, the absence of co-representation effects before four years of age can be explained relatively to the task complexity. It could be that the partner's co-representation can be more or less challenging due to the task itself and, more specifically, to the skills entailed in successfully complementing the partner's action (e.g., problem-solving, flexibility, action planning skills, motor skills). Puzzle-solving tasks, which can be more or less complex, tend to

---

<sup>46</sup> Infants looked reliably longer when the agent points at the red screen in (2) since the red screen from their perspective corresponds to the green screen from the agent's perspective.

<sup>47</sup> Milward and colleagues (2014) found that adult-like interference effects between self and other were only measured in 4- and 5-year-olds (not in 2- and 3-year-olds).

<sup>48</sup> Milward, Kita & Apperly (2016) made a follow-up study assessing individual differences which revealed that theory of mind, helping children to clearly differentiate and separate task representations of self and other, and inhibitory control were predictors of the ability to avoid interference effects resulting from the partner's co-representation.

confirm this explanation (see Chapter 2, 2.1.2.3). In other words, co-representation effects should be measured in studies involving less complex tasks, notably tasks requiring motor competencies instead of planning skills. Even in their third year of life, while children do integrate normative factors in a game for instance (25-27 months, see Warneken & Tomasello, 2007, Gräfenhain et al., 2009) and use gestures and verbal indicators to coordinate (30 months, see Hamann et al., 2012), they display quite limited planning skills (e.g., Meyer et al., 2016).

More generally, the main point that arises from this discussion of the altercentric bias hypothesis is that it favors *sharing* at a lower-level. In fact, if young children are able to *adopt* the other's perspective but show difficulties in being able to *use* it in reasoning before four years of age due to the task complexity, they probably *share* at a lower-level. Young children might share at a lower-level thanks to some social-cognitive processes such as visual perspective-taking which might support a *belief-like* tracking system efficient in implicit false-belief tasks before the development of explicit ToM reasoning. It could be that young children use visual perspective-taking and motor representations to encode goals, mapping what they observe onto their own motor repertoire using motor simulation (Becchio, Sartori & Castiello, 2010) or that they supplement motor simulation with teleological reasoning, as a basis for predicting the behavior of the other agent. This conclusion, on the one hand, runs counter to the maximalist position and on the other hand is compatible with the third proposition in our trilemma. In other words, it seems that infants *perform motorically shared* cooperative activities.

To recap, I questioned whether young children do engage in shared cooperative activities (which is the last part of our trilemma) by discussing two alternatives. On the one hand, I discussed Burge's sensing-action attribution scheme which excludes completely any form of mentality and any sort of sharedness. I concluded that while Burge's sensing-action attribution scheme might be used to account for certain types of parallel activities (which from an external point of view look like shared cooperative activities), the kinds of activities children engage in with others cannot be reduced to such parallel activities. On the other hand, I introduced Southgate's altercentric bias hypothesis which is compatible with the main three accounts (nativism, empiricism, dual theory of mind) but which, regarding our topic, seems to accord better with the empiricist account and the dual theory of mind

account. My main conclusions are that Burge's scheme cannot be used to cast doubt on the third proposition of our trilemma since it cannot provide a convincing deflationary account of children's cooperative activities and that Southgate's altercentric bias hypothesis indirectly maintains the sharedness but at a low-level (motor), suggesting that young children might perform *motorically shared* cooperative activities.

\*\*\*

The aim of Part II was to evaluate whether observations in childhood are compatible with the maximalist account of joint action, i.e., to confront joint actions as they are defined by maximalists – in terms of high-level cognitive skills – with developmental psychology studies. To do so, I introduced a trilemma resulting from the confrontation between maximalism and minimalism (Rakoczy, 2018) i.e., (1) Shared cooperative activities require sophisticated cognitive abilities to represent and reason about mental states; (2) Young children do not have those sophisticated cognitive abilities; (3) They engage in shared cooperative activities. In order to solve this trilemma, I examined each claim independently.

One can challenge the first claim by adopting an empiricist account (eliminating the mental aspect of maximalists) or a dual theory of mind account (offering a minimal version of theory of mind) Both of these positions offer alternatives to the strong high-level cognitive pre-requisites of maximalist accounts. When engaging in joint action, agents might use teleological reasoning or a minimal theory of mind to ascribe goals rather than intentions in order to predict the behavior of other agents and adjust to what they are doing.

The second claim can be challenged by adopting a nativist position, according to which very young are already equipped with mindreading abilities. Nativism offers a mentalistic interpretation of children's false-belief tasks understanding. According to nativism, the success gap observed between implicit and explicit false-belief tasks can be explained by the fact that children's competence is masked by task demands (verbal or executive function skills) in explicit tasks. In other words, children meet the cognitive requirements for engaging in shared cooperative activities set by maximalists. However, this mentalistic approach is itself challenged by some recent studies.

Challenging the third claim consists in testing the “shared” character of the activities children engage in with others. It is notably hard for researchers (empiricists and dual theory of mind advocates) who assert that children access others’ perspective to explain how they can ignore their own perspective in implicit false-belief tasks. One way of challenging the third proposition consists in considering that the activities children engage in with others could be explained without involving representational abilities or shared representations. However, an explanation in terms of Burge’s sensing-action attribution scheme (which excludes completely any form of mentality) must be ruled out as it does only apply to parallel activities and the kinds of activities children engage in with others cannot be reduced to such parallel activities. Another way of challenging the third claim consists in considering that, in order to share representations, children should be able to manage several perspectives at the same time which seems to be difficult. In fact, the introduction of Southgate’s altercentric bias hypothesis which indirectly maintain the sharedness but at a low-level (motor), suggests that young children might perform motorically shared cooperative activities.

To conclude, the second claim, i.e., young children do not have those sophisticated cognitive abilities, could not be successfully challenged. From my discussion of the two others claims, I concluded that (i) developmental data do not support the maximalist account of joint actions, (ii) the teleological reasoning and the two mindreading systems accounts offer two good candidate accounts of the kind of cognitive abilities that support joint action in childhood, (iii) the altercentric bias hypothesis would allow motorically shared cooperative activities in young children.



*Part III*

*A Joint Action Matrix*



## ***Introduction to Part III***

The specificity of joint actions lies in the fact that they require *shared representations* which, according to maximalism, would have to be intentions and, according to minimalism, could be goal representations (Part I). An examination of the human ontogeny of joint action tends to favor a minimalist account of joint actions, meaning that there are some joint actions which do not require understanding others in mental terms but simply understanding their actions in goal terms (Part II).

While performing joint action, co-agents need at the same time to track others' actions, represent their goals and to plan and execute actions of one's own to coordinate on a common goal. Tracking others' actions may involve motor representations competing with the representations underpinning action production, leading to the emergence of interference effects which hinder the joint action performance. Conversely, the congruency between the agent and her co-agent may in certain situations facilitate the task performance. In fact, the observation of actions congruent with one's own facilitates task performance, while the observation of incongruent actions leads to visuomotor interference effects (Clarke et al., 2019). Previous studies have shown that, faced with this issue, co-agents' strategies depend on the trade-off between the *task requirements* and the *uncertainty of the information available about the partner* and improve through learning and adaptation (for a recent review, see Chackochan & Sanguineti, 2019). In other words, the key element in joint actions seems to be that way co-agents encode one another and adopt strategies to handle both the task requirements and information relative to the partner.

My aim in Part III will be to develop a typology of joint actions and their underlying mechanisms that should help us better understand what types of joint actions young children are capable of and when this capacity emerges during development. This typology, or matrix, is organized around two axes: (1) the type of shared representation involved and (2) the level of sharedness implied by such shared representations. The type of shared representation refers to the action complexity (i.e., does this joint action require the use of fully developed mentalizing and metarepresentational abilities?) constituting a first matrix axis, presented in Chapter 6. The level of sharedness implied by

such representation refers to the co-agents' strategies (i.e., identical or complementary) in the joint action, constituting a second matrix axis, presented in Chapter 7.



## *Chapter 6*

### *Does joint action's complexity impact the type of shared representations?*

The specificity of joint actions lies in the fact that they involve representations shared between co-agents. As presented in Part I, in philosophy of action, maximalist approaches take it that joint action requires co-agents to share intentions and this, in turn, demands theory of mind capacities since agents should be able to attribute mental states (e.g., intentions) to themselves and to their partners (for a review, see Butterfill & Sebanz 2011). Other researchers, minimalists, consider that some kinds of joint action do not require these high-level cognitive processes but rather the ability to track the goals of others' actions using either motor simulation (Sinigaglia & Butterfill 2016; Butterfill, 2019), teleological reasoning (Gergely & Csibra 2003; Csibra & Gergely 2007, 2013), learned statistical correlations between actions and outcomes (Elsner & Aschersleben 2003; Leslie 1984) or a combination of these processes.

Young children develop theory of mind abilities around four years of age. If they are able to perform some kinds of joint actions before that age, it seems that they use representations of goals in order to participate in minimal joint actions (Part II). How then can young children form and use these representations when participating in joint actions? What kind of joint actions are they capable of engaging in given that they rely solely on representations of goals? In this chapter, I will first introduce a typology of goals and discuss the kinds of processes that can be involved in computing the goal of an action and their potential limitations (Section 1). I will then consider the same issue from a developmental perspective (Section 2). Given that different types of processes may be responsible for computing goals, which of these processes are actually used by young children? Finally, I will consider the role played by goal tracking in joint action (Section 3).

#### **6.1 Tracking and representing goals**

As pointed out by Sinigaglia and Butterfill (2016), the term goal can be used either to refer to the outcomes to which actions are directed or to refer to mental states of agents, i.e., the intentions or pro-attitudes in virtue of which their actions are directed at a certain outcome. In this chapter, unless otherwise specified, I will use goal in the first non-mentalistic sense. Similarly, I will use goal-ascription to refer to the process or processes through which one identifies the outcome to which an action is directed and goal representation to refer to the representation of the outcome to which an action is directed.

### ***6.1.1 A simple typology of goals***

The idea that there are many different levels of description of actions and that goals themselves are organized hierarchically is widely accepted both in philosophy (e.g., Davidson, 1980; Goldman, 1970, Searle, 1983) and in the motor control domain (e.g., Hamilton & Grafton, 2007; Kilner, 2011). While the number of levels identified may vary depending on authors and interests, there is general agreement that there is no one-to-one mapping between actions and goals at different levels. The same goal can be broken in different series of subgoals and realized by many different actions and the same action can serve to realize many different goals. Here, I offer a simple typology of goals arranged from more concrete to more abstract.

Among concrete goals, some goals can be immediate when others can be superordinate. A goal is superordinate when it requires the implementation of a sequence of actions to be achieved. For instance, the goal of grasping a mug is a concrete and immediate goal which requires one main motor action sequence (direct goal-realization relation) whereas the goal of drinking a glass of water is a concrete superordinate goal which requires a sequence of motor actions (mediated goal-realization relation). Let's take the example of Mathilde who wants to drink a glass of sparkling water when there are two water bottles (one sparkling and one still) on the table. Mathilde's goal needs to be broken down in a series of motor sub-goals and a corresponding sequence of actions to be performed. Her main goal of drinking a glass of sparkling water can be decomposed in a series of steps such as, (a) grasping the appropriate bottle, (b) pouring water in a glass, (c) grasping the glass, (d) bringing the glass to her mouth, and (e) swallowing the water. In this case, contrary to grasping a mug, the main

goal (drink a glass of still water) is decomposed in a sequence of actions with motor sub-goals (a, b, c, d, e) necessary to achieve in order to reach the main goal.

Abstract goals involve types of actions which cannot be directly implemented through motor representations and require a higher-level form of planning than simply motor planning. In fact, the performance of abstract goals implies a more indirect relation between the goals and their motor realization. For instance, the goal of going on vacation is an abstract type of goal involving a specification of concrete sub-goals based on non-motor planning before being deployed in motor representations (indirect goal-realization relation). The type of action(s) needed to perform the goal of going on vacation depends on the individual who forms this goal and engages in practical deliberation to further specify the goal and the means to toward that goal. For instance, it might involve other individuals (e.g., going on vacation with friends or children), depend on contextual elements (e.g., having a two-week vacation or a school vacation determines the dates of departure and return) and different means might be used to reach the chosen location (e.g., plane, train, or car).

In other words, in the case of abstract goals, the relation between the goal and its realization is much less direct than in the case of concrete goals.

### ***6.1.2 Goal ascription using motor representations***

According to motor theories of goal-ascription, some goal-tracking processes involve only motor processes and representations (Butterfill, 2019: 12). This approach is premised on three main ideas. The first idea is that motor representations can represent the goal of an action, i.e., the outcome toward which the action is directed, and not just bodily movements, joint displacements or end states. The second idea is that motor representations have control and monitoring functions. Representations of outcome trigger planning-like motor processes that compute the motoric means to that outcome (control) as well as computing sensory expectations or predictions concerning the effects of the action (monitoring). The third idea is that when observing rather than preparing an action, motor representations are also activated and planning-like processes can be used in reverse mode to compute goals from means.

The claims that motor representations can represent action outcomes and that they have control and monitoring functions are among the central tenets in much of the recent motor control and motor cognition literature (Jeannerod 1997, 2006; Prinz 1990; Rizzolatti and Sinigaglia 2008; Rosenbaum 2009; Wolpert, Ghahramani, and Jordan 1995) and are well-supported by empirical evidence. Butterfill and Sinigaglia point out that the claim ‘the motor representations can represent action outcomes’ is supported by evidence that “markers of motor processing are correlated with action outcomes rather than narrowly kinematic or dynamic features of action” (2014: 122). For instance, cortical motor neurons that are active during hand grasping in monkeys, also become active during grasping with pliers, both normal pliers and “reverse pliers”, requiring finger opening, instead of their closing, to grasp an object, once the monkeys have been trained to use these tools, suggesting that these neurons code grasping actions, independently of the means used or of the kinematics of the movements involved in grasping (Umiltà et al., 2008). Similarly, the idea that motor representations have control and monitoring functions is central to computational theories of motor control. According to influential theories of human motor control (Jordan & Wolpert, 1999; Wolpert & Ghahramani, 2000; Wolpert, 1997; Jeannerod, 1997; Frith et al., 2000), motor control involves two main types of internal models. Inverse models (or controllers) compute the motor commands needed for achieving a desired outcome given the current state of the system and of the environment. Efference copies of these commands are fed to forward models (also called predictive models) that represent the causal flow of a process in a system and can thus generate a prediction of the sensory consequences of performing these “commands”. Inverse and forward models are coupled through a series of “comparators” where comparators can be defined as mechanisms that compare two signals and use the result of the comparison for various kinds of regulation.

Finally, the idea that motor representations are activated not just when preparing or executing an action but also when observing someone else acting, is supported by a wide range of empirical evidence. In a series of single neuron recording experiments on macaque monkeys investigating the functional properties of neurons in area F5, Rizzolatti and his colleagues discovered so-called mirror-neurons, i.e., sensorimotor neurons that fire both during the execution of purposeful, goal-related actions by the monkey and when the monkey observes similar actions performed by another agent

(Fogassi & Gallese, 2002; Rizzolatti & Craighero, 2004; Rizzolatti & Sinigaglia, 2007). A large body of neuroimaging experiments in humans have investigated the neural networks engaged during action generation and during action observation, revealing the existence of an important overlap in the cerebral areas activated in these two conditions (for reviews, see Grèzes & Decety 2001; Jeannerod 2006). Behavioral studies have also shown that observing one action can cause an interference with the simultaneous performance of a second action by the agent that is similar to the interference that would occur if the agent had to perform the two actions simultaneously (Kilner et al., 2003).

Let us imagine that Cathie wants to turn on the light in the living room; the action outcome is that the light is turned on, her target is the light switch, the type of action a pressing. The motor representation represents the action outcome (the light is turned on). However, it is not sufficient for motor representations to represent information about the action's outcome; they also need to play a role in planning and monitoring the action. As Sinigaglia and Butterfill put it: "motor representations of outcomes trigger planning-like processes which generate predictions" (2016: 153). These planning processes yield a specific motor plan tailored to relevant aspects of the context. For instance, the way the pressing is realized (the effector used, the sequence of joint displacements, the speed of the movement) depends in part on the height and distance of the light switch relative to the agent. This more detailed motor representation will also allow Cathie to anticipate future states of the environment, generate expectations about the sensory consequences of her bodily movements and, if needed, adjust her movements during action execution. For instance, Cathie can predict that the light will be on once she presses on the light switch, but if it is not the case, she might probably press on it several times to check whether it is broken.

Paul observes Cathie. For Paul to represent Cathie's goal consists in several steps: (1) identifying her action target (the light switch); (2) tracking information present in the scene (e.g., Cathie's finger position, the direction toward which her movement is oriented) and retrieving actions learned to be associated with that target; (3) encoding the motor parameters to generate a prediction of the sensory consequences of the observed action (several actions can be encoded); (4) predicting the

sensory consequences of the most probable action<sup>49</sup>. Both Cathie (to perform her goal) and Paul (to understand Cathie's goal) compare the actual outcome they perceive through sensory feedback (e.g., visual, tactile) with their predictions to adjust their goal representation in first-person (Cathie) or in third-person (Paul).

Sinigaglia and Butterfill (2016) thus take it that goal ascription can be achieved motorically since motor representations represent not just movements but also action outcomes, since motor representations of outcomes trigger motor control processes whose function is to insure that the movements performed will bring about the represented goal, and since in observers these same motor planning processes are activated and can be used to predict which movements are likely to be performed.

However, there are limitations to the kind of goal ascriptions motor representations can help achieve. The goals encoded by motor representations are (1) motor goals rather more complex or abstract goals (e.g., going on vacation) and (2) must be present in the agent's and observer's motor repertoires (Rizzolatti & Craighero, 2004). Could other goal ascription processes help overcome these limitations and allow for the ascription of new, unfamiliar or improbable goals as well as more abstract goals?

### ***6.1.3 Goal ascription using learned statistical correlations between actions and outcomes***

Another way of considering goal ascription is to reason in terms of action-effect associations. On this view, individuals use action-effect associations – stored statistical information about previously observed co-occurrence of observed behaviors (actions) and their outcomes (effects) – to ascribe goals to currently observed behaviors. These action-effect associations are “bidirectional” (Csibra & Gergely, 2013: 2), meaning that stored associations can be used to predict the goal of an ongoing action, or to anticipate the behavior adopted by an individual whose goal is known. Such statistical

---

<sup>49</sup> Kilner (2011) describes five steps in the process of action understanding (here, drinking a cup of tea). “The first step is visual processing and identification of the object as a cup. The second step is the retrieval of actions that we have learned to be associated with that object. The third step is the selection of the most probable actions given the goal. Note here that more than one action can be selected but that the likelihood of the action can be signaled through the strength of that action's representation...The fourth step is the encoding of the motor parameters to generate a prediction of the sensory consequences of the observed action. Again multiple actions can be encoded as before. The fifth step is the prediction of the sensory consequences of the most probable action. Here only the most probable action is encoded” (ibid, 354).

information may contribute to gradual learning about goal-directed actions, including more complex motor actions (e.g., superordinate goals i.e., goal-directed behaviors requiring a sequence of actions to reach the desired effect) and the use of social norms (e.g., politeness) implicitly involved in the action-effect associations.

Action-effect associations establish a link between the representation of the action and the representation of desired effects in the observer's cognitive system (Csibra & Gergely, 2007; Elsner, 2007); the activation of one of the associated representations elicits the other representation<sup>50</sup>. The observer detects *agency cues* displayed by the agent which support the identification of the action goal (for a recent review, see Elsner & Adam, 2020). For instance, pointing is immediately interpreted as a communicative gesture to orientate the gaze of others in a direction (e.g., toward an object). From there, the process can be decomposed in three phases combining *bottom-up and top-down processes*<sup>51</sup>. First, the observer processes information about the initial situation and the beginning of the movement (bottom-up process). Second, the observer maps the bottom-up information into her cognitive action representation (e.g., Elsner & Hommel, 2001). Concretely, it is crucial that from an early age, infants gain experience in contact with agents who display agency cues to form strong action-effect associations. The observation of repeated action-effect instantiations creates stored associations which, once learned, can operate as prior beliefs constraining the interpretation of observed behavior bidirectionally. Confronted with unfamiliar agents or actions, the solicited stored action-effect associations are either weak (because they have not been instantiated very often) or contain only fragmentary feature representations. Third, based on the observer's stored action-effect associations, the bottom-up information activates stored action features and transitional probabilities which can trigger top-down influences (i.e., expected goal state) that can be measured via predictive gaze-shifts (Csibra, 2007; Southgate, 2013). In the case of unfamiliar actions or agents, the visible agency cues

---

<sup>50</sup> "...the 'idea' of the goal state, i.e., the desired effect, automatically activates the corresponding action, while the activation of an action elicits the anticipation of the distal effect associated with it" (Csibra & Gergely, 2007: 5).

<sup>51</sup> For instance, Elsner & Adam (2020) argue that "the impacts of action-experience and of agency cues on infants' action-goal prediction can be explained from an action-event perspective. In particular, we assume an interplay between bottom-up processing of the perceivable action features and top-down influences based on action-event schemata that store previously experienced features of the agent, movement, and goal state, as well as their transitional probabilities" (ibid, 14).

might activate a corresponding familiar action-effect association with similar initial features enabling a predictive gaze-shift<sup>52</sup> after the first familiarization trials. For the observer to attribute goals or predict actions based on action-effect associations, he/she needs to implicitly rely on the assumption that “...an action is directed towards the same goal state that has been produced by earlier, similar actions, and that the a goal state (an effect) will be achieved by an actor in a similar way, as it was achieved before” (Csibra & Gergely, 2007: 6). In other words, action-effect associations enable the observer to predict the goal of an observed action or to anticipate behaviors while the agent’s goal is known, even though the action is absent from his/her motor repertoire, the observer relying on statistical information from prior observations which operates as prior beliefs.

The advantage of goal ascription using action-effect associations, in comparison with the motor theory of goal ascription, is that the observed action does not need to be present in the observer’s motor repertoire. In fact, the observer processes bottom-up information (e.g., agency cues, action features) which activate stored action-effect associations; such action-effect associations operate as prior beliefs constraining the interpretation of observed behavior via top-down influence. However, goal ascription using action-effect associations encounters a difficulty, namely the case of totally new observed actions or improbable actions which do not share any similar or common action features with stored action-effect associations.

#### ***6.1.4 Goal ascription using teleological reasoning***

One answer to the challenge raised by new observed and improbable actions could be the use of teleological reasoning. Gergely & Csibra (2003) present the teleological stance as establishing a teleological “explanatory relation among three relevant aspects of current and future reality: the action, the (future) goal state, and the current situational constraints” (ibid, 289). In fact, they argue that teleological representations relate three elements (the action, the goal state and the situational constraints) connected by an efficiency principle which generates explanations and predictions about the observed action (ibid, 289). The teleological stance appeals to a principle of rationality or

---

<sup>52</sup> The observer gaze is coded differently based on whether it arrives at the target object before (predictive gaze-shift), simultaneously with (tracking gaze), or after (reactive gaze) the agent had contacted the target object (e.g., Adam & Elsner, 2018).

efficiency that presupposes that, actions function to bring about future goal states, and that these goal states are realized by the most efficient action available to the agent within the constraints of the situation (Gergely & Csibra, 2003; Csibra & Gergely, 2013). This teleological stance allows an observer to infer the goal of an agent based on the observation of the actions engaged in and of the constraints of the situation.

Csibra and Gergely (2013) consider that teleological reasoning enables three kinds of inference:

- (1) The prediction of action in situations in which the environment changes but the goal does not;
- (2) The prediction and attribution of goals to ongoing actions in advance of their completion;
- (3) The inference of situational constraints from an observed goal-directed action.

As an example of (1), an observer can predict that an agent will adjust the trajectory of their arm if while they are reaching for an object, the object starts moving. From the goal of the agent's action together with the situational change, one can predict the now more efficient action alternative. As an example of (2), an observer seeing the trajectory of an agent's hand and the way they shape their hand can infer that the goal of the action is to grasp the glass on the table rather than the pen also lying on the table as the observed behavior is an efficient action toward that outcome in that situation. Finally, the assumption of efficiency can help us infer situational constraints. As an illustration of (3), an observer seeing an agent making a detour to reach the door of a room rather than taking the shortest path will be in a position to infer that some obstacle is present on that path, since absent such a situational constraint, taking a detour would not be an efficient action.

Csibra and Gergely (2007) distinguish between two ways in which teleological action interpretation allows the prediction of future events, corresponding to two direction of inference one can use. On the one hand, interpreting an ongoing action can involve an 'action-to-goal' inference ('what is the function of this action?') allowing us to predict the outcome of the action. On the other, it can also involve an inference from 'goal-to-action' ('what action would achieve that goal?') allowing us to anticipate future actions. To take their example, if one interprets observed ongoing behavior in the kitchen as coffee making, one can predict the appearance of cups of coffee in the kitchen in the near future, but in addition, once one has identified the goal of an action, one can also predict the next steps

of the coffee making action. In fact, as they insist, both kinds of inference and both kinds of prediction are intertwined in perception since observers constantly revise goals and sub-goals attributed to actors on the basis of verified or disconfirmed action predictions. In addition, Csibra and Gergely claim that these two types of inferences also serve another key epistemic function: the social learning of actions. Observing others perform actions that are novel to us can help one discover novel and more efficient means to achieve a goal, as well as discover novel goals and artefact functions. As Csibra and Gergely (2007) point out, social learning is a distinctive function served by teleological reasoning compared to other procedures for goal attribution: “while action-effect associations and simulation procedures are generally well suited to serve on-line action monitoring and prediction, social learning of new means actions and artefact functions requires the inferential productivity of teleological reasoning” (ibid, 60).

Motor representations can support goal ascription but there are important limitations to the kinds of goals that can be represented motorically. For instance, in the case of new, unfamiliar or improbable actions or actions involved in abstract goals, reasoning based on prior knowledge or the adoption of the teleological stance may be needed to infer an action’s goal. Previously learned information about the statistical co-occurrence of observed actions and their outcomes, prior knowledge of the preferences or idiosyncrasies of the observed agent or knowledge of the behavioral norms that apply in a given situation can constrain an observer’s interpretation of an observed action. This may help overcome some of the limitations associated with motor-based goal ascription. For instance, this knowledge can be deployed in cases where the observed action is not part of the motor repertoire of the observer or when inferring a higher-level goal beyond the immediate motor goal of an action. Prior knowledge will not help, however, if the action is novel. In such cases, teleological reasoning using the principle of efficiency might still help us infer the goal of an action from the action observed and the constraints of the situation in which it takes place using the principle of efficiency. Thus, while motor representations may play an important role or indeed be sufficient for functional goal ascription in the case of simple motor goals, the more complex or abstract the goal of an action is, the more motor representations will need to be supplemented with other sources of information to allow for goal ascription.

## 6.2 Goal tracking in young children: what processes?

In the previous section, I introduced three types of mechanisms available to humans for goal ascriptions, involving respectively motor simulations, the exploitation of stored action-effect associations and teleological reasoning. In this section, I will discuss these goal-ascription mechanisms from a developmental perspective. How do infants and young children track goals? When in the course of development do these different mechanisms become available to children for goal-ascription? While goal-ascription achieved via teleological reasoning or the exploitation of stored action-effect associations relies on general inferential processes, the motor simulation procedure relies specifically on motor processes. Do children first achieve goal-ascription through general inferential processes or do they initially rely on motor processes for goal ascription?

### 6.2.1 *The Motor Theory of Goal Tracking (Butterfill, 2019)*

Butterfill (2019) contrasts what he calls the Simple View – the view according to which humans, both adults and children, rely on general inferential processes for goal ascription with his own view, the Motor Theory of Goal Tracking, according to which goal-tracking can rely on motor processes. He defends the following conjecture:

*The Developmental Motor Conjecture* In the first 9 months of life, all pure goal tracking is explained by the Motor Theory. Other goal-tracking processes emerge later in development.

(Butterfill 2019: 13)

In other words, according to this conjecture, motor simulation is the only procedure that infants (under nine months of age) can use to inform their inference about others' goals. More precisely, Butterfill argues that "...they can only track the goals of an action if they can represent a similar enough<sup>53</sup> action motorically at the time the action occurs" (ibid, 14). Butterfill argues that conjecture is

---

<sup>53</sup> Here, "similar enough" means either that two actions are both of the same kind (e.g., reaching actions) or such that the differences make no difference for the purposes of goal tracking in that context (e.g., reaching and pre-reaching).

supported by a body of evidence showing that, “at least in the first 9 months of life<sup>54</sup>, infants can only track goals of an action they can represent motorically at the time the action occurs” (ibid, 6). Explaining this limit in children’s goal-tracking abilities appears to raise a challenge for the Simple View. While defenders of the Simple View might agree that acquiring motor abilities provides new knowledge of means-ends relations (thus enhancing goal-tracking abilities), they have no obvious way of explaining why motor skills impact so much the goal-tracking process.

In their first year of life, children have quite limited motor capacities (and thus a limited motor repertoire). It is legitimate to wonder whether, as Butterfill’s Developmental Motor Conjecture proposes, infants can only track actions’ goals when they can represent a similar enough action motorically at the time the action occurs (Butterfill, 2019). Studies with young children have shown that first-person experience modulates effectively their action perception and notably their representations of actions in terms of goals (Gergely et al., 1995; Sommerville et al., 2005; Loucks & Sommerville, 2012; Needham et al., 2002, 2017). Some studies have found experimental ways to enhance infants’ abilities to act and tested how much their newly acquired motor skills impact their representation of goals (Needham et al., 2002; 2017). In these studies, experimenters put ‘sticky mittens’ on a group of 3-month-old infants who could play with objects (with a control group who did not wear ‘sticky mittens’). Using the paradigm of Woodward (2009), they found that only 3-month-

---

<sup>54</sup> Around nine to twelve months of age, some evidence indicates infants develop an understanding of goals more as abstract relations that organize physical actions which probably consolidate and favor the use of other processes than motor simulation in their reasoning about the means-outcome relation perceived in observing actions. For instance, they use looking events toward a focus of attention (e.g., an object) to condition their expectations about observed agent’s instrumental actions. They can notably expect that an agent will act on the object she just attended to (Phillips et al., 2002). Young children also show the ability to represent more distal goals i.e., goals requiring the performance of a sequence of actions. For instance, they understand an action on an intermediary object (e.g., a box or a tool) as directed to the access of another object (e.g., a toy inside the box or an object out of reach) rather than at the box or the tool itself (Woodward & Needham, 2008; Sommerville et al., 2008). From twelve months of age, evidences show that infants use intuitions about probabilities to shape expectations about the future (Bonatti, 2008; Téglás et al., 2007) and that young children use this predictive reasoning about other agents’ beliefs, goals and behaviors (Woodward, 1998; Onishi & Baillargeon, 2005; Kovács, Téglás & Endress, 2010). For instance, by twelve months of age, children take into account the knowledge or ignorance of the agent (e.g., she perceived or not that the object she is looking for felt down) and gesture appropriately to provide information (e.g., point to the object actual location) necessary in the agent’s goal performance (Liszkowski, Carpenter & Tomasello, 2008). At this age, while their motor abilities get better, researchers even measured variations in infants’ fine-grained goal representation based on their own motor capacities. For instance, Loucks & Sommerville (2012) showed 12-months-olds’ sensitivity to a causally inappropriate prehension in the frame of a goal-directed action is related to the precision of their own prehension abilities.

olds who had worn the mittens showed enhanced abilities to track the goals of others' actions. Conversely, studies have shown that children's own ability to perform actions influences their ability to detect actions impossible to perform (Reid et al., 2005; Southgate et al., 2008). Southgate and colleagues (2008) revealed that six- to eight-month-olds ascribe a goal to a biologically impossible reaching and grasping action while noting that some eight-month-olds whose motor abilities are more developed detect the movement's impossibility<sup>55</sup>.

If the studies presented above seem to support Butterfill's position, several studies have shown that even three-month-olds are already capable "of extracting goal-related information from displays involving simple geometric shapes" (Butterfill, 2019: 13; Luo, 2011). These results have been replicated several times and seem, at first glance, to disprove Butterfill's Developmental Motor Conjecture. In fact, it is improbable that the observation of moving geometric shapes involves a process of motor simulation. However, Butterfill argues that when infants, in the first 9 months of life, might appear to be tracking goals which they are unable to represent motorically, they are merely tracking targets.

According to the Motor Theory of Goal Tracking, under nine months of age, infants only use motor simulation in order to infer the goals of others' observed actions; infants do not possess the resources to use other processes yet.

### ***6.2.2 The Simple View (Gergely & Csibra, 2003)***

The advocates of the Simple View consider that motor simulation is not necessary for goal understanding and that instead, when observing actions, goal understanding may be a precondition of motor activation (e.g., de Klerk, Southgate and Csibra, 2016; Southgate, Johnson, Karoui and Csibra, 2010). For instance, Southgate and colleagues (2010) presented 9-month-olds with part of an action.

---

<sup>55</sup> Note that Southgate et al. (2008) take the fact that younger infants appear to extend goal attribution even to biomechanically impossible actions so long as they are physically efficient as evidence that the notion of 'goal' is unlikely to be derived directly from infant's experience. One might reply, however, that children at that age are already capable of reaching and grasping actions are thus already possessing at least coarse-grained motor representations of these actions. Yet, with more personal experience reaching for, and grasping objects, their motor representations are further refined and become more fine-grained. Thus, an alternative explanation of the findings of this experiment would be that biologically impossible reaching actions are sufficiently similar to the coarse-grained motor representations of the younger children to allow for motorically based goal ascription but that once children acquire more fine-grained motor representations, more stringent criteria for similarity apply.

For this action to be interpreted as goal-directed, the infants would need to predict an outcome for the action. The infants' prediction was measured using the attenuation of the sensorimotor alpha signal during observation of action. In this study, two conditions were presented: the occluder condition and the mimed condition. In the occluder condition, a grasping movement was presented such that infants should predict the presence of a target object behind the occluder. Conversely, in the mimed condition, the same grasping movement was presented but without the occluder such that infants could see that there was no target object. In fact, infants exhibited evidence of motor activation only if the observed action permitted them to infer a likely outcome, i.e., only in the occluder condition. This result provides evidence for on-line action prediction in infancy; more precisely, according to the authors' interpretation, "motor activation that is dependent on the ability to infer the existence of an object behind the occluder implies that this activation is the result of, rather than a precondition of, goal understanding" (Southgate et al., 2010: 358).

In another study, Southgate and Vernetti (2014) found that 6-month-olds take into account the agent's knowledge when predicting their actions. This study revealed that the presence of motor activation was related to action prediction as infants expected the agent to reach for the box basing their action prediction on what the agent believed to be the case (motor activation was present when the agent had a false belief about the ball's presence but not when the agent had a false belief about the ball's absence). In other words, it seems that young children rely on predictive processes which trigger motor representations in order to pass false-beliefs tasks. In what follows, I will present evidence from developmental psychology in favor of two types of processes, namely processes using means-goals associations and teleological reasoning.

#### *6.2.2.1 Means-goals associations*

Let's consider a concrete goal such as in the example of Mathilde; Anna is observing Mathilde. To infer not just Mathilde's immediate motor goal (e.g., grasping a water bottle) but also the larger goal it serves (e.g., having a drink), Anna can rely on top-down information such as prior beliefs or assumptions (e.g., usually, when people grasp a water bottle and have a glass, they help themselves with a drink) to interpret the observed situation. In addition, Anna can use her knowledge about

Mathilde (e.g., Mathilde is polite) or behavioral rules or norms related to a specific context (e.g., code of conduct when guests are coming over). For instance, in such cases, Anna will not anticipate that Mathilde will drink from the bottle but rather use a glass. In the case of abstract goals such as going on vacation, an observer may infer the goal of the observed agent by connecting observations of the action's target (e.g., dragging a *wheeled suitcase*) or of the action's location (e.g., showing a passport *at the airport*) with prior beliefs or assumptions in order to represent the agent's goal. Again, action-effect associations do not necessitate any first-person experience, meaning that, an individual who never went on vacation might have watched movies and have identified action features related to this abstract goal and thus, be able to ascribe a goal to an observed agent.

As mentioned previously (6.1.3), bottom-up information activates stored action-effect associations (e.g., through action features) which, then, can trigger top-down influences (i.e., expected goal state) which can be measured via predictive gaze-shifts (Csibra, 2007; Southgate, 2013). Studies with adults revealed that inferences are privileged in the understanding of novel, unfamiliar or improbable actions (such as operating a light switch with the knee), when motor simulation is more useful in other situations (Brass et al., 2007). In fact, some researchers even think that the development of children's understanding of goal-directed actions might primarily emerge through repeated observations of actions and their perceptual effects (e.g., Hunnius & Bekkering, 2014). In fact, during their first year of life, infants learn associations between actions and produced effects while observing others perform such actions (Paulus et al., 2012) and while performing such actions themselves (Verschoor et al., 2013). More specifically, previous studies have shown that infants perform predictive gaze-shifts (see footnote n°52) toward goals of familiar human agents and actions before they perform them toward goals of unfamiliar human agents or actions (e.g., Woodward, 1998, 1999; Adam et al., 2016; Cannon, & Woodward, 2012; Henrichs, Elsner, Elsner, Wilkinson, & Gredebäck, 2014). The measure of predictive gaze-shifts with familiar agents or actions suggests that young children base their predictions on frequency information. For instance, Woodward (1998) revealed that 6-month-olds infants, who *repeatedly* saw a hand grasping the same object, anticipate that the hand will grasp again the object even when the spatial location of objects is rearranged.

Some studies measure the mu rhythm as a neural index of action prediction in childhood<sup>56</sup>. Monroy and colleagues (2019) found, measuring mu rhythm in toddlers' brains, that the motor system of 19-month-olds can use knowledge from statistical learning to predict upcoming actions. In their study, toddlers were trained at home with videos of unfamiliar action sequences giving prominence to statistical regularities (deterministic i.e., 100% predictable vs random actions). During the test phase, increased motor activity was observed preceding the deterministic but not the random actions, showing that infants could predict upcoming events based on statistical regularities learned through observation. In other words, it seems that the infant motor system can use knowledge from statistical learning to predict upcoming actions and thus, that motor simulation is not totally independent from the observer's prior beliefs or assumptions emphasized in the Simple View.

However, infants do not consider any end state as the action effect. Indeed, infants take into consideration action features such as the hand position (grasping versus touching in an unfamiliar way) to generate expectations. When observing a hand touching an object in an unfamiliar way (e.g., back-of-hand touching), infants only anticipate the same effect in a new arrangement if the action generates a new and salient effect (Adam & Elsner, 2018; Király, Jovanovic, Prinz, Aschersleben, & Gergely, 2003); otherwise, they do not generate expectation of the same ending in a new spatial arrangement (Woodward, 1999). For instance, 6- and 8-month-olds would form action-goal expectations only if a back-of-hand touching would dislocate the target object as a salient effect (Jovanovic et al., 2007; Király et al., 2003). In another study, Biro and colleagues (2014) found that 12-month-old infants could interpret a novel action as goal-directed even in the absence of an outcome, when that they had been provided with prior experience of that action accompanied with a salient visible outcome in another context but not when they had been provided with prior experience of that action but without the salient effect. In fact, the ability to predict an action based on observed

---

<sup>56</sup> A body of research studies the correlation between the measure of sensorimotor mu- and beta-oscillations attenuations and the generation of action prediction (Fox et al., 2016; Vanderwert et al., 2013; Braukmann, Bekkering, Hidding, Poljac, Buitelaar & Hunnius, 2017). Mu rhythm is considered as an index of predictive activity with a decrease in its power occurring in motor regions prior to the onset of an observed action in adults and infants (Marshall and Meltzoff, 2011; Southgate, Johnson, Osborne, & Csibra, 2009). As mu rhythm desynchronization is present even in infants at a different frequency (5-7 Hz), it suggests that the mu rhythm represents predictive activity in our mirror neuron system (Southgate et al., 2009; motor activation was previously demonstrated to be involved when an observer expect an agent to act; e.g., Caetano et al., 2007, Kilner et al., 2004, Southgate and Begus, 2013).

effects or to infer unseen movements from incomplete demonstration tends to improve by the end of the second year of life; before 19 months of age, toddlers usually fail to emulate unseen movements necessary to produce specific effects (Elsner, 2007).

#### 6.2.2.2 Teleological reasoning

According to the teleological reasoning approach, goals are typically inferred through the application of a rationality or efficiency principle relating actions, goals and situational constraints, such that observed behavior is interpreted as an action directed to a particular goal if it is judged to be the most efficient means available to the agent for achieving this goal in the given environment<sup>57</sup>. Gergely & Csibra (2007) claimed that infants use teleological reasoning based on a classic experiment by Gergely and colleagues (1995). In Gergely and colleagues (1995), 12-month-old infants were habituated to events in which the agent (a small circle) moves toward a goal (a large circle). Some infants saw a barrier blocking the most direct path such that the agent had to “jump” over the barrier. Other infants saw the same “jumping” motion without the barrier. Infants who were familiarized to the barrier-jumping condition looked significantly less at a new, direct path to the goal compared to the jumping-without-barrier condition; these results tend to indicate that they expect agents to be goal-directed and

---

<sup>57</sup> Studies with adults measured perceptual biases derived from the expectations related to the efficiency principle (Hudson, Nicholson, Ellis, & Bach, 2016; Hudson, Nicholson, Simpson, Ellis & Bach, 2016; Hudson, M., Bach, P. & Nicholson, 2018; Hudson, McDonough, Edwards & Bach, 2018). In Hudson, McDonough and colleagues (2018), adults were observing a hand starting to reach for an object with a straight or arched trajectory. The observed actions were either efficient (reaching straight when the path was clear or arched over an obstacle) or inefficient (straight towards an obstacle or arched over empty space). Before the movement was completed, the hand disappeared and participants were asked to report the hand’s last seen position. Results revealed that participants were biased by their expectations of efficient action: “Straight reaches were reported to have reached higher when an obstacle was blocking its path. ...Conversely, high arched reaches were reported lower when no obstacle was present, and corrected towards the straighter, more energy-efficient trajectory...” (Hudson et al., 2018: 2). In other words, it seems that the teleological reasoning is perceptually represented, providing a reference trajectory based on the efficiency principle which informs and even biases the action perception. Furthermore, McDonough, Hudson & Bach (2019) tested on what stimulus features these predictions of efficient action depend; they found that, when observing others’ behaviors, these expectations of efficient action depend on bottom-up cues to intentionality, which have been previously shown, in adults and children, to be related to the detection of two types of cues – the object’s semantics and its trajectory and motion profile – in the observed behaviors (e.g., Leslie, 1994; Rakison & Poulin-Dubois, 2001). Indeed, previous studies have shown a sensitivity to the object’s semantics (comparing a hand with a ball for instance; e.g., Avenanti, Annella, Candidi, Urgesi & Aglioti, 2012) and to motion cues (e.g., self-propulsion and change of direction, biological motion trajectory and speed of movement) which enable them to distinguish intentional agents from inanimate objects (Leslie, 1994; Morewedge, Preston & Wegner, 2007; Press, 2011). Results from McDonough, Hudson & Bach (2019) show that (1) these bottom-up cues modulate predictive biases towards efficient action in adult action observation and (2) these expectations of efficient action which impact the observer’s predictions emerge from the attribution of intention to the observed actions. In other words, there is a top-down influence which informs their perceptual representation of the observed action.

move in an efficient way. Other studies have shown that, from 10-12 months of age, children are able to predict and ascribe goals to agents even before their achievement (e.g., Csibra et al., 2003; Daum et al., 2008; Southgate & Csibra, 2009; Wagner & Carey, 2005) and to infer situational constraints from observed goal-directed action thanks to the efficiency principle (see, Csibra, 2008; Skerry et al., 2013; Gergely & Csibra, 2013). For instance, infants infer the presence of a barrier when an agent gets close to his goal jumping, or they expect agents to select the most efficient means available to achieve their goals. From early on, children attribute intentionality to observed behaviors which generates expectations following the action efficiency principle. For example, they are surprised when the agent they attributed intentionality to violates their expectations of action efficiency by not trying to avoid an obstacle (Gergely & Csibra, 2003).

In a study investigating the contribution of teleological reasoning and frequency learning to action prediction, Paulus and colleagues (2011) presented 9-month-old infants and adults with an agent who repeatedly walked to a goal while taking the longer of two possible paths – the shorter and more efficient path was impassable. In the subsequent test phase, both paths were passable; at the first test trial, both infants and adults anticipated that the agent would take the longer path but in successive trials, only infants kept anticipating movements to the longer path even though the agent took the more efficient path. In other words, it seems that frequency learning (means-goals associations) dominates action prediction and that teleological reasoning might gain strength later on. Again, these results show that motor experience is not necessary for the predictive tracking of action kinematics, and it suggests that infants may be able to use their extensive experience with observing others' actions to generate real-time action predictions.

I agree with Butterfill's Motor Theory of Goal Tracking that motor simulation is the main resource that young children use in order to ascribe goals to others and predict their actions. The Motor Theory of Goal Tracking explains how young children visually process and identify the agent's target (Target Tracking Guess) and then motor simulation allows children to retrieve actions that they have learned to be associated with that identified target. However, I think that this picture of the role of motor simulation in goal tracking should be nuanced in two ways.

First, which action is selected also depends on the probability of that action given the target object and the context. This suggests that general inferential processes (i.e., means-goals associations and teleological reasoning) of the kind postulated by the Simple View also to play a role in goal identification and may indeed condition the use of motor representations. In fact, Pomiechowska and Csibra (2017) found that “the involvement of sensorimotor cortices during action processing is conditional on a particular (instrumental) action interpretation, and that action interpretation relies on inferential processes and top-down mechanisms that are implemented outside of motor system” (ibid, 84). In this study, the researchers used an action-observation task with adults, assessing motor activation (as indexed by the suppression of the EEG mu rhythm) in response to actions typically interpreted as instrumental (e.g., grasping) or referential (e.g., pointing) towards an object. Results revealed that only the observation of instrumental actions results in enhanced mu suppression. In addition, exposure to grasping actions failed to elicit mu suppression when they were preceded by speech, suggesting that the presence of communicative signals modulated the interpretation of the observed actions. Furthermore, there is evidence that while interacting with others, other-relevant items are better encoded than non-task-relevant items (Joint Memory effect) and based on the emergence of that effect, task co-representation could also emerge for non-motor tasks (e.g., Elekes et al., 2016). For instance, Elekes and colleagues (2016) found an enhanced recall performance to stimuli relevant to the co-actor also when the participants’ task required non-motor responses (counting the target words) instead of key-presses. In another experiment, they found that direct visual access to the co-actor and his actions was found to be unnecessary to evoke the joint memory effect in case of the non-motor, but not in case of the motor task, meaning that prior knowledge of the co-actor's target category is sufficient to evoke deeper incidental encoding. In other words, these findings indicate that the capacity of task co-representation extends beyond the realm of motor tasks; simulating the other's motor actions is not necessary in this process.

Second, thinking of goal reasoning in the context of false-beliefs tasks raises the question of whether representations are tied to self and other (Southgate, 2020). Indeed, young children do not possess a fully developed self-concept (self-referencing develops around the age of two, see Rochat, 2003) but they show perspective-taking abilities earlier on age (Sodian, Thoermer and Metz, 2007;

Song and Baillargeon, 2008). Thus, do young children encode goals with a “subjective” component? One preliminary answer would be that young children allocate different representational strength to self and other representations. For instance, Southgate (2019) suggests that young children’s performance in false-belief tasks could be due to the spontaneous ability to take someone else’s perspective, the other-representation having more strength than the self-representation. If this is the case, studies should reveal that from early on age, infants privilege the other representation and do not act egocentrically (e.g., Buttelmann, Carpenter & Tomasello, 2009). I will discuss this question in chapter 7, but let’s consider the implication of reasoning about goals in the special case of joint action beforehand.

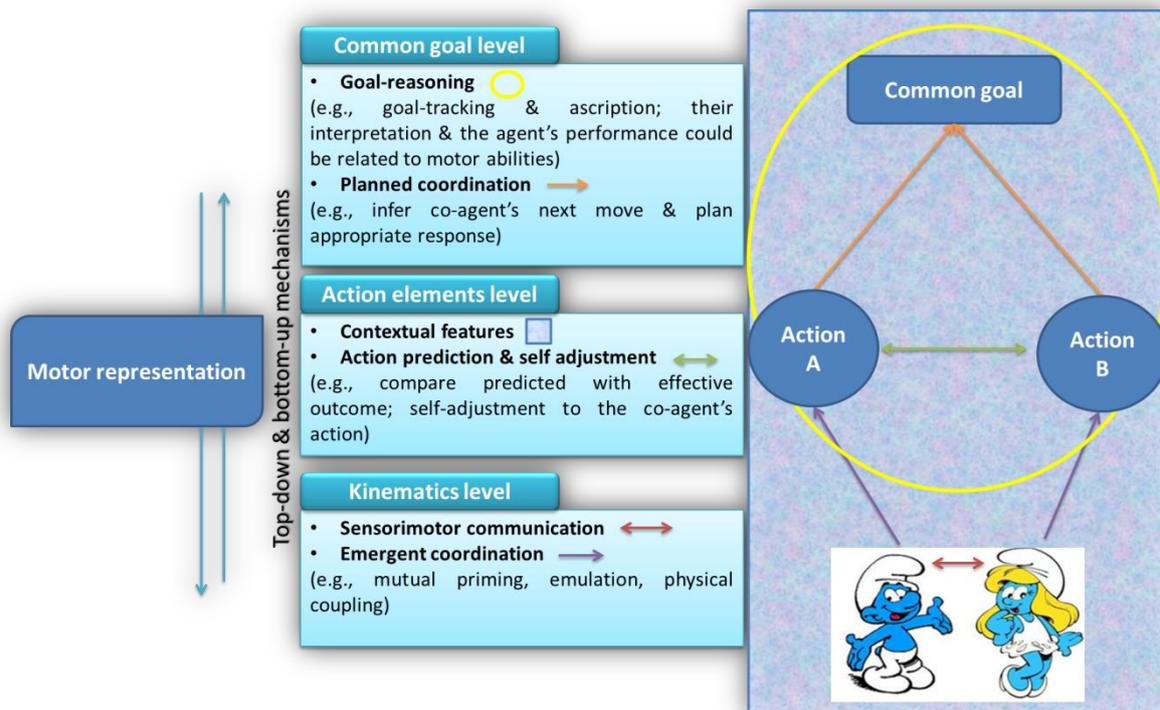
### ***6.3 Reasoning about goals in joint action***

In the case of joint actions, agents can rely on goal-reasoning for themselves and their partner. However, joint actions present two main specificities: (1) in joint action agents share a common goal and (2) joint action implies a specific form of interactivity allowing a more efficient coordination of co-agents.

Let us imagine two smurfs who decide to smurf together as a common goal. In order to smurf, they will need to respectively perform actions A and B. When the smurfs decide to smurf together, their joint action needs be analyzed at several levels. At the common goal level, the smurfs need to reason about what actions are needed to meet the common goal and how each of these actions, here actions A and B, contribute to the common goal in order both to perform their respective action and interpret their partner’s goal; they also need to make use of their planning resources to infer their partner’s next move and plan appropriate responses to it. At the action elements level, the smurfs need to include contextual features in their planning (e.g., situational constraints) and to rely on action prediction and self-adjustment so that actions A and B are performed in the most efficient way, given the context, to reach the common goal. Finally, at the kinematic level, the smurfs make use of sensorimotor

communication<sup>58</sup> and emergent coordination (mutual priming, emulation, physical coupling) to make their respective goals more salient and facilitate coordination.

Joint action can thus be described at three levels – a common goal level, an action elements level and a kinematic level – which work together thanks to top-down and bottom-up processes; motor representations being the point of convergence between these mechanisms (Cretu, Ruddy, Germann, Wenderoth, 2019). How do motor representations get involved in joint actions?



**Fig. 2:** Two smurfs smurfing described in the light of a three-level joint action model. All the levels work together via bottom-up and top-down mechanisms with motor representations as their convergence point.

### 6.3.1 Tracking goals in joint action

During their joint action performance, co-agents track each other’s goal. To do so, they may rely on motor representations, notably in the case of common goals requiring actions that involve motor processes such as lifting up a table or painting a wall together. In such cases, agents need to get coordinated and adjust their motor processes based on the incorporation of information about their co-

<sup>58</sup> Sensorimotor communication is described as a “natural form of communication that does not require any prior convention or any specific code...[which] amounts to the continuous and flexible exchange of bodily signals, with or without awareness, to enhance coordination success; and it is versatile, as sensorimotor signals can be embedded within every action” (Pezzulo et al., 2019).

agent (e.g., her strength to lift the table or her position in front of the wall), the environment (e.g., situational constraints such as if there is an obstacle to avoid or if the wall is outside and it is raining) and each other contribution to the goal performance which can impact, for instance, the action timing (e.g., synchronic contribution when lifting up the table together or diachronic contribution if co-agents divide up tasks, one painting the wall with an undercoat of beige and the other then adding another coat of blue paint on top of the undercoat).

In the case of abstract goals, such as celebrating a birthday or meeting at the airport, the advantage of joint actions is that agents can communicate verbally and motorically. Celebrating a birthday is a type of action that could be defined as the acknowledgement of a significant day (the date of birth of an individual) with a social gathering or enjoyable activity. The performance of this goal implies that agents explicitly agree on the realization of the goal and commit to the joint action (e.g., each commits to perform his/her part of the action by, for instance, by bringing a cake, booking a location, or simply coming to the surprise party). The common point between motor and abstract goals is that they both require the identification of the outcome(s) which the observed means are a best available way of achieving. Their difference is that some goals are more abstract than others and require an explicit level of agreement and commitment before being deployed at a lower-level in motor representations. In such cases, functional goal ascription is not sufficient; co-agents rely on representational goal ascription (see Part I). Compared with individual actions, joint actions with abstract goals could require a certain level of planning skills and the ability to understand the norms and commitments involved.

### ***6.3.2 Sharing motor representations in joint action***

Studies have shown that joint action is based on active prediction of the partner's action effects and that, there is an integration of motor representations of both the agent's and the partner's (predicted) actions as a *unitary motor representation, i.e., as a dual-person (dyadic) motor plan* (for a recent review, see Sacheli, Arcangeli & Paulesu, 2018; della Gatta, Garbarini, Rabuffetti, Viganò, Butterfill and Sinigaglia, 2018). For instance, della Gatta and colleagues (2018) adapted a bimanual paradigm

(the circle-line drawing paradigm<sup>59</sup>) to contrast two agents acting in parallel with two agents performing a joint action. Participants were required to draw unimanually lines or circles while observing circles or lines being drawn by an experimenter. The only difference between the parallel and the joint condition was that, in the joint condition, participants were instructed to perform the task together with the experimenter, as if their two drawing hands gave shape to a single design. In the joint condition, the researchers found an ovalization of the participant's drawing suggesting that participants in joint actions can represent collective goals motorically<sup>60</sup>. However, this ovalization was absent in the parallel condition. These findings indicate that an interpersonal motor coupling effect may occur in joint but not in parallel action.

In the case of joint action, co-agents also use sensorimotor communication to make their action goals more salient for their co-agent, notably via specific kinematic cues in order to be more predictable for each other (for a recent review, see Pezzulo et al., 2019; see also, Vesper et al., 2010; Pezzulo, 2011a; Pezzulo and Dindo, 2013; Sacheli et al., 2013; Candidi et al., 2015). In fact, agents tend to produce specific kinematic cues in the context of joint action. In a recent study, kinematic cues oriented toward action prediction were shown to be exclusively present in the joint action context compared to a teaching context in which kinematic cues were optimized to orient the learner's attention (McEllin, Knoblich and Sebanz, 2018). In this study, McEllin and colleagues (2018) used a xylophone playing task revealing that agents modulate their movement height differently to support spatial (in the ascent of the movement) and temporal (in the descent of the movement) predictions in

---

<sup>59</sup> In this bimanual paradigm, participants were instructed to draw either lines or circles with their left hand and the other drawing with their right hand. Previous studies found that, when an individual is bimanually drawing lines and circles, the line drawing hand tends to ovalize its trajectories (Piedimonte et al., 2014; the baseline experiment in della Gatta and colleagues, 2018). The ovalization in the individual performance in a bimanual action is interpreted as a motor coupling effect i.e., a consequence of representing the goal of drawing both a circle and a line.

<sup>60</sup> "We suggest that ovalization in the Joint Action condition has fundamentally the same source as ovalization in individual performance of a bimanual action: it is a consequence of representing the goal of drawing both a circle and a line...Representing this goal would trigger motor processes in the participant concerning both circle and line drawing actions, somewhat like those which would occur where the participant performing both drawing actions herself. These motor processes should interfere with each other, somewhat as they do in bimanual action. Of course, in joint action the hands belong to different individuals: this may explain why the interference is stronger in bimanual action than in unimanual joint action. But the critical point for us is that in both cases, bimanual action and joint action, interference is a consequence of representing motorically the goal of drawing both a circle and a line" (della Gatta et al., 2018: 59).

joint action coordination. During a joint action, from one to the other, sensorimotor communication is as important as data gathering abilities enabling co-agents to organize visually their environment and to be able to reason probabilistically (Bernadyn & Feigensohn, 2018).

The coordination required by joint action can rely on the use of a unitary motor representation (dual-person motor plan) and on the interactivity of co-agents who communicate at different levels to achieve their common goal. This interpretation of the joint action phenomenon is thus compatible with the Motor theory of Goal Tracking. However, what about more abstract joint actions? How do these elements help in categorizing joint actions?

### ***6.3.3 Categorizing joint actions***

Our discussion of the use of goal representations (instead of mental representations) and of the use of motor representations in goal reasoning suggests a first criterion for distinguishing joint actions based on the representational nature of the shared representations they require.

#### *6.3.3.1 Joint action level of complexity: abstractness and advance planning*

Joint actions can vary based on their level of complexity. The complexity of a joint action is a function of the degree of abstraction of their goals and of the amount of advance planning their successful performance requires. More concrete goals can typically be encoded and tracked motorically. In contrast, encoding more abstract goals requires further representational resources and ascribing these goals to others involves attributing them representational mental states, thus demanding robust theory of mind skills. In addition, the successful performance of a joint action may require varying degrees of advance planning. In some situations, little or no advance planning is needed as the situation or familiarity with the action make it clear how the joint action should proceed and what the division of labor should be and emergent processes are sufficient to insure coordination. In other cases, advance planning and deliberation are necessary to devise a plan of action, distribute roles and ensure that co-agents are coordinated. While these two dimensions of complexity often go together, they can also come apart.

I will now describe three types of joint actions based on the action complexity. Highly complex joint actions are characterized by the abstractness of their goals and the need for advance planning. For

instance, playing Kemps card game can be considered as a highly complex action as the goal of the partner is quite abstract (e.g., the partner's goal can be to confuse the other team by using fake signals to make the other team yell "COUNTER-KEMPS" or by occasionally picking up cards you do not need; the point is to be able to dissociate when the partner's goal is to confuse the other team from when it is to send signals to you) and advance planning is needed in order to win the game (e.g., if I see that my partner has progressively taken three jacks, I want to help him/her to be discreet in his/her strategy to prevent the other team from saying "COUNTER-KEMPS" so I take the last jack, wait for the other team to be distracted and play the jack so that he/she can get it and I can say "KEMPS"). Joint actions of low complexity are characterized by the concreteness of their goal and low demands for advance planning. For instance, lifting up a table together constitutes an example of a joint action of low complexity, as agents use cues from the situation and the interaction itself to adjust to each other (e.g., each placing herself at the opposite side of the table, mirroring each other in their posture and grasp of the table). However, note that there can be joint actions of intermediary complexity since abstractedness and advance planning can be considered as continuous variables. For instance, celebrating a birthday, which is a quite abstract goal, can require more or less advance planning based on the agreement between the participants; moving a car which suddenly broke down together, which is a quite concrete goal, can require more or less advanced planning skills based on the context and the access to the reliable information about the partner while performing the joint action. In other words, there are joint actions which are not categorized by either high or low abstractedness and advance planning.

Importantly, this first criterion takes into account the fact that complexity is relative to the agents involved, meaning that the more the action is performed by an agent the more effortless it will become to perform for herself. For instance, regular Kemps card game players won't need to engage in much advance planning as they will probably have built relevant play routines.

#### *6.3.3.2 Joint action level of complexity modulates the nature of the shared representations*

My hypothesis is that the nature of the representations shared by co-agents will depend on the complexity of the joint actions. Highly complex joint actions (e.g., Kemps card game) require the

representation of mental states whereas joint actions of low complexity (e.g., lifting up a table) require the representation of goals. Thus, joint actions of low complexity mainly rely on the use of motor representations. What about joint actions of intermediary complexity? I think such intermediary complex joint actions are cases where there may be no need to appeal to mental states but that the motor representations involved must be supported by set of sufficiently developed further cognitive abilities.

What are these further cognitive abilities and when do they appear? First, children need to develop a richer understanding of goals. There is evidence that around nine to twelve months of age, infants develop an understanding of goals as abstract relations that organize physical actions. Understanding goals as abstract relations probably consolidates and favors the use of other processes than motor simulation in their reasoning about means-outcomes relations. For instance, they use looking events toward a focus of attention (e.g., an object) to condition their expectations about observed agent's instrumental actions. They can notably expect that an agent will act on the object she just attended to (Phillips et al., 2002). Young children also show the ability to represent goals requiring the performance of a sequence of actions. For instance, they understand an action on an intermediary object (e.g., a box or a tool) as directed to the access of another object (e.g., a toy inside the box or an object out of reach) rather than at the box or the tool itself (Woodward & Needham, 2008; Sommerville et al., 2008). Second, they must develop more powerful predictive processes. Developmental evidence shows that from twelve months of age infants use intuitions about probabilities to shape expectations about the future (Bonatti, 2008; Téglás et al., 2007) and that young children use this predictive reasoning about other agents' beliefs, goals and behaviors (Woodward, 1998; Onishi & Baillargeon, 2005; Kovács, Téglás & Endress, 2010). For instance, by twelve months of age, children take into account the knowledge or ignorance of the agent (e.g., whether she perceived or not that the object she is looking for felt down) and gesture appropriately to provide information (e.g., point to the object actual location) necessary for the agent's goal performance (Liszkowski, Carpenter & Tomasello, 2008).

Third, children also need to develop their abilities to plan proactively to accommodate their partner's actions and to use norms. These abilities develop later in childhood. A body of research

studying children's abilities to plan their actions in a joint action context with situational constraints shows that planning abilities get significantly better from 3½ years of age with a big improvement at 5 years of age (e.g., Meyer et al., 2016; Paulus, 2016). For instance, in Meyer and colleagues study (2016) with children aged from 2½ to 5, children engaged with an adult experimenter in a joint cup-stacking. They were sat opposite each other at a table and separated by a window frame requiring children to reach around the frame on one of its sides to hand over the cups. Children passed cups to their partner who had only one hand available. Children's response choices were assessed (i.e., whether they passed the cup on the free or occupied side to their partner). They found that only by 3½ children started significantly handing the cup more to the experimenter's available hand. Also, in their third year of life, children integrate normative factors (25-27 months of age, see Warneken & Tomasello, 2007; Gräfenhain et al., 2009), they use gestures and verbal indicators to coordinate (30 months, see Hamann et al., 2012). For instance, around three years of age, children start understanding that they are committed with their partner to a joint activity and try to reengage the partner if she stopped playing (Warneken & Tomasello, 2007).

My first criterion – the joint action complexity – constitutes the first axis of the joint action matrix (the second axis will be presented in chapter 7). The nature of the representation shared between co-agents is modulated along this axis. Joint actions of low complexity rely on shared motor representations when highly complex joint actions require shared mental representations (e.g., shared intentions). Joint actions of intermediate complexity mainly require the use of motor representations but they also integrate other skills (e.g., planning skills, norms integration, common knowledge) that emerge later in childhood and are presumably also used by adults.

Reasoning about goals demands that one takes into account bottom-up information (e.g., subtle kinematics present in the agent's action, situational constraints or dynamic features of the action) as well as apply top-down information (e.g., means-goal associations, knowledge about the agent, prior beliefs or assumptions about norms) to the specific situation in order to represent the agent's goal. I discussed two main positions (the Simple View and the Motor Theory of Goal Tracking) regarding the *processes* – general inferential processes about means-goals associations or motor simulation – involved in tracking minimal goals (notably under nine months of age). I agreed with the Motor

Theory of Goal Tracking about the fact that motor simulation is the main resource that young children use in order to ascribe goals to others and predict their actions. However, I added to this position, (1) the idea that the activation of a motor representation of an action in the observer also depends on the probability of that action in the given context and (2) that motor representations may involve a “subjective” encoding component (self for the observer herself and other for the observed agent) whose influence on experience is indexed by the self-concept development in the observer. On the basis of this discussion, I proposed a first criterion for the categorization of joint action, based on their level of complexity (itself characterized in terms of abstractness and demands in advance planning). In other words, I propose that the nature of the representation shared between co-agents is modulated by the joint action complexity.



## Chapter 7

# *Does co-agents' strategy to maximize their performance impact the sharedness characteristics of shared representations?*

In chapter 6, I focused on how the complexity of their goals affect the nature of the shared representations needed for joint action. Collective goals can be assessed as more or less complex based on two criteria – abstractness and demands for advance planning. Consequently, three types of joint actions can be distinguished: low complexity (low abstractedness and low advance planning), intermediary complexity (high abstractedness and low advance planning or the reverse) and high complexity (high abstractedness and advance planning) joint actions. The hypothesis proposed in chapter 6 is that the complexity of joint actions determines the nature of the representations shared by co-agents. More specifically, joint actions of low and intermediate complexity only require co-agents to share representations of goals (via motor representations and/or inferential processes) whereas highly complex joint actions require the representation of mental states.

When co-agents share representations of goals, studies have shown that there is an integration of motor representations of both the agent's and the partner's (predicted) actions as a *unitary motor representation, i.e. as a dual-person (dyadic) motor plan* (for a recent review, see Sacheli, Arcangeli & Paulesu, 2018). However, how do the agent's and the co-agent's motor representations get integrated in this unitary motor representation? This question can be decomposed in two sub-questions. First, how do co-agents *share* representations? Second, what is the impact of relying on a *dyadic* motor plan? To answer these questions, one should investigate, always relatively to the type of joint action considered, (1) what strategy is used by the agent to integrate the motor representation of his partner, (2) whether there is a representational difference between self and other representations in

their integration and (3) whether this difference is related to the co-agents' adopted strategies (e.g., identical or complementary) in a joint action.

Studies have shown that the integration of their co-agent(s) in their own planning can facilitate or interfere with the agent's performance due to co-representation effects (e.g., Sebanz, Knoblich and Prinz, 2003; Samson, Apperly, Braithwaite, Andrews and Scott, 2010). Co-representation effects have not been measured in children before age 4 (e.g., Milward, Kita & Apperly, 2014), which could suggest that children need to have a fully developed theory of mind to perform truly 'joint' actions (and not merely coordinated individual actions). However, if children do not co-represent their partner during a joint activity, they should not be able to complement her action in any type of joint activity before the apparition of a robust theory of mind. Analyzing puzzle-solving tasks (Chapter 2, 2.1.2.3) revealed that children's ability to perform complementary joint tasks (i.e., ability to complement their partner's action) vary based on the skills – problem-solving, flexibility, action planning skills, motor skills – required by the task itself to successfully complement the partner's action (these skills develop progressively during childhood), meaning that theory of mind may not be necessary to perform joint actions. This analysis is interesting in view of other studies that have shown that co-agents' strategies depend on the trade-off between the *task requirements* and the *uncertainty of the information available about the partner* and that the speed at which they develop stable strategies and the optimality of these strategies depend on the amount of information uncertainty (for a recent review, see Chackochan & Sanguineti, 2019). Indeed, in contrast with the interferences linked to the presence of co-representation effects, other studies have shown that co-agents tend to adopt specific strategies (e.g., identical or complementary) which maximize their performance and can even lead to a better performance jointly than individually (e.g., faster, more accurate; for a recent review, see Pezzulo et al., 2019). It could be that the adoption of such strategies is a solution to deal with co-representation effects, notably regarding the information available about the partner, while integrating their co-agent(s) in the agent's own planning. The aim of this chapter is to develop this hypothesis, that is, that the adoption of strategies by co-agents as a way to maximize co-agents' performance is relative to perspective-taking abilities. In the case of identical strategies, co-representation effects may be facilitating the joint action performance while in the case of complementary strategies, co-agents need to avoid interference

created by co-representation effects to maximize their performance – these two types of cases require different levels of perspective-taking abilities which will be described in this chapter. The strategies used by co-agents to maximize their performance form the second axis of the proposed joint action matrix.

### **7.1 The co-agents' self- and other-representations in joint action**

Co-representation effects, mostly referred to as interference effects in the literature about the joint Simon effect<sup>61</sup>, are interpreted as evidence of the emergence of a representation of the co-agent or as an index of emergence of shared representations (e.g., Sebanz et al., 2003, 2006; Knoblich and Sebanz, 2008) but they could be, more generally, interpreted in terms of self-other integration (Ruissen and de Bruijn, 2016) in a unitary motor representation. Co-representation effects emerge under some circumstances intrinsic to the joint action phenomenon (see Appendix 1) due to the contextual congruency (spatial or perspectival) between co-agents, leading to facilitation or interference effects in performance. For example, in contextual congruency, studies showed that people are faster to execute an instructed action if that action is congruent with an action afforded by an object reachable by someone else, suggesting that they spontaneously encode the object's affordance for the other (Costantini et al., 2011). In contrast, the observation of others' actions can give rise to representations that interfere with one's own task performance. For instance, Brass and colleagues (2000) found that participants who were instructed to produce finger movements in response to symbolic cues responded more quickly when simultaneously observing irrelevant finger movements that were physically congruent to the ones they were instructed to produce, and more slowly when simultaneously observing irrelevant finger movements that were physically incongruent to these. The emergence of co-representation effects might be a sign that there is an overlap between the cognitive systems responsible for representing self- and other-perspectives<sup>62</sup>. This may, on the one

---

<sup>61</sup> For more details about interference effects, see Chapter 2 (notably the footnote n°22 about the joint Simon effect) and the Appendix 1.

<sup>62</sup> From early on, humans seem to rely on shared mechanisms between self and other to act and to interpret information about the world. A large body of research has studied self-other shared mechanisms involving low-level cognitive processes as, for instance, in the motor domain (e.g., Decety & Sommerville, 2003) as well as

hand, contribute to a perspectival confusion between self and other especially in young children (Steinbeis, 2016) and, on the other hand, motivate the need for additional mechanisms to dissociate self from other representations (Kampis & Southgate, 2020).

The variety of joint actions leads to contexts that are more or less congruent between co-agents – modulating the *uncertainty of the information available about the partner* (Chackochan & Sanguinetti, 2019) and requiring that co-agents adopt specific strategies (for a recent review, see Pezzulo et al., 2019). Indeed, it is probable that co-agents adopt identical forms of interaction when they are placed in a congruent situation and have a good access to information about their partner. In contrast, it is probable that co-agents adopt complementary forms of interaction when they are placed in an incongruent context and have poor access to information about their partner. Since congruent and incongruent contexts respectively lead to the emergence of facilitation or interference in the agent's performance due to co-representation effects, (1) they might not require the same control over self and other representations and (2) they probably require that co-agents adopt different strategies and roles that maximize their performance.

Some joint action contexts might be more congruent than others. On the one hand, it is probable that complementary forms of interaction in joint action probably lead to more contextual incongruency than identical forms of interaction. This suggests that there should be more interference effects in the case of joint actions requiring co-agents to perform complementary actions (compared with those requiring them to perform identical agents) and that co-agents would need to have control over self and other representations so that they can maintain or confront self and other representations to adopt such complementary roles. On the other hand, co-agents try to maximize their performance and thus, based on their abilities and the context, they might adopt the most favorable strategy to achieve the common goal at stake.

---

involving more high-level cognitive processes as in the linguistic and memory domains (e.g., Forgács et al., 2019, 2020; Howard & Woodward, 2019). More specifically, in the motor domain, such shared mechanisms coding for self and other actions has been shown to be involved both in action execution and action observation and prediction (e.g., Southgate & Verneti, 2014; Brass et al., 2001; Kampis et al., 2015).

### ***7.1.1 Identical and complementary forms of interaction***

While performing joint actions, co-agents adopt identical or complementary forms of interaction.

In the case of identical forms of interaction, co-agents are placed in a context where they have access to information about their partner and they can readily synchronize each other's performance by aligning their representations. For instance, when co-agents lift up a table together, co-agents both have visual access to the object they will manipulate and to their partner in front of them. In such a case, they can even imitate each other in order to grasp and lift the table in synchrony via the integration of motor representations<sup>63</sup>. Previous studies, measuring mimicry, priming and automatic imitative actions, have shown that humans resonate with one another efficiently, (Heyes, 2011). Co-agents might continuously influence each other via their actions, where these actions function as by-product signals sensed by their partner (without any intentional communication or awareness), and the interaction dynamics itself can provide elements such as affordances<sup>64</sup> (e.g., Kelso et al., 2013; Tucker & Ellis, 2004) to help co-agents mutually predict each other, mutually align their internal representations and synchronize (Friston and Frith, 2015; Koban et al., 2019; Pezzulo, 2013a). In this perspective, basic forms of social coordination, alignment and synchronization emerging within the interaction itself give rise to a coupling phenomenon as in other physical phenomena (e.g., coupled oscillators or pendulums; see Schmidt et al., 1990, Bennett et al., 2002)<sup>65</sup>. When adopting identical

---

<sup>63</sup> Synchronization can be simultaneous as in the case of lifting up a table but it can also be diachronical. For example, two pianists can synchronize their performance to each other's timing with asynchronies of 30 ms (diachronically; Keller et al., 2007).

<sup>64</sup> Agents can use sensorimotor communication (for a recent review, see McDonough, Costantini, Hudson, Ward & Bach, 2020). For instance, co-agents identify affordances offered by or as an effect of their co-agent's actions; the identification of affordances result from the elicitation of motor programs adequate to interact with such objects (e.g., the observation of graspable objects eliciting grasping actions, Tucker & Ellis, 2004). Furthermore, affordances can be adaptable since other agents have an impact on their environment. For instance, heavy objects can be perceived as lighter in the presence of other individuals. The adaptability of affordances in the interaction with others is typical of teamwork (Bacharach, 2006). However, there are cases in which co-agents do not need to use sensorimotor communication. For instance, Vesper and colleagues (2013) tested coordination strategies in an asymmetric task whereby dyads were asked to coordinate jumping actions of different distances (i.e., by landing in temporal synchrony) without being able to see or sense their partner but informed of their partner's jumping distance before each trial. Researchers found that co-agents were able to coordinate by adopting complementary strategies, i.e., the person with the shorter jump adapted their actions to their partner's task demands by waiting longer before jumping, while the person with the longer jump adapted less.

<sup>65</sup> Note that such a description could also apply to cases whereby two or more individuals execute the same or similar actions in parallel without joint goals such as walking side-by-side (van Ulzen et al., 2008), coordinating finger movements (Oullier et al., 2008), lifting glasses from each other's trays (Pezzulo et al., 2017) or rocking

forms of interaction, co-agents might be placed within congruent situations, which lead to the emergence of facilitation due to co-representation effects. In addition, it could be that in such situations co-agents' representations are not tied to self and other, meaning that co-agents would not need to dissociate self and other representations. In other words, while adopting identical forms of interaction, the alignment of co-agents' motor representations might afford an agent-neutral stance (i.e. a stance that does not distinguish self and other contributions in the joint plan), as if co-agents' representations were superimposed. In this case, co-agents' spontaneous tendency to take the other perspective might be sufficient to allow the performance of such joint actions and even facilitate it. If this is the case, (i) studies should show that co-agents' performance is better jointly than individually in simultaneous joint actions compared with sequential ones since co-agents' performance is related to the interaction itself (it is the case, see Pezzulo et al., 2017); (ii) spontaneous perspective-taking abilities might be sufficient to perform such kind of joint actions. In fact, previous studies point to the connection between spontaneous perspective-taking abilities and the participation in a joint action (Furlanetto et al., 2013; Mazzarella et al., 2012; Tversky & Hard, 2009; Freundlieb, Kovács, & Sebanz, 2016). More specifically, Freundlieb and colleagues (2016) found indeed that “spontaneous VSP [visuospatial perspective]-taking can effectively facilitate and speed up spatial alignment processes accruing from dynamic interactions in multiagent environments” (ibid, 401).

In contrast, in the case of complementary forms of interaction, co-agents need to complement each other's actions. In order to do so, they might maintain or confront self and other representations<sup>66</sup> to adjust properly in the joint action context. The incongruency of complementary forms of interaction should lead to interference due to co-representation effects (e.g., motor resonance here is not suitable; for a review, see Knoblich, Butterfill, & Sebanz, 2011). Each agent has a representation of the whole joint action but must also maintain representations of self and other actions as separated blocks, which imbricated one with another, enable the performance of the joint action in its entirety. The

---

next to each other (Richardson et al., 2007). However, the difference with joint actions lies in the fact that co-agents form joint goals and maximize their joint performance.

<sup>66</sup> It is important here to dissociate the need to maintain self and other perspectives from the need to confront self and other perspectives which are described as two separate levels of perspective-taking abilities (e.g., Moll & Meltzoff, 2011a,b); I will detail the different levels of perspective-taking abilities later in this chapter.

maintenance or confrontation of such self and other representations could be a way for co-agents to keep in mind the purpose of their joint action as a whole and to adjust based on other factors (e.g., different perceptual or informational access, situational constraints). For instance, two agents decide to move a car which suddenly broke down, one agent pushes it while the other agent sits behind the wheel, steering and applying the brakes as necessary; the two agents complement each other's actions, i.e., each has a part to play in the joint action performance and both need the involvement of their co-agent to perform the joint action. The agent who sits behind the wheel needs to take into consideration the fact that the agent who pushes the car does not have the same perceptual access to the road, the maintenance of self and other representations allows her to adjust and apply the brake if necessary for instance. The co-agent who seats behind the wheel needs to identify how her co-agent who pushes the car perceives the road to act upon her representation. Another example could be painting a wall together where two co-agents need to complement each other's actions. In contrast with the broken car example, co-agents need to confront self and other representations to adjust appropriately regarding the wall they are painting together. In this case, co-agents both need to understand that the wall may simultaneously look different from different perspectives. For instance, it could be that the sunlight reflection on the wall makes a part of it appear already painted when it is not; the co-agent who perceives the wall correctly is then able to paint it or to indicate to her partner who is not aware of the sunlight reflection effect that she should paint it. In complementary cases, the spontaneous tendency to take the other perspective is not sufficient to allow the performance of such joint action. In fact, co-agents need to exert cognitive control to dissociate self from other representations and adjust properly to each other. In joint actions whereby co-agents adopt complementary forms of interaction, the alignment of co-agents' representations has the form of an imbrication. While adopting complementary forms of interaction, co-agents need to maintain or confront self and other representations, meaning that their representations need to be tied to self and other. In complementary forms of interaction, co-agents probably encounter an incongruent context leading to uncertainty about the information available about the partner and the emergence of interference effects. In this case, it could be that (1) co-agents need to be able to avoid such interference in order to maximize their performance as they do not need to imitate their partner but rather to complement his/her action

relatively to the common goal; (2) it takes more than spontaneous perspective-taking abilities to perform such kind of joint actions.

### ***7.1.2 Co-agents' performance maximization strategies in joint action***

Agents show co-representation effects resulting from the co-representation of their partner, which impacts their performance during joint action. If such effects do affect their performance, it seems probable that agents adopt roles which better suit the joint action in order to improve their performance as a function of context congruency (leading or not to the emergence of interference effects). Previous studies have shown that a dyad – two agents performing a joint action<sup>67</sup> – can perform better together than one agent performing the same task alone<sup>68</sup> (Reed et al., 2016). In their joint action, co-agents adopt strategies which maximize their performance and conversely, minimize errors and efforts for each other. In fact, the strategies adopted by co-agents are not determined arbitrary but notably as a function of (lack of) perceptual access to the co-agent's actions (Curioni et al., 2019). Co-agents also readily represent the difficulty of their co-agent's potential actions, plan joint actions to accommodate the co-agent's difficulty and use first hand motor experience to enhance the response selection and planning of joint actions (Ray, de Grosbois & Welsh, 2017).

On the one hand, incomplete or unreliable information about the partner slows down the speed at which co-agents perform the action and creates an interactive strategy alternating leader-follower roles<sup>69</sup>. In such cases, leaders often adapt their action kinematics in communicative ways in order to signal their roles, convey task information to followers (Candidi, Curioni, Donnarumma, Sacheli & Pezzulo, 2015; Pezzulo, Donnarumma & Dindo, 2013; Sacheli, Tidoni, Pavone, Aglioti & Candidi,

---

<sup>67</sup> In the literature, researchers studied contexts in which dyads share a common goal via the control of isometric force (Masumoto & Inui, 2013, 2014), the reaching for the same fixed (Reed et al., 2006; Takagi, Beckers & Burdet, 2016) or moving target (Takagi, Usai, Ganesh, Sanguineti & Burdet, 2018; Ganesh, Takagi, Osu, Yoshioka, Kawato & Burdet, 2014; Takagi, Ganesh, Yoshioka, Kawato & Burdet, 2017), or the operating of a tool (van der Wel, Knoblich & Sebanz, 2011).

<sup>68</sup> It could be that dyads perform better together than individually because they share efforts (i.e., less motor effort might lead to less motor noise and errors, O'Sullivan, Burdet & Diedrichsen, 2009) or they are more accurate since they share their estimation of external events (Ganesh, Takagi, Osu, Yoshioka, Kawato & Burdet, 2014).

<sup>69</sup> The co-agent is considered as a leader if she initiates the action and provides most effort (Masumoto & Inui, 2014), she sets the pace and requires that the co-agents adjust more to her than the opposite (Pezzulo, Donnarumma, Dindo, D'Ausilio, Konvalinka & Castelfranchi, 2018; Curioni et al., 2019).

2013) until followers seem to understand the rules (Pezzulo & Dindo, 2011; Candidi et al., 2015)<sup>70</sup>. The leader's emphasized action kinematics is meant to be used by the follower as a reference point for her to align her behavior. For instance, let's consider two drummers playing together; one of the drummers could emphasize his beat to give a reference point to the other drummer for his movements. Conversely, more complete and reliable information leads to more synchronous behaviors which can be qualified as identical forms of interaction (see Fig. 3 below). It could be that factors such as familiarity with the co-agent and a congruent environment favor the better performance of a joint action context. In fact, studies have shown that joint training improves subsequent individual performance (Ganesh, Takagi, Osu, Yoshioka, Kawato & Burdet, 2014).

On the other hand, a complementary strategy emerges spontaneously when co-agents are confronted with a joint action which might be expected to complicate motor planning and efficient task execution (for a recent review, see Pezzulo et al., 2019; Reed et al., 2006; Masumoto & Inui, 2013). For instance, Reed and colleagues (2006) tested haptically coupled dyads (co-agents' motions were physically linked) in a target acquisition task. In the coupled dyad condition, an agent's limb motion responded to the output of two motor-control systems (namely, the agent's own and his/her partner's). Researchers found that co-agents adopted a complementary strategy (one agent controlling acceleration and the partner controlling deceleration). The adoption of a complementary strategy enables coupled dyads to perform better (i.e. faster) in the task in comparison with their individual performance. As Pezzulo and colleagues (2020) puts it: "strategy flexibility may emerge as a stable solution to interactive problems, where two complementary strategies can afford better coordination than two identical strategies" (14). In fact, some findings provide partial support for the hypothesis that visuomotor interference effects can be reduced when two physically incongruent actions are

---

<sup>70</sup> Candidi and colleagues (2015) similarly to Pezzulo and Dindo (2011) found that leaders stop sending coordination signals when they understand that followers understood the "rules" they try to convey. For instance, in Pezzulo and Dindo (2011), co-agents (a Builder who leads and a Helper who follows) were asked to jointly build a tower made of colorful bricks but only the Builder knew the tower design (e.g., red blocks first, blue blocks later). In this study, the Helper picks up a block and passes it to the Builder; together, they lift and place the block together on top of the tower (blocks are heavy and require to be lifted together). The Builder can signal (e.g., by pointing) the correct color of block so that the Helper can align his representation of how the tower should look like. For instance, if the tower should be composed of red blocks and blue blocks at the top, the Builder will stop pointing at the red blocks once he inferred that the Helper understood the rule.

represented as mechanically interdependent contributions to a joint action goal (Clarke et al., 2019). In other words, complementary roles can be considered as optimal forms of interaction which minimize the need to take heed of the partner when it enables co-agents to maximize their performance.



**Fig. 3:** The strategies adopted by co-agents based on the completeness and reliability of information about the partner they have access to. In red, incomplete and unreliable information about the partner leads to complementary strategies. In yellow, incomplete or unreliable information about the partner leads to leader-follower strategies. In green, complete and reliable information about the partner leads to identical strategies.

In sum, it is not that some joint actions systematically require the adoption of identical or complementary forms of interaction but rather that co-agents adopt (without being aware of it) the strategy or roles which maximize their performance. The variety of joint actions leads to contexts that are more or less congruent between co-agents; it is probable that complementary forms of interaction are more frequent in incongruent configurations and identical forms of interaction are more frequent in congruent configurations.

The variety of joint actions leads to contexts that are more or less congruent between co-agents. The congruency of the context modulates the quality (completeness and reliability) of the information available about the partner but also requires adopting specific strategies either to emphasize or reduce the impact of co-representation effects within performance. The more congruent a joint action context is, the more co-agents will adopt identical strategies which emphasize facilitation effects due the spontaneous ability to take their partner's perspective. In contrast, the more incongruent a joint action context is, the more co-agents will adopt complementary strategies which reduce interference effects thanks to the dissociation between self and other perspectives (maintenance or confrontation). In the next section, I will introduce different levels of perspective-taking abilities which appear at different

ages during childhood, and I will explore the relationship between the emergence of perspective-taking abilities and the strategies at stake in joint action performance.

## **7.2 The integration of self and other representations in childhood**

As mentioned in chapter 2 (2.1.2.2 and footnote n°22), previous studies have shown that co-representation effects do not emerge before four years of age (Milward et al., 2014<sup>71</sup>). However, if children do not co-represent their partner during a joint activity, they should not be able to complement her action in any type of joint activity before the apparition of a robust theory of mind. In view of the different tasks used to test children's joint action abilities (see Chapter 2, 2.1.2), it seems that children's ability to perform complementary joint tasks vary based on the skills – problem-solving, flexibility, action planning skills, motor skills – required by the task itself to successfully complement the partner's action. As children complement their partner's action before the apparition of theory of mind abilities, I suggest considering perspective-taking abilities (and its different levels) as a good candidate to support the integration of the partner's representation in a dyadic plan.

### ***7.2.1 Visual Perspective-taking abilities***

A body of research has focused on visual perspective-taking (VPT) abilities and more specifically, on how agents represent the visual perspectives of others when such perspectives differ from their own (Flavell, Everett, Croft, & Flavell, 1981; Keysar, Lin, & Barr, 2003; Samson, Apperly, Braithwaite, Andrews, & Scott, 2010). Taking someone else's perspective can be defined as the act of viewing a scene or understanding from an alternate point-of-view<sup>72</sup>. For instance, in a joint action context (e.g., a simple team game), agents represent their partner's point of view and in fact, studies have shown that they perform better at playing the game when their partner's point of view does not differ from their own (Surtees et al., 2016). Furthermore, the literature on perspective-taking abilities draws a key

---

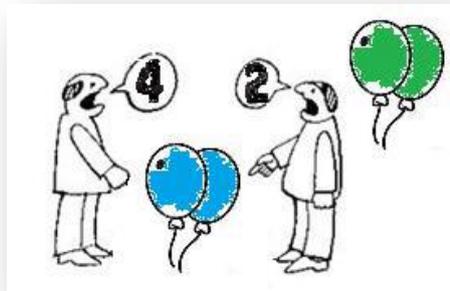
<sup>71</sup> As a reminder, Milward, Kita & Apperly (2014) adapted the Bear Dragon task (Kochanska, Murray, Jacques, Koenig, & Vandegest, 1996) in which a pair of children were instructed by puppets to point to a picture (they were required either to point when one puppet told them or to inhibit pointing when instructed to by the other puppet. In Milward and colleagues' (2014) paradigm, in the same condition, both children were required to respond to the same puppet and in the different condition, children were required to respond to different puppets.

<sup>72</sup> Note that in studies testing visual perspective-taking abilities, participants were presented with stimuli of an avatar in a room.

distinction between two levels of visual perspective-taking (level-1 VPT and level-2 VPT; Apperly & Butterfill, 2009; Flavell et al., 1981; Surtees, Apperly & Samson, 2016; Surtees, Butterfill, & Apperly, 2012).

#### 7.2.1.1 Level-1 visual perspective-taking abilities

Level-1 VPT refers to the ability to judge *whether* objects are seen by another (Flavell et al., 1981a); experimentally, a perspective refers here to ‘a restriction on *which* things can be seen’ (Perner, Brandl & Garnham, 2003; see Fig. 4 below).



**Fig. 4:** Illustration of level-1 VPT contexts. The agent on the left side uses level-1 VPT abilities if he understands that the agent in front of him (on the right side) only perceives the blue balloons (not the green ones).

A classical test of Level-1 VPT is the original Dot Perspective Task (Samson et al., 2010; Qureshi, Apperly & Samson, 2010) where participants are presented with stimuli of a human avatar and an array of dots in a room. On each trial, participants are instructed to take the avatar’s perspective (prompted with HE or SHE) or their own perspective (prompted with YOU). Then, a single digit is shown in the middle of the screen, followed immediately by a stimulus containing the avatar and a certain number of dots. Participants were asked to respond “Yes” or “No”, depending on whether the digit matches the number of dots in the stimulus in relation to the perspective they were instructed to take. On “self” trials (prompted with YOU), participants must respond based on the number of dots they see in the stimulus. On “other” trials (prompted with HE or SHE), they must respond based on the number of dots the avatar sees. The analysis of reaction times and error rates shows a consistency effect: participants are faster and more accurate when both perspectives are consistent (the participants

and the avatar can see the same number of dots) than when they are inconsistent. In addition, in inconsistent trials, participants are slower and make more errors both when reporting the number of dots the avatar can see (egocentric interference) and when reporting the number of dots they can see (altercentric interference). Some evidence suggests that level-1 perspective is efficiently processed outside of cognitive control. For instance, Samson and colleagues (2010) found that altercentric interference occurs even if participants were repeatedly instructed to judge their own perspective across an entire block or the entire experiment. In addition, Qureshi, Apperly and Samson (2010) found that participants computed the avatar's irrelevant perspective even when also completing a secondary task that loaded executive function (Qureshi, Apperly, & Samson, 2010). These results suggest that participants could not voluntarily ignore the irrelevant perspective of the avatar and that such computation occurred outside of cognitive control.

When do level-1 VPT abilities appear in childhood? Several studies have tested level-1 VPT abilities with different paradigms and found that such abilities emerge around fourteen months of age when implicit measures are used (e.g., looking time paradigms; Sodian, Thoermer and Metz, 2007; Song and Baillargeon, 2008) and two years of age using explicit measures (e.g., interactive or verbal measures; Moll and Tomasello, 2006). Moll and Tomasello (2006) tested 18- and 24-month-olds in an interactive task where children were supposed to hand to an adult a toy that was visible for the child but hidden from the adult's view. They found that only 24-month-olds seemed to understand that when an adult asked for help in searching for an object, she was asking for the object that only the child could see. In contrast, Sodian and colleagues (2007) tested 12- and 14-month-olds using a looking-time paradigm where infants were familiarized to a person reaching for her toy (she explicitly says "Where is my toy?" and "Here" while always reaching for the same toy). In the testing phase, an occluder (either transparent or opaque) was placed in front of the person such as she could or could not see one of the two toys on stage (but the infants could see both toys). Results revealed that only fourteen-month-olds looked longer at a person reaching for and grasping a new object when the old goal-object was visible than when it was invisible to the person (but visible to the infant).

My hypothesis is that what is at stake in level-1 VPT is an understanding of the referential nature of looking<sup>73</sup> which may be related to the progressive development of the child's cognitive sense of self<sup>74</sup>. Children's self-awareness emerges gradually (for a recent review, see Rochat, 2018). At birth infants already demonstrate the rudiments of an implicit sense of sense, differentiating their embodied self from other entities. By the end of the second month, infants show clear signs that they also have a sense of how their own body is situated in relation to other entities in the environment, displaying, for instance, systematic self-defensive reactions to objects looming toward them (Rochat & Goubet, 1995, 2000). During their first months, they also develop an awareness that their actions have an effect (Rochat & Striano, 1999), a larger motor repertoire and a more precise body schema (Filippetti, Johnson et al., 2013). For instance, 4-month-olds start reaching for objects within their reach, but tend to inhibit reaching when an object is either too large for grasping or slightly out of reach (Rochat, Goubet, & Senders, 1999). A second main level of self-awareness, explicit or objectified self-awareness, appears around 18 to 20 months of age. At that age, infants begin to recognize themselves in the mirror and to mark contrasts between themselves and other people in their verbal production. This second main level of self-awareness allows infants to consider others as persons, to over-imitate (many imitation games are recommended to 18-month-olds as they tend to imitate adults around them a lot; see Agnetta and Rochat, 2004). The emergence of this explicit self-awareness is also linked to social awareness (mirror self-recognition task with norm condition<sup>75</sup>, see Rochat, Broesch, & Jayne, 2012). For instance, children will place a sticker they previously saw on their parent's forehead or cheek on the same spot on their own face (from 16 months of age, Kampis, Grosse Wiesmann, Koop

---

<sup>73</sup> For more details, please refer to Part I, footnotes n°17 and n°18.

<sup>74</sup> Around 18 months of age, typical children express a conceptual or objectified sense of self (i.e. they can recognize and evaluate themselves, e.g., "that's me in the mirror, look how tall I am"; see Rochat & Zahavi, 2011) which emerges in conjunction with a general symbolic competence (i.e. a linguistic ability to refer to things differentiating signifier and signified across domains). In other words, it could be that such double ability (better cognitive sense of self and general symbolic competence) enables young children to understand that the others' looks are "about" something (instead of just interpreting the gaze direction as a cue for prediction).

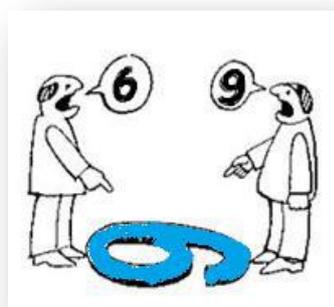
<sup>75</sup> Rochat, Broesch, & Jayne (2012) designed two mirror task conditions. In the classic mirror task condition, the child was marked prior to mirror exposure (only the child was marked). In the norm condition, the child, experimenter, and accompanying parent were marked prior to the child's mirror exposure. Results indicate that in both conditions children pass the test in comparable proportion, with the same increase as a function of age. However, in the norm condition, children displayed significantly more hesitation while removing the mark, often touching it without removing it or, if so, promptly putting the mark back onto their forehead.

& Southgate, 2020). However, it takes time for objective self-awareness to consolidate. It is initially restricted to the present and only slowly becomes extended over time. For example, Povinelli (2001) found that 3 year-olds viewing themselves in a live video do tend to reach for a large sticker they see on top of their own head. In contrast, when the video is not live (they view the replay of the same video taken 3 minutes prior), they do not reach for the large sticker. Povinelli reports the commentary of a 3 year-old viewing herself on a TV with a sticker on her forehead; she says: “it’s Jennifer... it’s a sticker” and then adds: “but why is she wearing my shirt?” (Povinelli, 2001: 81). Furthermore, when asked who was on the TV, it is only by 4 years that the majority of children say “Me” rather. Finally, a third level of self-awareness, which following Rochat (2018) we may call reflective and evaluative self-awareness, starts emerging around 36 months of age with the development of metacognitive abilities, when children start showing awareness of how they are perceived and evaluated by others. Reflective and evaluative self-awareness is linked notably to social emotions such as pride or shame, self-management, inequity aversion (sharing, ownership), the development of deception (e.g., lying behavior), and moral identity.

The second layer of self-awareness which appears around 18-20 months of age – involving a better cognitive sense of self and a general symbolic competence – could sustain the level-1 VPT capacities exhibited by young children.

#### 7.2.1.2 Level-2 perspective-taking abilities

Level-2 VPT refers to the ability to judge *how* the other sees those objects (Flavell et al., 1981b); experimentally, a level-2 perspective refers to ‘a restriction on *how* certain things appear’ (Perner, Brandl & Garnham, 2003; see Fig. 5 below).



**Fig. 5:** Illustration of level-2 VPT contexts. The agent on the left side uses level-2 VPT abilities if he understands that the agent in front of him (on the right side) perceives a nine while himself perceives a six (he does not perceive the stimulus in the same way).

Level-2 VPT abilities can be tested using an adaptation of the Dot Perspective Task including numerals (e.g., 0, 6, 8 or 9) on a wall or on a table with an avatar sitting on the other side of the table. Some numbers appear to the avatar and the participant, either, identically both on a wall and on the table (e.g., 0, 8), or identically on a wall but differently on a table (e.g., 6, 9; Surtees, Butterfill and Apperly, 2012). For instance, when the numeral is placed on a table, the avatar can see a 6 while the participant sees a 9, whereas they both see it as a 9 on a wall. Interestingly, no interference was observed in self-perspective performance when the avatar sees the number differently from the participant (for example a 9 which appears as a 6 for the avatar). In other words, there is no altercentric effect in level-2 perspective-taking tasks. These results contrast with findings from level-1 perspective-taking tasks, suggesting that level-2 perspective-taking requires controlled processing, using executive resources (Surtees et al., 2012, 2016<sup>76</sup>).

If only level-2 VPT abilities require costly executive resources, then level-1 and level-2 VPT abilities should develop at different ages during childhood. When do level-2 VPT abilities appear in childhood? In traditional Level-2 VPT tasks, preschoolers were able to pass the test only by four years of age (Flavell et al., 1981; Masangkay et al., 1974). In contrast, other studies revealed that level-2 VPT abilities could be found by 36 months of age when a different task is used (Moll & Meltzoff, 2011a). In their task, Moll & Meltzoff (2011a) used color filters and children had to recognize how an object looked to an adult when she saw it through a color filter. Furthermore, in their discussion, Moll & Meltzoff point out that their task does not capture perspective-taking at the same level as traditional tasks<sup>77</sup> and propose to draw a distinction between level-2A and a level-2B VPT abilities:

---

<sup>76</sup> “While it was clear that participants processed and correctly judged both level-1 and level-2 perspectives when directly asked to do so, the indirect measure [interference effects: altercentric or egocentric effects] indicated that only level-1 PT was triggered outside of cognitive control” (Surtees, Samson and Apperly, 2016: 103).

<sup>77</sup> “In all these other tasks, children have to simultaneously ‘confront,’ to borrow Perner, Stummer, Sprung, and Doherty’s (2002) term, two different perspectives on one and the same thing. In appearance–reality tasks, two different conceptual perspectives are confronted: The self-same object can be construed as a rock from an ‘appearance perspective’ and as a sponge from a ‘reality perspective.’” Similarly, in the alternative-naming task, children have to acknowledge that a given object, for example, a rabbit, can be labeled both as a ‘‘rabbit’’ and as a ‘‘bunny’’ (Doherty & Perner, 1998). Likewise, in Masangkay et al.’s (1974) turtle task children have to

We thus would argue that what has been called “Level 2 visual perspective-taking” needs to be redefined and differentiated into two distinct levels: (a) the ability to identify or produce specific ways of seeing an object and (b) the understanding that an object may simultaneously look different from different perspectives. Hence, the classic Level 1 versus Level 2 distinction may have to be extended to a third level that allows children to “confront” perspectives and understand that a given object can simultaneously be seen/construed differently from different standpoints.

(Moll & Meltzoff, 2011a: 672)

In others words and following Moll & Meltzoff’s (2011a,b) terminology, three levels of VPT abilities can be distinguished. Level-1 VPT abilities may be related to the objectified self (Rochat, 2018) which emerges around 18 months of age and enables young children to *take others* perspective. Level-2A VPT abilities may be related to reflective self which emerges around 36 months of age (Rochat, 2018) and enables children to *take another’s* perspective that differs from their own view of an object (without “confronting” perspectives). Level-2B VPT abilities may be related to the emergence of theory of mind abilities and executive functions around 4 years of age, enabling children to *confront* perspectives with another (be these actual perspectives held by concrete individuals or possible perspectives that “one” could hold<sup>78</sup>) about a same object<sup>79</sup>.

In a joint action context, co-agents represent each other thanks to their perspective-taking abilities (Table 1) which may lead to the emergence of either facilitation or interference effects

---

acknowledge that the turtle can simultaneously look “upside down” from one visual perspective but “right side up” from another. In other words, what is put to a test in all these tasks is an understanding that there can be two different judgments, construals, or (re)presentations of one and the same object or event at the same time” (Moll & Meltzoff, 2011a: 670-671).

<sup>78</sup> “Such a simultaneous confrontation of perspectives, however, is not necessary in the color filter tasks (Moll and Meltzoff, 2011[a]). To succeed in these tasks, the child needs to recognize how the adult sees an object but not how that compares to their own perception of it. They can ignore the fact that what looks, for example, green to the adult looks blue to them-because they are not asked to contrast or confront the others' perspective with their own at that time” (Moll & Meltzoff, 2011b: 299).

<sup>79</sup> “While 2A is well in place by 3 years of age, 2B emerges at around 4½ years, as has been established and replicated in numerous false belief and similar theory of mind tasks (see e.g., Wellman, Cross, and Watson, 2001)” (Moll & Meltzoff, 2011b).

depending on whether there is a spatial or perspectival congruence or incongruence between co-agents.

Action Strategy	Identical	Complementary	
VPT Level	Level-1	Level-2	Level-3
Self-other perspectives	Taking other perspective	Maintaining self and other perspectives	Confronting self and other perspectives

**Table 1:** Relationship between action strategy adopted by co-agents and the different levels of VPT abilities they require due to co-representation effects (either to facilitate or avoid interference within performance)<sup>80</sup>. Congruent contexts favor the quality of information accessed about the partner and lead to the emergence of performance facilitation due to co-representation effects. In such congruent contexts, co-agents adopt identical strategies which can rely simply on level-1 VPT i.e., on the ability to take the partner’s perspective (without any need of dissociating self from other perspectives). Incongruent contexts favor the poor quality of information accessed about the partner and lead to the emergence of performance interference due to co-representation effects. In such incongruent contexts, co-agents adopt complementary strategies which can rely either on level-2A VPT or level-2B VPT based on the need to maintain or confront self and other perspectives within the joint action.

### 7.2.2 *Forms of interaction and strategies in childhood*

Presumably, in congruent contexts, level-1 VPT abilities (which produce an altercentric effect outside of any cognitive control) may be the cognitive mechanism leading to the emergence of facilitation effects. In contrast, in incongruent contexts, children would have to deploy level-2A and level-2B VPT to successfully coordinate and limit interference effects. If this hypothesis is true, then with regard to children’s self-awareness development, (1) children should be progressively able to manage more varied joint action contexts as they develop new levels of VPT abilities; (2) younger children should be more proficient in congruent contexts compared with incongruent ones and (3) while initially children probably adopt leader-follower strategies to learn and adjust their behaviors tending toward the adoption of identical strategies, later on, they might be able to adopt complementary strategies in joint action contexts.

---

<sup>80</sup> Even though this relationship is represented using a table, it should be considered as a continuum whereby, in terms of effort, the optimal strategy is the identical one and the less optimal strategy is the complementary one.

At an early age, infants interact almost exclusively with their caregiver or parent. In fact, Bakeman and Adamson (1984) found that 18-month-olds' play consists of 25% joint activity with their mothers, whereas only 7% was with a peer. Many studies have shown that human adults do not interact with young children as they would with another adult; for instance, they emphasize their gestures, their facial expressions, accentuate their intonation. With young infants, adults interact in naturalistic games (see Chapter 2, 2.1.2.1) and seem to adopt leader-follower strategies (e.g., social games, see Fantasia et al., 2014). While infants are learning new skills and improving them through practice, they probably adopt identical strategies with their caregiver or parent. The fact that young children barely succeed in coordinating with peers at above-chance at around 24 months (Warneken, Chen & Tomasello, 2006; see also Brownell & Carriger, 1990) suggest that they cannot compensate for their peer's margin of error (e.g., due to lack of motor abilities) while they themselves still need to improve their performance. My hypothesis is that children initially adopt leader-follower strategies, first being only able to follow their caregiver or parent and later on, being able to lead in symbolic plays ( $\approx$  18-24 months of age, e.g., tea parties plays<sup>81</sup>), in more complex games including rules ( $\approx$  25-27 months of age; e.g., if rules are not respected, children intervene, see Hamann et al., 2012). When infants simply imitate their caregiver, I don't think that they perform joint actions; I think they rather develop their motor skills and self-awareness in their interaction with others. My point is that, the imitation of their caregiver does not yet enable infants to integrate self and other representations in a dyadic motor plan. The integration of self and other representations requires both the ability to take the other perspective that infants show early on age and the development of self-awareness which emerges progressively.

The adoption of complementary strategies entails the ability to complement their partner's action and to avoid potential interference effects due to the contextual incongruency (co-representation effects are measured at 4-5 years of age, see Milward et al., 2014). More specifically, in joint go/no-go tasks, the assessment of individual differences revealed that *theory of mind* helps children to clearly

---

<sup>81</sup> Note that self-other understanding (e.g., talking about their own and others' actions and internal states, taking adults' perspective or using personal pronouns, see Brownell et al., 2006) is correlated with a more skillful cooperation with peers and notably during symbolic play (i.e., children use objects to represent or symbolize other objects, e.g., a banana becomes a cell-phone or play tea parties).

differentiate and separate task representations of self and other, *inhibitory control skills* were predictors of the ability to avoid interference from co-representation effects leading to more accuracy in their turn-taking performance and the precision in *action prediction* was a predictor of the action timing variability during the play (Milward, Kita & Apperly, 2016; Meyer et al., 2015). However, comparing the experimental designs of puzzle-solving tasks (see 2.1.2.3) revealed that (1) the task itself (more or less demanding) might make a big difference in the age at which children show the ability to complement their partner's action and that (2) the tasks must be analyzed using three levels of VPT abilities instead of two.

Puzzle-solving tasks reveal the ability to complement the partner's action – testing proactive planning without previous experience – from 2 ½ years of age (Warneken et al., 2014; Meyer et al., 2016; Paulus, 2016) but the flexibility of the adjustment to the partner starts to emerge only from 3 ½ years of age (note that 5-year-olds only were flexible in their adjustment in less than 60% of the trials; Meyer et al., 2016). For instance, in Meyer and colleagues (2016), children were instructed to pass cups to their partner who had only one hand available (alternating over time) to build a tower. Proactive planning was assessed at the very start of the joint-action game when children decide which side to pass the first cup. In fact, 70% of all children passed the cup to the joint-action partner on her free-hand side. In contrast, the flexibility of the adjustment was assessed across the cup-stacking task while the experimenter alternated her hand available (children were flexibly adjusting when they handed the cup to the experimenter's free hand). In Moll & Meltzoff's terminology, in the proactive planning assessment, researchers measured level-1 VPT abilities since children could take the other perspective without the need to maintain their own whereas in the flexibility adjustment assessment, researchers measured level-2A VPT abilities since it required children to maintain self and other perspective across the experiment to succeed. In daily life, when children perform helping behaviors, when they point to an object their parent is looking for, they rely on level-1 VPT abilities but when they need to physically get involved as in opening a cabinet for an adult whose arms are full, they rely on level-2A VPT abilities. A contrasting example could be the experimental design tested in Paulus (2016) whereby 3-, 5-, and 7-year-old children and adults were instructed to hand the experimenter a tool with which she could activate 1 of 2 different effects on the apparatus (based on the orientation of

the tool during its insertion). In order to elicit a specific effect, participants should plan their grasping and reaching actions in such a way as to enable the partner to efficiently handle the tool, i.e., anticipating the final end state of the joint activity. Adults demonstrate joint action planning from the beginning, 5- and 7-year-olds improve their joint action performance over time but 3-year-olds children do not adjust their behavior to accommodate the other's action (they also do not improve over time and first-hand experience with the task before acting with a partner does not facilitate performance in the joint action task). This task requires that the agents "confront" self and other perspectives in order to appropriately complement the partner's actions and reach the desired effect; in other words, it requires level-2B VPT abilities.

In the literature, the absence of co-representation effects before 4 years of age has been interpreted as young children's inability to represent and integrate their partner in their joint plans without fully developed theory of mind abilities (Milward et al., 2014). Such an interpretation is in line with the classical idea that young children are ego-centered before being able to be alter-centered (e.g., the three mountains tasks in Piaget and Inhelder, 1956). In contrast, I propose, following Southgate's altercentric hypothesis that, the absence of co-representation effects may be due to the fact that children are alter-centered before becoming able to maintain a self-perspective with the emergence and consolidation of objective self-awareness. As we saw, the development of perspective-taking abilities – level-1, -2A and -2B – provides support for the idea that children are alter-centered before being ego-centered (Southgate, 2020). Their altercentric bias, discussed in Chapter 5, enables young children to see through the eyes of others and perform some minimal joint actions from an early age. Indeed, in joint action performance, it seems that young children use visual perspective-taking and supplement motor simulation with teleological reasoning, as a basis for predicting, and reasoning about, the other behavior.

The variety of joint actions leads to contexts that are more or less congruent between co-agents. The congruency of the context modulates the quality (completeness and reliability) of the information available about the partner. Consequently, depending on the congruency of the context, different strategies must be adopted to either emphasize or reduce the impact of co-representation effects within performance. The more congruent a joint action context is, the more co-agents will adopt identical

strategies which emphasize facilitation effects due the spontaneous ability to take their partner's perspective (level-1 VPT abilities). In contrast, the more incongruent a joint action context is, the more co-agents will adopt complementary strategies which reduce interference effects thanks to the dissociation between self and other perspectives (maintenance and confrontation, respectively with level-2A and -2B VPT abilities).

Analyzing children's perspective-taking abilities revealed that they are connected with the development of self-awareness. Three levels of VPT abilities have been described. I examined which of these VPT abilities were involved in the strategies (identical or complementary) deployed by young co-agents in joint action. In the light of this literature, I conclude that young children are alter-centered (Southgate, 2020), i.e., they have a tendency to encode events from the visual perspective of others and they might not have a competing self-perspective (failure to maintain a representation of the event from their own perspective). In human ontogeny, altercentrism may serve as a way to learn about others and about the world through the lenses of others (e.g., inferring behavioral rules, Kamps et al., 2013). In the joint action context, altercentrism enables children to perform minimal joint actions adopting identical strategies without the need of dissociating self from other perspective and later on, while developing self-awareness and notably level-2A perspective-taking abilities, perform minimal joint actions adopting complementary strategies only with the need to maintaining self and other perspectives.

Future research could focus on the mechanisms involved in modulating the relative strength of self and other perspectives, as a consequence of the altercentric bias is that representations from the perspective of others might initially outweigh representations from one's own perspective.

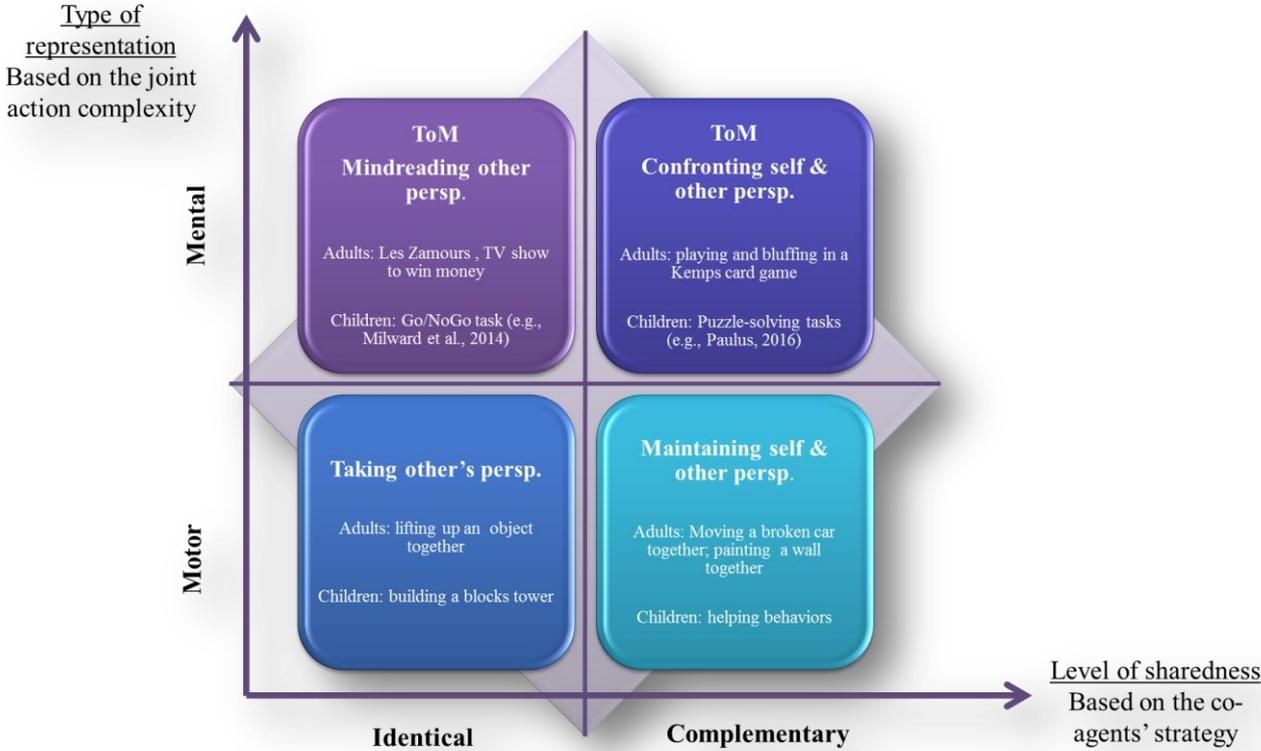
\*\*\*

Understanding how co-agents encode and integrate each other via the use of shared representations is probably one of the most challenging issues in an investigation of the social-cognitive underpinnings of joint action. In Part III, I proposed a joint action matrix with two axes, namely the type of representations involved in joint action and the level of sharedness implied by such representations.

In Chapter 6, I presented the first axis of the matrix and proposed that the type of shared representation co-agents need to form (shared goal representations vs. shared intentions) depends on the complexity of the action's goal (itself characterized in terms of abstractness and demands for advance planning). I discussed two main positions (the Simple View and the Motor Theory of Goal Tracking) regarding the *processes* – general inferential processes about means-goals associations or motor simulation – involved in goal tracking and action prediction. I defended an intermediary position considering that while motor simulation is the main resource that young children use in order to ascribe goals to others and predict their actions, the activation of a motor representation of an action in the observer also depends on the probability of that action in the given context. Concretely, if the joint action's complexity modulates the type of shared representations involved, it gives us a categorization of joint actions such that joint actions of low complexity rely on shared motor representations when highly complex joint actions require shared mental representations (e.g., shared intentions). Joint actions of intermediate complexity mainly require the use of motor representations but they also integrate other skills (e.g., planning skills, norms integration, common knowledge) that emerge later in childhood and are presumably also used by adults.

In Chapter 7, I presented the second axis of the matrix proposing that the way in which representations are shared and their level of sharedness depend on the strategies co-agents need to use to maximize their performance. The variety of joint actions leads to more or less contextual congruency between co-agents. Co-agents' congruency modulates the quality (completeness and reliability) of the information available about the partner and thus, requires the adoption of different strategies (identical or complementary) to either emphasize or reduce the impact of co-representation effects during performance. The more congruent a joint action context is, the more co-agents will adopt identical strategies which emphasize facilitation effects. In contrast, the more incongruent a joint action context is, the more co-agents will adopt complementary strategies which reduce interference effects thanks to the dissociation between self and other perspectives. In Chapter 7, I presented three levels of VPT abilities and examined the extent to which they were involved in the strategies that co-agents deploy in joint action. I argued that children's perspective-taking abilities are connected with the development of self-awareness – young children being fundamentally alter-centered – and that

their ability to participate in joint actions reflects their management of self- and other-related representations. In congruent joint action contexts, children are able to perform minimal joint actions adopting identical strategies without the need to dissociate self from other perspective since they have a tendency to encode events from the visual perspective of others and they might not have a competing self-perspective to manage. Later on, as self-awareness and level-2A perspective-taking abilities develop, they can perform minimal joint actions adopting complementary strategies which require them maintaining self and other perspectives but not to confront them. Finally, around the same age at which more robust theory of mind abilities emerge, children develop level-2B perspective-taking abilities which enable them to perform joint actions adopting complementary strategies which require confronting self and other perspectives. In other words, different joint action and contexts may require co-agents to adopt different strategies to maximize their performance and the level of sharedness of shared representations depends in turn on the strategy used. The joint action matrix with its two axes is presented in Fig. 6.



**Fig. 6:** A joint action matrix. This matrix presents four cases of joint actions based on two main axes, namely the type of representation involved in joint action based on the joint action complexity and the level of sharedness implied by such representations based on the co-agents' strategy to maximize their

joint action performance. Some joint actions require encoding others in goal terms, which can be done via motor representations; depending on what is the best strategy for joint action performance, co-agents might only need to take the other perspective (using level-1 VPT abilities) or they might have to maintain self and other perspective (using level-2A VPT abilities). Some joint actions require encoding others in mental terms, using mental representations (thus, theory of mind abilities); depending on what is the best strategy for joint action performance, co-agents might only have to mindread the other perspective<sup>82</sup> or they might have to confront self and other perspectives (using level-2B VPT abilities).

---

<sup>82</sup> All the examples given in this matrix were discussed in the thesis except the example of Les Zamours. Les Zamours is a French game show adapted from the American game The Newlywed Game. Each show confronts three couples for about thirty minutes; the facilitator asks questions about the candidates' married life (e.g., what is your spouse favorite color? What part of your body does your spouse like the most?). Women respond first to questions previously asked to their spouses, each corresponding response earns a certain number of points. The roles are then reversed; it is the men who must give the same answer as their spouses to gain points. The couple with the highest score can participate in the final. The final takes the form of a questionnaire on the tastes of the spouse. The first contestant has forty-five seconds to answer seven questions correctly, knowing that the game is lost if the time is up or if more than four mistakes have been made. It appears clearly that the game requires that the spouses mindread each other perspective in order to answer correctly.



## *Conclusion*

Maximalists defend variants of the *shared intention hypothesis* i.e., the hypothesis according to which joint actions require shared intentions. For agents to form such shared intentions, they need to engage high-level cognitive abilities in order to encode and incorporate the co-agent(s) in mental terms.

In contrast, minimalists argue that some kinds of joint action do not require fully developed mentalizing and meta-representational abilities (i.e., a robust theory of mind); they propose that some joint actions might require only goal representations (instead of intentions). Observations from developmental psychology reveal that young children use certain abilities such as turn-taking, action prediction, joint attention, coordination skills or normative factors to perform actions which look like joint action before the age of four i.e., before they have a robust theory of mind. In other words, minimal cognitive abilities as proxies for, or precursors of, fully developed cognitive abilities (which could be found in children's early social competencies) and contextual features such as intentions-in-action reading, the use "coordination smoothers" or the ability to understand the motivational state of others might be sufficient to perform such joint actions.

The analysis of whether observations in childhood are compatible with maximalist accounts of joint action, (i.e., the confrontation of joint actions as they are defined by maximalists – in terms of high-level cognitive skills – with developmental psychology studies) revealed that (1) developmental data do not support maximalist accounts of joint actions, (2) the teleological reasoning and the two mindreading systems accounts offer two good candidate accounts of the kind of cognitive abilities that support joint action in childhood, (3) the altercentric bias hypothesis would allow *motorically shared* cooperative activities in young children.

From there, I defined a joint action matrix with two axes, corresponding respectively to the type of representations involved in joint action and to their level of sharedness.

With respect to the first axis of the matrix, I proposed that the type of shared representation co-agents need to form depends on the complexity of the action's goal (itself characterized in terms of abstractness and demands for advance planning). I distinguished three broad types of joint actions

along that axis: low complexity (low abstractedness and low advance planning), intermediary complexity (high abstractedness and low advance planning or the reverse) and high complexity (high abstractedness and advance planning) joint actions. I proposed that joint actions of low complexity can rely on shared motor representations but that highly complex joint actions require shared mental representations (e.g., shared intentions). Joint actions of intermediate complexity mainly require the use of motor representations but also integrate other skills (e.g., planning skills, norms integration, common knowledge) that emerge later in childhood and are presumably also used by adults.

In the second axis of the matrix, the level of sharedness of representation depends on the strategies co-agents must resort to in order to maximize their joint performance. Co-agents adopt different strategies to either increase co-representation effects to maximize performance in congruent contexts (identical strategies) or to reduce them to maximize performance in incongruent contexts (complementary strategies). The adoption of identical and complementary strategies requires different forms of management of self and other perspectives. I argued that children's perspective-taking abilities are connected with the development of self-awareness – young children are initially alter-centered – and that their ability to participate in joint actions reflects their management of self- and other-related representations. Children are first able to perform minimal joint actions adopting identical strategies without the need of dissociating self from other perspective since they have a tendency to encode events from the visual perspective of others and they might not have a competing self-perspective to manage. Later on, as self-awareness and level-2A perspective-taking abilities develop, they can perform minimal joint actions adopting complementary strategies which require them maintaining self and other perspectives but not to confront them. Finally, around the same age at which more robust theory of mind abilities emerge, children develop level-2B perspective-taking abilities which enable them to perform joint actions adopting complementary strategies which require confronting self and other perspectives.

In conclusion, rather than simply asking whether young children are capable of joint action or not, one should take into account the fact that joint action vary in complexity and can impose more or less stringent representational and sharing demands. As the proposed matrix (Fig. 6) illustrates, while success at more complex forms of joint action depends on the progression of mindreading and

perspective-taking abilities, children may successfully engage in minimal forms of joint action before these abilities are fully developed.

### **Future directions**

Future research can look deeper into many specific points that were discussed in this thesis. First, researchers can test to what extent motor simulation can be independent from other goal-tracking processes presented in Chapter 6. Second, researchers can focus on the emergence of different strategies (identical or complementary forms of interaction) in children when participating in a live motor cooperative activity and test the role played by certain cognitive abilities (e.g., inhibitory control skills, theory of mind capacities, self-awareness abilities) in the adoption of one strategy over another. Third, researchers can investigate the levels of perspective-taking abilities which children develop to further test the validity of the distinction between three levels of perspective-taking abilities. Fourth, researchers can examine whether shared representations have a subjective component based on the different levels of perspective-taking abilities involved. This question can be addressed in regards with the idea that representations are tied to self and other (e.g., Kovács, 2016 using “belief-files” or Apperly & Butterfill, 2009 using “registration”) or that there can be specific self and other representational *strength* and *format* (Southgate, 2019).



## *Appendix 1*

### *The joint Simon effect*

This paradigm referred to as the ‘social’ or joint Simon task has been extensively used to study joint action (e.g. Dittrich, Rothe, & Klauer, 2012; Iani, Anelli, Nicoletti, Arcuri, & Rubichi, 2011; Klemnova & Liepelt, 2016; Kuhbandner, Pekrun, & Maier, 2010; Liepelt, Wenke, & Fischer, 2013; Liepelt, Wenke, Fischer, & Prinz, 2011; Stenzel et al., 2012, 2014; Tsai & Brass, 2007; Vlainic, Liepelt, Colzato, Prinz, & Hommel, 2010; Welsh, 2009; Welsh et al., 2013) and is considered as an evidence of co-representation of the partner in a joint configuration.

Researchers identified several parameters, specific to the joint action context, which are necessary for the emergence of the joint Simon effect. Many factors are involved in the identification of the other individual as a co-agent by the agent herself such as the co-agent’s active involvement in the task (Sebanz et al., 2007), the attribution of agency to the co-agent (Stenzel, et al., 2012, 2014; Wen & Hsieh, 2015), the perceived interdependence between co-agents (e.g. Colzato, de Bruijn, & Hommel, 2012; Iani et al., 2011; Ruys & Aarts, 2010; Ruissen and de Bruijn, 2016) in a response-uncertainty context (Karlinsky, Lam, Chua and Hodges, 2017).

#### *Response uncertainty and the need for agent discrimination*

Karlinsky, Lam, Chua and Hodges (2017) found that the joint Simon effect only emerges when there is a response uncertainty (i.e. if the task consists in a turn-taking task with 100% predictability of turns, no effect is observed). In this study, Karlinsky and colleagues (2017) designed two conditions: participants predictably alternated turns on consecutive trials, their stimulus presented either on 100% of their turns (fully predictable group) or on 83% of their turns (response-uncertainty group, 17% no-go catch trials). The joint Simon effect was absent in the fully predictable condition which suggests that the need to execute a response and by extension the degree of preparation seems crucial to co-representation.

When the context is fully predictable, it seems that agents do perform parallel tasks and not a joint task, they just need to get ready alternatively and thus individually, there is no challenge to overcome. In other words, they do not need to consider the other as a partner (no joint intention) and the prediction is not directed toward the agent next to them but the context which is already fully predictable.

### *Identification of co-agent*

In joint action, contrary to the case of parallel actions, the other agent needs to be identified as a co-agent by the agent herself. Studies have shown that the joint Simon effect only emerges when the other agent is considered as displaying intentionality and agency (Stenzel, et al., 2012, 2014; Wen & Hsieh, 2015), actively involved in the task (Sebanz et al., 2007) and in a context of cooperation rather than competition (e.g. Colzato, de Bruijn, & Hommel, 2012; Iani et al., 2011; Ruys & Aarts, 2010; Ruissen and de Bruijn, 2016).

A number of studies suggest that the joint Simon effect only emerges when participants are partnered with a human co-agent but not with a computer (for a recent review, see Sahāi et al., 2019). In fact, a growing body of research in human-machine interaction reveals that the emergence of the joint Simon effect requires two components, namely the attribution of intentionality and the attribution of agency to the co-agent (Stenzel et al., 2014). In their study, Stenzel and colleagues (2014) have three conditions which vary the agency<sup>83</sup> and the intentionality displayed in the co-agent's actions. The co-agent either intentionally presses down on the response key (agency+/intentionality+ condition) or passively places his/her finger on the response key which automatically moved up and down (agency-/intentionality- condition) or seemed to control the response key with a brain computer interface while placing their finger besides the response key (agency- /intentionality+ condition). The joint Simon effect only emerges in the agency+/intentionality+ condition, meaning that the attribution of agency is necessary for the emergence of the effect. The attribution of agency to the co-agent seems to be crucial to identify the other agent as a co-agent (who can participate and make a difference).

---

<sup>83</sup> Agency refers to the causal relationship between agent and action effect.

Sebanz and colleagues (2007) showed that it is not sufficient for the co-agent to be present without acting to observe the joint Simon effect, meaning that the social context is crucial. Indeed, the joint Simon effect only emerges when both partners are actively involved and acting in the task revealing that individuals' performance at motor planning is affected by the action of his/her task partner. The fact that I need my co-agent to be actively involved in the task could be interpreted as a requirement to form a shared intention (if I do not need my partner, then it is not a joint action) and it could also both be a significant sign that both my co-agent is committed to the joint task, we form a (mutual) belief about each other involvement and "mutual responsiveness" in adjusting to each other to perform the joint task. Besides, the reachability of the co-agent's response affects the emergence of the joint Simon effect. In a recent study, Iani and colleagues (2019) found that the effect emerges only when participants could reach their co-agent's response with the use of a tool. The reachability of the co-agent's response could be interpreted such as a form of commitment to mutual support, i.e. if my co-agent's response is reachable, I could help him/her in the joint task.

#### *Cooperation rather than competition*

Previous studies have shown that the joint Simon effect only emerges in a context of cooperation rather than competition (e.g. Colzato, de Bruijn, & Hommel, 2012; Iani et al., 2011; Ruys & Aarts, 2010; Ruissen and de Bruijn, 2016).

In fact, the perceived interdependence of co-agents affects the emergence of the joint Simon effect; being in a competitive context with the other agent does not allow the emergence of interference effects. For instance, participants show the joint Simon effect only when they perform the task in a positive interdependence relationship with their co-agent i.e. when their co-agent is friendly and cooperative (not when their co-agent is intimidating and competitive; Hommel et al., 2009). Other studies revealed that negative interdependence such as competitive contexts (e.g. the fastest and more accurate agent in the pair receive a monetary reward) inhibit the emergence of the joint Simon effect (Iani et al., 2011, 2014) and even when participants were confronted to a competitive context before being involved in the joint Simon task (e.g. in a Tetris task on a computer beforehand; Ruissen and de

Bruijn, 2016)<sup>84</sup>. The interdependence between agents shows how important it is to identify the other agent as a cooperative partner to “cement” the agents together as co-agents (which could refer as the interdependence condition in Bratman’s account, the collective commitment in Tuomela’s account or the common knowledge about the conditional commitment of co-agents to participate in the joint action in Gilbert’s account; see Chapter 1).

---

<sup>84</sup> Note that the fact that participants were placed in a competitive or cooperative context beforehand and separately from the joint Simon task (playing Tetris game) rules out the possibility of *fully* explaining the modulating results in terms of attentional processes since attending to the other’s task share is equally required in both cooperative and competitive game play (Ruissen and de Bruijn, 2016).



## *Appendix 2*

### *Kemps card game*

Kemps card game usually involves a deck of 52 cards and **four players** divided into **teams** composed of two players disposed such that the partners sit across from each other with a playing surface in the middle.

The game is organized with several hands and in each hand, the team who loses gets a letter of the word KEMPS; the game ends when one team can spell the whole word K-E-M-P-S meaning that they lose the game.

To win a hand, there are two possibilities: (1) you obtain four cards of a kind (for example, four queens) and your partner says “KEMPS” or vice versa; (2) you or your partner say “COUNTER-KEMPS” when a player from the other team had four of a kind, before his partner says “KEMPS”. Note that if a player says “KEMPS” but his partner does not have four cards of a kind or if a player says “COUNTER-KEMPS” but the other team does not have four cards of a kind, they lose the hand. In fact, as soon as the words “Kemps!” or “COUNTER-KEMPS” are said, the hand is over.

If a player says “KEMPS” his partner must reveal his cards. If the partner does not have four of a kind, then the team that incorrectly called “KEMPS” receives a letter. If a player says “COUNTER-KEMPS” it means that he suspects the opposite team of having “KEMPS”. If one of the opposing players does in fact have four of a kind, then the team who was “countered” receives a letter. However, if “COUNTER-KEMPS” is said and the opposing team does not have “KEMPS” then the team who said it receives the letter.

Before the game starts, the teams are formed and each two player team goes away and privately agrees a signal (e.g., body movement, gesture, hand signal) that they can secretly use to tell each other when they have four of a kind. No words or table talks are allowed. Note that partners can confuse the other team using fake signals.

When the game starts, each player is dealt with four cards and four cards are turned face up on the playing surface. All players swap one of their cards for one of the central cards at any time. If a player picks up more than one card, he must discard an equal number of different cards to have four cards in his hand. Note also that if two players want the same face up card, the first player to touch it gets it. When there is no further swaps desired by players, they clear the four cards and turn up four new central cards. If at some point, no one wants any of the face up cards but there are not more fresh cards to deal, the hand ends but no one gets a letter.

The game ends when one team can spell the whole word K-E-M-P-S meaning that they lose the game.



## *Bibliography*

- Adam, M., & Elsner, B. (2018). Action effects foster 11-month-olds' prediction of action goals for a non-human agent. *Infant Behavior and Development*, 53, 49-55.
- Adam, M., Reitenbach, I., Papenmeier, F., Gredebäck, G., Elsner, C., & Elsner, B. (2016). Goal saliency boosts infants' action prediction for human manual actions, but not for mechanical claws. *Infant Behavior and Development*, 44, 29-37.
- Adams, B. J., Hooper-Bùi, L. M., & Strecker, R. M. (2011). Raft formation by the red imported fire ant, *Solenopsis invicta*. *Journal of Insect Science*, 1
- Aggelopoulos, N. C. (2015). Perceptual inference. *Neuroscience & Biobehavioral Reviews*, 55, 375-392.
- Agnetta, B., & Rochat, P. (2004). Imitative games by 9-, 14-, and 18-month-old infants. *Infancy*, 6(1), 1-36.
- Aknin, L. B., Van de Vondervoort, J. W., & Hamlin, J. K. (2018). Positive feelings reward and promote prosocial behavior. *Current Opinion in Psychology*, 20, 55-59.
- Allen, M., Perry, C., & Kaufman, J. (2018). Toddlers prefer to help familiar people. *Journal of Experimental Child Psychology*, 174, 90-102.
- Alonso, F. M. (2009). Shared intention, reliance, and interpersonal obligations. *Ethics*, 119(3), 444-475.
- Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116(4), 953-1023.
- Apperly, I.A. (2013). Can theory of mind grow up? Mindreading in adults, and its implications for the neuroscience and development of mindreading, in Baron-Cohen, S., Tager-Flusberg, H. & Lombardo, M. (Eds.) *Understanding Other Minds Perspectives from developmental social neuroscience*, Oxford University Press, (3rd Edn.), 520 p: 72-92.
- Avenanti, A., Annella, L., Candidi, M., Urgesi, C., & Aglioti, S. M. (2013). Compensatory plasticity

- in the action observation network: virtual lesions of STS enhance anticipatory simulation of seen actions. *Cerebral Cortex*, 23(3), 570-580.
- Avril, A., Purcell, J., & Chapuisat, M. (2016). Ant workers exhibit specialization and memory during raft formation. *The Science of Nature*, 103(5-6), 36.
- Bacharach, M. (2006). *Beyond individual choice: teams and frames in game theory*. Princeton University Press.
- Baess, P., Weber, T., & Bermeitinger, C. (2018). Sharing different reference frames: How stimulus setup and task setup shape egocentric and allocentric Simon Effects. *Frontiers in Psychology*, 9, 2063.
- Baillargeon, R., Buttelmann, D., & Southgate, V. (2018). Invited commentary: Interpreting failed replications of early false-belief findings: Methodological and theoretical considerations. *Cognitive Development*, 46, 112-124.
- Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology*, 67, 159-186.
- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, 14(3), 110-118.
- Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child development*, 1278-1289.
- Bakker, M., Sommerville, J. A., & Gredebäck, G. (2016). Enhanced neural processing of goal-directed actions after active training in 4-month-old infants. *Journal of Cognitive Neuroscience*, 28(3), 472-482.
- Barresi, J., & Moore, C. (1996). Intentional relations and social understanding. *Behavioral and Brain Sciences*, 19(1), 107-122.
- Barrett, T. M., Traupman, E., & Needham, A. (2008). Infants' visual anticipation of object structure in grasp planning. *Infant Behavior and Development*, 31(1), 1-9.
- Becchio, C., Cavallo, A., Begliomini, C., Sartori, L., Feltrin, G., & Castiello, U. (2012). Social grasping: From mirroring to mentalizing. *Neuroimage*, 61(1), 240-248.

- Becchio, C., Manera, V., Sartori, L., Cavallo, A., & Castiello, U. (2012). Grasping intentions: from thought experiments to empirical evidence. *Frontiers in Human Neuroscience*, 6, 117.
- Becchio, C., Sartori, L., & Castiello, U. (2010). Toward you: The social side of actions. *Current Directions in Psychological Science*, 19(3), 183-188.
- Belden, J., Mansoor, M. M., Hellum, A., Rahman, S. R., Meyer, A., Pease, C., Pacheco, J., Koziol, S., & Truscott, T. T. (2019). How vision governs the collective behaviour of dense cycling pelotons. *Journal of the Royal Society Interface*, 16(156), 20190197.
- Bennett, M., Schatz, M. F., Rockwood, H., & Wiesenfeld, K. (2002). Huygens's clocks. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 458(2019), 563-579.
- Bernadyn, T., & Feigenson, K. A. (2018). Data gathering ability contributes to visual organization and probabilistic reasoning. *Heliyon*, 4(3), e00582.
- Biro, S., Verschoor, S., Coalter, E., & Leslie, A. M. (2014). Outcome producing potential influences twelve-month-olds' interpretation of a novel action as goal-directed. *Infant Behavior and Development*, 37(4), 729-738.
- Boccadoro, S., Cracco, E., Hudson, A. R., Bardi, L., Nijhof, A. D., Wiersema, J. R., ... & Mueller, S. C. (2019). Defining the neural correlates of spontaneous theory of mind (ToM): An fMRI multi-study investigation. *NeuroImage*, 203, 116193.
- Böckler, A., Knoblich, G., & Sebanz, N. (2012). Effects of a coactor's focus of attention on task performance. *Journal of Experimental Psychology: Human Perception and Performance*, 38(6), 1404.
- Böffel, C., & Müsseler, J. (2020). Taking time to take perspective? Rapidly changing reference frames in the avatar-Simon task. *Acta Psychologica*, 204, 103005.
- Bolt, N. K., Poncelet, E. M., Schultz, B. G., & Loehr, J. D. (2016). Mutual coordination strengthens the sense of joint agency in cooperative joint action. *Consciousness and Cognition*, 46, 173-187.
- Bonatti, L. (2008). At the origin of rationality: how intuitions of probabilities shape predictions about the future. In *16th International Conference on Infant Studies*, Vancouver, Canada.

- Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica*, 106(1-2), 3-22.
- Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: inferential processes versus action simulation. *Current Biology*, 17(24), 2117-2121.
- Bratman, M. (1987). *Intention, plans, and practical reason* (Vol. 10). Cambridge, MA: Harvard University Press.
- Bratman, M. (1992). "Shared Cooperative Activity", *The Philosophical Review*, 101: 327–341.
- Bratman, M. (1993). "Shared Intention", *Ethics*, 104: 97–113.
- Bratman, M. (2009a). "Shared Agency", in *Philosophy of the social sciences: philosophical theory and scientific practice*, C. Mantzavinos, ed., Cambridge, UK; New York: Cambridge University Press, 41–59.
- Bratman, M. (2009b). "Intention, Practical Rationality, and Self-Governance", *Ethics*, 119: 411–443.
- Bratman, M. E. (2007). *Structures of agency: Essays*. Oxford University Press.
- Bratman, M. E. (2013). *Shared agency: A planning theory of acting together*. Oxford University Press.
- Bratman, M. E. (2018). *Planning, time, and self-governance: essays in practical rationality*. Oxford University Press.
- Bratman, M., 2006, "Dynamics of Sociality," *Midwest Studies in Philosophy*, XXX: 1–15.
- Bratman, Michael, 1999, *Faces of Intention: Selected Essays on Intention and Agency*, Cambridge: Cambridge University Press.
- Braukmann, R., Bekkering, H., Hidding, M., Poljac, E., Buitelaar, J. K., & Hunnius, S. (2017). Predictability of action sub-steps modulates motor system activation during the observation of goal-directed actions. *Neuropsychologia*, 103, 44-53.
- Brownell, C. A. (2011). Early developments in joint action. *Review of Philosophy and Psychology*, 2(2), 193-211.
- Brownell, C. A., & Carriger, M. S. (1990). Changes in cooperation and self-other differentiation during the second year. *Child Development*, 61(4), 1164-1174.
- Brownell, C., & Carriger, M. (1991). Collaborations among toddler peers: Individual contributions to

- social contexts. In L. Resnick, J. Levine, & S. Teasley (Eds.). *Perspectives on socially shared cognition*. Washington, D.C.: APA.
- Brownell, C. A., Ramani, G. B., & Zerwas, S. (2006). Becoming a social partner with peers: Cooperation and social understanding in one- and two-year-olds. *Child Development, 77*(4), 803-821.
- Brownell, C. A., Svetlova, M., & Nichols, S. (2009). To share or not to share: When do toddlers respond to another's needs?. *Infancy, 14*(1), 117-130.
- Brownell, C., Nichols, S., & Svetlova, M. (2013). Converging developments in prosocial behavior and self-other understanding in the second year of life: The second social-cognitive revolution. In M. Banaji & S. Gelman (Eds.), *Navigating the social world: What infants, children, and other species can teach us*. New-York: Oxford University Press.
- Buetti, S., Lleras, A., & Moore, C. M. (2014). The flanker effect does not reflect the processing of “task-irrelevant” stimuli: Evidence from inattention blindness. *Psychonomic Bulletin & Review, 21*(5), 1231-1237.
- Bukowski, H., Hietanen, J. K., & Samson, D. (2015). From gaze cueing to perspective taking: Revisiting the claim that we automatically compute where or what other people are looking at. *Visual Cognition, 23*(8), 1020-1042.
- Burge, T. (2018). Do infants and nonhuman animals attribute mental states?. *Psychological Review, 125*(3), 409.
- Burns, P., Riggs, K. J., & Beck, S. R. (2012). Executive control and the experience of regret. *Journal of Experimental Child Psychology, 111*(3), 501-515.
- Burnside, K., Wright, K., & Poulin-Dubois, D. (2018). Social orienting predicts implicit false belief understanding in preschoolers. *Journal of Experimental Child Psychology, 175*, 67-79.
- Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition, 112*(2), 337-342.
- Buttelmann, F., & Kovács, Á. M. (2019). 14-Month-olds anticipate others' actions based on their belief about an object's identity. *Infancy, 24*(5), 738-751.

- Butterfill, S. A. (2010). Joint action without shared intention. Manuscript.
- Butterfill, S. A. (2012). Joint action and development. *The Philosophical Quarterly*, 62(246), 23-47.
- Butterfill, S. A. (2017). Joint action: a minimalist approach. In J. Kiverstein (ed.), *Routledge Handbook on the Social Mind* (pp. 357-369). Routledge.
- Butterfill, S. A. (2019). Goals and targets: a developmental puzzle about sensitivity to others' actions. *Synthese*, 1-22.
- Butterfill, S. A. & Pacherie, E. (2020). Towards a blueprint for a social animal. In A. Fiebich (ed.), *Minimal cooperation and shared agency* (pp. 111-125). Dordrecht: Springer.
- Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language*, 28(5), 606-637.
- Butterfill, S. A., & Sebanz, N. (2011). Joint action: What is shared?. *Review of Philosophy and Psychology*, 2(2), 137-146.
- Cacchione, T., Schaub, S., & Rakoczy, H. (2013). Fourteen-month-old infants infer the continuous identity of objects on the basis of nonvisible causal properties. *Developmental Psychology*, 49(7), 1325.
- Caetano, G., Jousmäki, V., & Hari, R. (2007). Actor's and observer's primary motor cortices stabilize similarly after seen or heard motor actions. *Proceedings of the National Academy of Sciences*, 104(21), 9058-9062.
- Campos, J. J., & Stenberg, C. R. (1981). Perception, appraisal and emotion: The onset of social referencing. In M. E. Lamb & L. R. Sherrod (Eds.), *Infant social cognition* (pp. 273-314). Hillsdale, NJ: Psychology Press.
- Candidi, M., Curioni, A., Donnarumma, F., Sacheli, L. M., & Pezzulo, G. (2015). Interactional leader–follower sensorimotor communication strategies during repetitive joint actions. *Journal of the Royal Society Interface*, 12(110), 20150644.
- Cannon, E. N., & Woodward, A. L. (2012). Infants generate goal-based action predictions. *Developmental Science*, 15(2), 292-298.
- Carey, S. (2009). *The origin of concepts*. Oxford: Oxford University Press.

- Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, 72(4), 1032-1053.
- Carpenter, M. (2009). Just how joint is joint action in infancy?. *Topics in Cognitive Science*, 1(2), 380-392.
- Carpenter, M., & Liebal, K. (2011). Joint attention, communication, and knowing together in infancy. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (p. 159–181). MIT Press.
- Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child development*, i-174.
- Carruthers, P. (2013). Mindreading in infancy. *Mind & Language*, 28(2), 141-172.
- Carruthers, P. (2016). Two systems for mindreading?. *Review of Philosophy and Psychology*, 7(1), 141-162.
- Carruthers, P. (2018). Implicit versus explicit attitudes: Differing manifestations of the same representational structures?. *Review of Philosophy and Psychology*, 9(1), 51-72.
- Cassidy, K. W., Fineberg, D. S., Brown, K., & Perkins, A. (2005). Theory of mind may be contagious, but you don't catch it from your twin. *Child Development*, 76(1), 97-106.
- Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., & Becchio, C. (2016). Decoding intentions from movement kinematics. *Scientific Reports*, 6(1), 1-8.
- Chackochan, V. T., & Sanguineti, V. (2019). Incomplete information about the partner affects the development of collaborative strategies in joint action. *PLoS Computational Biology*, 15(12), e1006385.
- Clark, A. (2015). Radical predictive processing. *The Southern Journal of Philosophy*, 53, 3-27.
- Clarke, S., McEllin, L., Francová, A., Székely, M., Butterfill, S. A., & Michael, J. (2019). Joint action goals reduce visuomotor interference effects from a partner's incongruent actions. *Scientific Reports*, 9(1), 1-9.
- Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, 9(4),

377-395.

- Colzato, L. S., De Bruijn, E., & Hommel, B. (2012). Up to “me” or up to “us”? The impact of self-construal priming on cognitive self-other integration. *Frontiers in Psychology*, 3, 341.
- Costantini, M., Committeri, G., & Sinigaglia, C. (2011). Ready both to your and to my hands: mapping the action space of others. *PloS One*, 6(4), e17923.
- Cretu, A. L., Ruddy, K., Germann, M., & Wenderoth, N. (2019). Uncertainty in contextual and kinematic cues jointly modulates motor resonance in primary motor cortex. *Journal of Neurophysiology*, 121(4), 1451-1464.
- Crivello, C., & Poulin-Dubois, D. (2018). Infants’ false belief understanding: A non-replication of the helping task. *Cognitive Development*, 46, 51-57.
- Csibra, G. (2007). Action mirroring and action understanding: an alternative account. In P. Haggard, Y. Rosetti, & M. Kawato (Eds.), *Sensorimotor Foundations of Higher Cognition. Attention and Performance XXII* (pp. 435-459). Oxford: Oxford University Press.
- Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science*, 1(2), 255-259.
- Csibra, G., & Gergely, G. (2007). ‘Obsessed with goals’: Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, 124(1), 60-78.
- Csibra, G., & Gergely, G. (2013). Teleological understanding of actions. In M.R. Banaji & S.A. Gelman (Eds), *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us* (pp. 38-43). Oxford: Oxford University Press.
- 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.
- Cuevas, K., & Paulus, M. (2016). The Development of Action Mirroring. *The British Journal of Developmental Psychology*, 34(1), 1.
- Curioni, A., Vesper, C., Knoblich, G., & Sebanz, N. (2019). Reciprocal information flow and role distribution support joint action coordination. *Cognition*, 187, 21-31.
- Daum, M. M., Prinz, W., & Aschersleben, G. (2008). Encoding the goal of an object-directed but

- uncompleted reaching action in 6-and 9-month-old infants. *Developmental Science*, 11(4), 607-619.
- Davidson, D. (1980). *Essays on Actions and Events*. Oxford, Clarendon Press.
- de Klerk, C. C., Hamilton, A. F. D. C., & Southgate, V. (2018). Eye contact modulates facial mimicry in 4-month-old infants: An EMG and fNIRS study. *Cortex*, 106, 93-103.
- de Klerk, C. C., Southgate, V., & Csibra, G. (2016). Predictive action tracking without motor experience in 8-month-old infants. *Brain and cognition*, 109, 131-139.
- Decety, J. and Sommerville, J.A. (2003) Shared representations between self and other: A social cognitive neuroscience view. *Trends in Cognitive Sciences*, 7, 527–533.
- 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.
- Denison, S., & Xu, F. (2010). Twelve-to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science*, 13(5), 798-803.
- Denison, S., Reed, C., & Xu, F. (2013). The emergence of probabilistic reasoning in very young infants: Evidence from 4.5-and 6-month-olds. *Developmental Psychology*, 49(2), 243.
- D'Entremont, B., Hains, S. M., & Muir, D. W. (1997). A demonstration of gaze following in 3-to 6-month-olds. *Infant Behavior and Development*, 20(4), 569-572.
- Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child development*, 85(5), 1777-1794.
- Di Gangi, V., Farroni, T., & Southgate, V. (2014) Infants exhibit a negativity bias when interpreting others' choices. In *4<sup>th</sup> Budapest CEU Conference on Cognitive Development*, Budapest, Hungary.
- Dittrich, K., Dolk, T., Rothe-Wulf, A., Klauer, K. C., & Prinz, W. (2013). Keys and seats: spatial response coding underlying the joint spatial compatibility effect. *Attention, Perception, & Psychophysics*, 75(8), 1725-1736.
- Dittrich, K., Rothe, A., & Klauer, K. C. (2012). Increased spatial salience in the social Simon task: a response-coding account of spatial compatibility effects. *Attention, Perception, &*

- Psychophysics*, 74(5), 911-929.
- Dolk, T., Hommel, B., Colzato, L. S., Schütz-Bosbach, S., Prinz, W., & Liepelt, R. (2011). How “social” is the social Simon effect?. *Frontiers in Psychology*, 2, 84.
- Dolk, T., Hommel, B., Colzato, L. S., Schütz-Bosbach, S., Prinz, W., & Liepelt, R. (2014). The joint Simon effect: a review and theoretical integration. *Frontiers in Psychology*, 5, 974.
- Dolk, T., Hommel, B., Prinz, W., & Liepelt, R. (2013). The (not so) social Simon effect: a referential coding account. *Journal of Experimental Psychology: Human Perception and Performance*, 39(5), 1248.
- Donnarumma, F., Dindo, H., & Pezzulo, G. (2017). Sensorimotor coarticulation in the execution and recognition of intentional actions. *Frontiers in Psychology*, 8, 237.
- Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). How (not) to measure infant Theory of Mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development*, 46, 12-30.
- Drea, C. M., & Carter, A. N. (2009). Cooperative problem solving in a social carnivore. *Animal Behaviour*, 78(4), 967-977.
- Dretske, F. (1988). Representational systems. In T. O'Connor & D. Robb (Eds). *Philosophy of Mind: Contemporary Readings*, (pp. 304-331). Routledge.
- Duh, S., Paik, J. H., Miller, P. H., Gluck, S. C., Li, H., & Himelfarb, I. (2016). Theory of mind and executive function in Chinese preschool children. *Developmental Psychology*, 52(4), 582.
- Elekes, F., Bródy, G., Halász, E., & Király, I. (2016). Enhanced encoding of the co-actor's target stimuli during a shared non-motor task. *Quarterly Journal of Experimental Psychology*, 69(12), 2376-2389.
- Elsner, B. (2007). Infants' imitation of goal-directed actions: The role of movements and action effects. *Acta Psychologica*, 124(1), 44-59.
- Elsner, B., & Adam, M. (2020). Infants' Goal Prediction for Simple Action Events: The Role of Experience and Agency Cues. *Topics in Cognitive Science*, 1-18.
- Elsner, B., & Aschersleben, G. (2003). Do I get what you get? Learning about the effects of self-

- performed and observed actions in infancy. *Consciousness and Cognition*, 12(4), 732-751.
- Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 229.
- Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, 2(10), 704-716.
- Fantasia, V., Fasulo, A., Costall, A., & López, B. (2014). Changing the game: exploring infants' participation in early play routines. *Frontiers in Psychology*, 5, 522.
- Fawcett, C. A., & Markson, L. (2010). Similarity predicts liking in 3-year-old children. *Journal of Experimental Child Psychology*, 105(4), 345-358.
- Fawcett, C., & Gredebäck, G. (2013). Infants use social context to bind actions into a collaborative sequence. *Developmental Science*, 16(6), 841-849.
- Feiman, R., Carey, S., & Cushman, F. (2015). Infants' representations of others' goals: Representing approach over avoidance. *Cognition*, 136, 204-214.
- Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 4, 215.
- Fiebich, A. (2020). *Minimal Cooperation and Shared Agency*. New York, États-Unis: Springer Publishing.
- Filimon, F., Nelson, J. D., Hagler, D. J., & Sereno, M. I. (2007). Human cortical representations for reaching: mirror neurons for execution, observation, and imagery. *Neuroimage*, 37(4), 1315-1328.
- Filippetti, M. L., Johnson, M. H., Lloyd-Fox, S., Dragovic, D., & Farroni, T. (2013). Body perception in newborns. *Current Biology*, 23(23), 2413-2416.
- Filippi, C. A., Cannon, E. N., Fox, N. A., Thorpe, S. G., Ferrari, P. F., & Woodward, A. L. (2016). Motor system activation predicts goal imitation in 7-month-old infants. *Psychological Science*, 27(5), 675-684.
- Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental*

- Psychology*, 17(1), 99.
- Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental psychology*, 17(1), 99.
- Fogassi, L., & Gallese, V. (2002). The neural correlates of action understanding in non-human primates. In M. I. Stamenov & V. Gallese (Eds.), *Mirror neurons and the evolution of brain and language* (pp. 13–35). Amsterdam: John Benjamins.
- Forgács, B., Gervain, J., Parise, E., Csibra, G., Gergely, G., Baross, J., & Király, I. (2020). Electrophysiological investigation of infants' understanding of understanding. *Developmental Cognitive Neuroscience*, 100783.
- Forgács, B., Parise, E., Csibra, G., Gergely, G., Jacquey, L., & Gervain, J. (2019). Fourteen-month-old infants track the language comprehension of communicative partners. *Developmental science*, 22(2), e12751.
- Fox, N. A., Bakermans-Kranenburg, M. J., Yoo, K. H., Bowman, L. C., Cannon, E. N., Vanderwert, R. E., Ferrari, P. F., & Van IJzendoorn, M. H. (2016). Assessing human mirror activity with EEG mu rhythm: A meta-analysis. *Psychological Bulletin*, 142(3), 291.
- Franks, N. R. (1989). Thermoregulation in army ant bivouacs. *Physiological Entomology*, 14(4), 397-404.
- Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2016). When do humans spontaneously adopt another's visuospatial perspective?. *Journal of Experimental Psychology: Human Perception and Performance*, 42(3), 401-412.
- Freundlieb, M., Sebanz, N., & Kovács, Á. M. (2017). Out of your sight, out of my mind: Knowledge about another person's visual access modulates spontaneous visuospatial perspective-taking. *Journal of Experimental Psychology: Human Perception and Performance*, 43(6), 1065.
- Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, 16(9), 1325-1352.
- Friston, K. J. (2005). Hallucinations and perceptual inference. *Behavioral and Brain Sciences*, 28(6), 764-766.

- Friston, K. J., & Frith, C. D. (2015). Active inference, communication and hermeneutics. *Cortex*, 68, 129-143.
- Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355(1404), 1771-1788.
- Furlanetto, T., Cavallo, A., Manera, V., Tversky, B., & Becchio, C. (2013). Through your eyes: incongruence of gaze and action increases spontaneous perspective taking. *Frontiers in Human Neuroscience*, 7, 455.
- Gamepe, A., Prinz, W., & Daum, M. M. (2016). Measuring action understanding: relations between goal prediction and imitation. *British Journal of Developmental Psychology*, 34(1), 53-65.
- Ganea, P. A., & Saylor, M. M. (2007). Infants' use of shared linguistic information to clarify ambiguous requests. *Child Development*, 78(2), 493-502.
- Ganesh, G., Takagi, A., Osu, R., Yoshioka, T., Kawato, M., & Burdet, E. (2014). Two is better than one: Physical interactions improve motor performance in humans. *Scientific Reports*, 4, 3824.
- Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences*, 7(7), 287-292.
- Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, 56(2), 165-193.
- Gerson, S. A., Bekkering, H., & Hunnius, S. (2016). Social context influences planning ahead in three-year-olds. *Cognitive Development*, 40, 120-131.
- Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 1/2-7 years old on a Stroop-like day-night test. *Cognition*, 53(2), 129-153.
- Gilbert, M. (1987). Modelling collective belief. *Synthese*, 73(1), 185-204.
- Gilbert, M. (1989). *On Social Facts*. Princeton: Princeton University Press.
- Gilbert, M. (1990). Walking together: A paradigmatic social phenomenon. *MidWest Studies in Philosophy*, 15, 1-14.

- Gilbert, M. (1994). Me, you, and us: Distinguishing “egoism,” “altruism,” and “groupism”. *Behavioral and Brain Sciences*, 17(4), 621-622.
- Gilbert, M. (1999). Obligation and joint commitment. *Utilitas*, 11(2), 143-163.
- Gilbert, M. (2000). *Sociality and Responsibility: New Essays in Plural Subject Theory*. Lanham: Rowman and Littlefield.
- Gilbert, M. (2006). *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*, Oxford: Oxford University Press.
- Gilbert, M. (2009). Shared intention and personal intentions, *Philosophical Studies*, 144: 167–87.
- Gilbert, M. (2013). *Joint commitment: How we make the social world*. Oxford: Oxford University Press.
- Gilbert, M. (2018). *Rights and demands: A foundational inquiry*. Oxford: Oxford University Press.
- Giroto, V. & Gonzalez, M. (2006). Norms and intuitions in the assessment of chance. In L. Smith & J. Vonèche (Eds.). (2006). *Norms in human development*, (pp.220-236). Cambridge: Cambridge University Press
- Goebel, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception*, 26(5), 427-438.
- Goldman, A. (1970). *A theory of human action*. New York: Prentice Hall.
- Goswami, U. (2008). *Cognitive development—The learning brain* (2nd ed.). Hove, UK: Psychology Press.
- Gräfenhain, M., Behne, T., Carpenter, M., & Tomasello, M. (2009). Young children’s understanding of joint commitments. *Developmental Psychology*, 45(5), 1430.
- Graham, J. M., Kao, A. B., Wilhelm, D. A., & Garnier, S. (2017). Optimal construction of army ant living bridges. *Journal of Theoretical Biology*, 435, 184-198.
- Grezes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Human Brain Mapping*, 12(1), 1-19.
- Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2017). Implicit and explicit false belief development in preschool children. *Developmental Science*, 20(5), e12445.

- Grosse Wiesmann, C., Kampis, D., Koop, S., & Southgate, V. The development of self-concept: A novel task to assess mapping others to oneself. In *9th Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics, Oslo, Norway*.
- Guagnano, D., Rusconi, E., & Umiltà, C. A. (2010). Sharing a task or sharing space? On the effect of the confederate in action coding in a detection task. *Cognition*, 114(3), 348-355.
- Hamann, K., Warneken, F., & Tomasello, M. (2012). Children's developing commitments to joint goals. *Child Development*, 83(1), 137-145.
- Hamilton, A. F., & Grafton, S. T. (2007). The motor hierarchy: from kinematics to goals and intentions. *Sensorimotor foundations of higher cognition*, 22, 381-408.
- Hamlin, J. K., & Wynn, K. (2012). Who knows what's good to eat? Infants fail to match the food preferences of antisocial others. *Cognitive Development*, 27(3), 227-239.
- Hamlin, J. K., Mahajan, N., Liberman, Z., & Wynn, K. (2013). Not like me= bad: Infants prefer those who harm dissimilar others. *Psychological Science*, 24(4), 589-594.
- Hammond, S. I., Müller, U., Carpendale, J. I., Bibok, M. B., & Liebermann-Finestone, D. P. (2012). The effects of parental scaffolding on preschoolers' executive function. *Developmental Psychology*, 48(1), 271.
- Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know?. *Animal Behaviour*, 61(1), 139-151.
- Hassin, R. R., Aarts, H. & Ferguson, M. (2005). Automatic goal inferences. *Journal of Experimental Social Psychology*. 41(2), 129-40.
- Hay, D. F. (1979). Cooperative interactions and sharing between very young children and their parents. *Developmental Psychology*, 15(6), 647.
- He, Z., Bolz, M., & Baillargeon, R. (2011). False-belief understanding in 2.5-year-olds: evidence from violation-of-expectation change-of-location and unexpected-contents tasks. *Developmental Science*, 14(2), 292-305.
- He, Z., Bolz, M., & Baillargeon, R. (2012). 2.5-year-olds succeed at a verbal anticipatory-looking false-belief task. *British Journal of Developmental Psychology*, 30(1), 14-29.

- Henderson, A. M., & Woodward, A. L. (2011). "Let's work together": What do infants understand about collaborative goals?. *Cognition*, 121(1), 12-21.
- Henderson, A. M., Wang, Y., Matz, L. E., & Woodward, A. L. (2013). Active experience shapes 10-month-old infants' understanding of collaborative goals. *Infancy*, 18(1), 10-39.
- Henrichs, I., Elsner, C., Elsner, B., Wilkinson, N., & Gredebäck, G. (2014). Goal certainty modulates infants' goal-directed gaze shifts. *Developmental Psychology*, 50(1), 100.
- Heyes, C. (2011). Automatic imitation. *Psychological Bulletin*, 137(3), 463.
- Hobson, R. P. (2005). What Puts the Jointness into Joint Attention?. *Joint attention: Communication and Other Minds: Issues in Philosophy and Psychology*, 185.
- Hohwy, J. (2013). *The predictive mind*. New-York: Oxford University Press.
- Hommel, B. (1996). SR compatibility effects without response uncertainty. *The Quarterly Journal of Experimental Psychology Section A*, 49(3), 546-571.
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24(5), 849-878.
- Howard, L. H., & Woodward, A. L. (2019). Human actions support infant memory. *Journal of Cognition and Development*, 20(5), 772-789.
- Hudson, M., Bach, P., & Nicholson, T. (2018). You said you would! The predictability of other's behavior from their intentions determines predictive biases in action perception. *Journal of Experimental Psychology: Human Perception and Performance*, 44(2), 320.
- Hudson, M., McDonough, K. L., Edwards, R., & Bach, P. (2018). Perceptual teleology: expectations of action efficiency bias social perception. *Proceedings of the Royal Society B: Biological Sciences*, 285(1884), 20180638.
- Hudson, M., Nicholson, T., Ellis, R., & Bach, P. (2016). I see what you say: Prior knowledge of other's goals automatically biases the perception of their actions. *Cognition*, 146, 245-250.
- Hudson, M., Nicholson, T., Simpson, W. A., Ellis, R., & Bach, P. (2016). One step ahead: The perceived kinematics of others' actions are biased toward expected goals. *Journal of Experimental Psychology: General*, 145(1), 1.

- Hughes, C., Jaffee, S. R., Happé, F., Taylor, A., Caspi, A., & Moffitt, T. E. (2005). Origins of individual differences in theory of mind: From nature to nurture?. *Child Development, 76*(2), 356-370.
- Hunnius, S., & Bekkering, H. (2014). What are you doing? How active and observational experience shape infants' action understanding. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1644), 20130490.
- Iani, C., Anelli, F., Nicoletti, R., & Rubichi, S. (2014). The carry-over effect of competition in task-sharing: Evidence from the joint Simon task. *PloS One, 9*(6).
- Iani, C., Anelli, F., Nicoletti, R., Arcuri, L., & Rubichi, S. (2011). The role of group membership on the modulation of joint action. *Experimental Brain Research, 211*(3-4), 439.
- Iani, C., Ciardo, F., Panajoli, S., Lugli, L., & Rubichi, S. (2019). The role of the co-actor's response reachability in the joint Simon effect: remapping of working space by tool use. *Psychological Research, 1-12*.
- Ishikawa, M. and Itakura, S. (2018) Observing others' gaze direction affects infants' preference for looking at gazing- or gazed-at faces. *Frontiers in Psychology, 9*, 1–7.
- Jacob, P. (2020). What Do False-Belief Tests Show?. *Review of Philosophy and Psychology, 11*(1), 1-20.
- Jeannerod, M. (1997) *The cognitive neuroscience of action*. Oxford: Blackwell.
- Jeannerod, M. (2006). *Motor cognition: What actions tell the self* (No. 42). Oxford University Press.
- Jordan, M. I., & Wolpert, D. M. (1999). Computational motor control. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 601–620). Cambridge, MA: MIT Press.
- Jovanovic, B., Király, I., Elsner, B., Gergely, G., Prinz, W., & Aschersleben, G. (2007). The role of effects for infants' perception of action goals. *Psychologia, 50*(4), 273-290.
- Kagan, J. (2013). *The second year: The emergence of self-awareness*. Harvard University Press.
- Kampis, D., & Southgate, V. (2020). Altercentric Cognition: How Others Influence Our Cognitive Processing. *Trends in cognitive sciences, 1-15*.
- Kampis, D., Parise, E., Csibra, G., & Kovács, Á. M. (2015). Neural signatures for sustaining object

- representations attributed to others in preverbal human infants. *Proceedings of the Royal Society B: Biological Sciences*, 282(1819), 20151683.
- Kampis, D., Somogyi, E., Itakura, S., & Király, I. (2013). Do infants bind mental states to agents?. *Cognition*, 129(2), 232-240.
- Kampis, D., Wiesmann, C. G., Koop, S., & Southgate, V. (2020). The social self: Infants spontaneously map others' face to their own at 16 to 26 months. *PsyArXiv Preprints* <https://psyarxiv.com/ky35p/>
- Karlinsky, A., Lam, M. Y., Chua, R., & Hodges, N. J. (2019). Whose turn is it anyway? The moderating role of response-execution certainty on the joint Simon effect. *Psychological Research*, 83(5), 833-841.
- Kärtner, J., Keller, H., & Chaudhary, N. (2010). Cognitive and social influences on early prosocial behavior in two sociocultural contexts. *Developmental Psychology*, 46(4), 905.
- Keller, P. E., Knoblich, G., & Repp, B. H. (2007). Pianists duet better when they play with themselves: on the possible role of action simulation in synchronization. *Consciousness and Cognition*, 16(1), 102-111.
- Kelso, J. S., Dumas, G., & Tognoli, E. (2013). Outline of a general theory of behavior and brain coordination. *Neural Networks*, 37, 120-131.
- Kenward, B., & Gredebäck, G. (2013). Infants help a non-human agent. *PloS One*, 8(9).
- Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, 55, 271-304.
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89(1), 25-41.
- Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological movement on action. *Current Biology*, 13(6), 522-525.
- Kilner, J. M., Vargas, C., Duval, S., Blakemore, S. J., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature Neuroscience*, 7(12), 1299-1301.
- Kilner, J.M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences*.

15, 352–357.

- King, S. L., Allen, S. J., Connor, R. C., & Jaakkola, K. (2016). Cooperation or dolphin ‘tug-of-war’? Comment on Kuczaj et al. and Eskelinen et al. *Animal Cognition*, 19(6), 1227-1229.
- Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, 104(30), 12577-12580.
- Király, I., Jovanovic, B., Prinz, W., Aschersleben, G., & Gergely, G. (2003). The early origins of goal attribution in infancy. *Consciousness and Cognition*, 12(4), 752-769.
- Klempova, B., & Liepelt, R. (2016). Do you really represent my task? Sequential adaptation effects to unexpected events support referential coding for the joint Simon effect. *Psychological Research*, 80(4), 449-463.
- Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1499), 2021-2031.
- Knoblich G., Butterfill S., & Sebanz, N. (2011). Psychological research on joint action: Theory and data. In Ross B., (Ed.). *The psychology of learning and motivation*. Burlington: Academic Press, Vol. 54: 59–101.
- Knudsen, B., & Liszkowski, U. (2012). Eighteen- and 24-month-old infants correct others in anticipation of action mistakes. *Developmental Science*, 15(1), 113-122.
- Koban, L., Ramamoorthy, A., & Konvalinka, I. (2019). Why do we fall into sync with others? Interpersonal synchronization and the brain’s optimization principle. *Social Neuroscience*, 14(1), 1-9.
- Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development*, 67(2), 490-507.
- Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others’ beliefs in human infants and adults. *Science*, 330(6012), 1830-1834.
- Krogh, L., Vlach, H., & Johnson, S. P. (2013). Statistical learning across development: Flexible yet

- constrained. *Frontiers in Psychology*, 3, 598.
- Kuczaj, S. A., Winship, K. A., & Eskelinen, H. C. (2015). Can bottlenose dolphins (*Tursiops truncatus*) cooperate when solving a novel task?. *Animal Cognition*, 18(2), 543-550.
- Kuhbandner, C., Pekrun, R., & Maier, M. A. (2010). The role of positive and negative affect in the “mirroring” of other persons' actions. *Cognition and Emotion*, 24(7), 1182-1190.
- Kulke, L., & Rakoczy, H. (2017). How reliable and valid are anticipatory looking measures in theory of mind task? In H. Rakoczy & L. Kulke (Chairs), *Are implicit theory of mind findings robust? Some doubts from converging non-replications across the lifespan*. Symposium conducted at the Society for Research in Child Development Biennial Meeting. Austin, Texas.
- Kulke, L., & Rakoczy, H. (2018). Implicit Theory of Mind—An overview of current replications and non-replications. *Data in Brief*, 16, 101-104.
- Kulke, L., Reiß, M., Krist, H., & Rakoczy, H. (2018). How robust are anticipatory looking measures of Theory of Mind? Replication attempts across the life span. *Cognitive Development*, 46, 97-111.
- Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science*, 16(9), 678-683.
- Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology*, 43(1), 186.
- Leslie, A. M. (1984). Spatiotemporal continuity and the perception of causality in infants. *Perception*, 13(3), 287-305.
- Leslie, A. M. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. *Mapping the mind: Domain specificity in cognition and culture*, 29, 119-48.
- Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science*, 1(2), 247-253.
- Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology*, 50(1), 45-85.

- Leslie, A.M. (2005). Developmental Parallels in Understanding Minds and Bodies. *Trends in Cognitive Sciences*, 9(10), 459-462.
- Lewis, D. (1969). *Convention: A Philosophical Study*. Hoboken, NJ: Wiley-Blackwell.
- Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Developmental Science*, 12(2), 264-271.
- Liebal, K., Carpenter, M., & Tomasello, M. (2010). Infants' use of shared experience in declarative pointing. *Infancy*, 15(5), 545-556.
- Liepelt, R., Wenke, D., & Fischer, R. (2013). Effects of feature integration in a hands-crossed version of the social Simon paradigm. *Psychological Research*, 77(2), 240-248.
- Liepelt, R., Wenke, D., Fischer, R., & Prinz, W. (2011). Trial-to-trial sequential dependencies in a social and non-social Simon task. *Psychological Research*, 75(5), 366-375.
- Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, 108(3), 732-739.
- Loucks, J., & Sommerville, J. A. (2012). The role of motor experience in understanding action function: The case of the precision grasp. *Child Development*, 83(3), 801-809.
- Low, J., & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science*, 24(3), 305-311.
- Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives*, 10(3), 184-189.
- Low, J., Drummond, W., Walmsley, A., & Wang, B. (2014). Representing how rabbits quack and competitors act: Limits on preschoolers' efficient ability to track perspective. *Child Development*, 85(4), 1519-1534.
- Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science*, 16(8), 601-608.
- Mahajan, N., & Wynn, K. (2012). Origins of "us" versus "them": Prelinguistic infants prefer similar others. *Cognition*, 124(2), 227-233.

- Marshall, P. J., & Meltzoff, A. N. (2011). Neural mirroring systems: Exploring the EEG mu rhythm in human infancy. *Developmental Cognitive Neuroscience*, 1(2), 110-123.
- Marshall, P. J., Saby, J. N., & Meltzoff, A. N. (2013). Imitation and the developing social brain: infants somatotopic EEG patterns for acts of self and other. *International Journal of Psychological Research*, 6(spe), 22-29.
- Masumoto, J., & Inui, N. (2013). Two heads are better than one: Both complementary and synchronous strategies facilitate joint action. *Journal of Neurophysiology*, 109(5), 1307-1314.
- Masumoto, J., & Inui, N. (2014a). A leader–follower relationship in joint action on a discrete force production task. *Experimental Brain Research*, 232(11), 3525-3533.
- Masumoto, J., & Inui, N. (2014b). Effects of speech on both complementary and synchronous strategies in joint action. *Experimental Brain Research*, 232(7), 2421-2429.
- Mazzarella, E., Ramsey, R., Conson, M., & Hamilton, A. (2013). Brain systems for visual perspective taking and action perception. *Social Neuroscience*, 8(3), 248-267.
- McDonough, K. L., Costantini, M., Hudson, M., Ward, E., & Bach, P. (2020). Affordance matching predictively shapes the perceptual representation of others' ongoing actions. *Journal of Experimental Psychology: Human Perception and Performance*. 46(8), 847–859.
- McDonough, K. L., Hudson, M., & Bach, P. (2019). Cues to intention bias action perception toward the most efficient trajectory. *Scientific Reports*, 9(1), 1-10.
- McEllin, L., Knoblich, G., & Sebanz, N. (2018). Distinct kinematic markers of demonstration and joint action coordination? Evidence from virtual xylophone playing. *Journal of Experimental Psychology: Human Perception and Performance*, 44(6), 885.
- Mead, G. H. (1934). *Mind, self and society* (Vol. 111). University of Chicago Press.
- Meltzoff, A. N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31(5), 838.
- Memelink, J., & Hommel, B. (2013). Intentional weighting: a basic principle in cognitive control. *Psychological Research*, 77(3), 249-259.
- Meyer, M., Bekkering, H., Haartsen, R., Stapel, J. C., & Hunnius, S. (2015). The role of action

- prediction and inhibitory control for joint action coordination in toddlers. *Journal of Experimental Child Psychology*, 139, 203-220.
- Meyer, M., van der Wel, R. P., & Hunnius, S. (2016). Planning my actions to accommodate yours: joint action development during early childhood. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1693), 20150371.
- Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2020). Two systems for thinking about others' thoughts in the developing brain. *Proceedings of the National Academy of Sciences*, 117(12), 6928-6935.
- Milward, S. J., Kita, S., & Apperly, I. A. (2014). The development of co-representation effects in a joint task: Do children represent a co-actor?. *Cognition*, 132(3), 269-279.
- Miss, F. M., & Burkart, J. M. (2018). Corepresentation during joint action in marmoset monkeys (*Callithrix jacchus*). *Psychological Science*, 29(6), 984-995.
- Mlot, N. J., Tovey, C. A., & Hu, D. L. (2011). Fire ants self-assemble into waterproof rafts to survive floods. *Proceedings of the National Academy of Sciences*, 108(19), 7669-7673.
- Mlot, N. J., Tovey, C., & Hu, D. L. (2012). Dynamics and shape of large fire ant rafts. *Communicative & Integrative Biology*, 5(6), 590-597.
- Moeller, M. P., & Schick, B. (2006). Relations between maternal input and theory of mind understanding in deaf children. *Child Development*, 77(3), 751-766.
- Moll, H., & Meltzoff, A. N. (2011a). How does it look? Level 2 perspective-taking at 36 months of age. *Child Development*, 82(2), 661-673.
- Moll, H., & Meltzoff, A. N. (2011b). Perspective-taking and its foundation in joint attention. In N. Eilan, H. Lerman, & J. Roessler (Eds.), *Perception, causation, and objectivity. Issues in philosophy and psychology* (pp. 286-304). Oxford, England: Oxford University Press.
- Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology*, 24(3), 603-613.
- Monroy, C. D., Meyer, M., Schröder, L., Gerson, S. A., & Hunnius, S. (2019). The infant motor system predicts actions based on visual statistical learning. *Neuroimage*, 185, 947-954.

- Moore, C., & D'Entremont, B. (2001). Developmental changes in pointing as a function of attentional focus. *Journal of Cognition and Development, 2*(2), 109-129.
- Morewedge, C. K., Preston, J., & Wegner, D. M. (2007). Timescale bias in the attribution of mind. *Journal of Personality and Social Psychology, 93*(1), 1.
- Mylopoulos, M., & Pacherie, E. (2017). Intentions and motor representations: The interface challenge. *Review of Philosophy and Psychology, 8*(2), 317-336.
- Mylopoulos, M., & Pacherie, E. (2019). Intentions: The dynamic hierarchical model revisited. *Wiley Interdisciplinary Reviews: Cognitive Science, 10*(2), e1481.
- Naish, K. R., Reader, A. T., Houston-Price, C., Bremner, A. J., & Holmes, N. P. (2013). To eat or not to eat? Kinematics and muscle activity of reach-to-grasp movements are influenced by the action goal, but observers do not detect these differences. *Experimental Brain Research, 225*(2), 261-275.
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*, 353-383.
- Needham, A. W., Wiesen, S. E., Hejazi, J. N., Libertus, K., & Christopher, C. (2017). Characteristics of brief sticky mittens training that lead to increases in object exploration. *Journal of Experimental Child Psychology, 164*, 209-224.
- Needham, A., Barrett, T., & Peterman, K. (2002). A pick-me-up for infants' exploratory skills: Early simulated experiences reaching for objects using 'sticky mittens' enhances young infants' object exploration skills. *Infant Behavior and Development, 25*(3), 279-295.
- Oates, J., & Grayson, A. (2004). Introduction: perspectives on cognitive and language development. In J., Oates, *Cognitive and Language Development in Children*. Open University; Malden, MA: Blackwell Pub, 8-20.
- Oh, H., Braun, A. R., Reggia, J. A., & Gentili, R. J. (2019). Fronto-parietal mirror neuron system modeling: visuospatial transformations support imitation learning independently of imitator perspective. *Human Movement Science, 65*, 121-141.
- Oktay-Gür, N., Schulz, A., & Rakoczy, H. (2018). Children exhibit different performance patterns in explicit and implicit theory of mind tasks. *Cognition, 173*, 60-74.

- Okumura, Y. *et al.* (2013) The power of human gaze on infant learning. *Cognition* 128, 127–133.
- Ondobaka, S., de Lange, F. P., Wittmann, M., Frith, C. D., & Bekkering, H. (2015). Interplay between conceptual expectations and movement predictions underlies action understanding. *Cerebral Cortex*, 25(9), 2566-2573.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?. *Science*, 308(5719), 255-258.
- O'Sullivan, I., Burdet, E., & Diedrichsen, J. (2009). Dissociating variability and effort as determinants of coordination. *PLoS Computational Biology*, 5(4), e1000345.
- Oullier, O., De Guzman, G. C., Jantzen, K. J., Lagarde, J., & Scott Kelso, J. A. (2008). Social coordination dynamics: Measuring human bonding. *Social Neuroscience*, 3(2), 178-192.
- Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, 107(1), 179-217.
- Pacherie, E. (2011). Framing Joint Action. *Review of Philosophy and Psychology*, 2 (2), 173- 192.
- Pacherie, E. (2011). Framing joint action. *Review of Philosophy and Psychology*, 2(2), 173-192.
- Pacherie, E. (2012). Action. In K., Frankish, & W., Ramsey, (Eds). *Cambridge Handbook of Cognitive Science*, (pp. 92-111). Cambridge: Cambridge University Press.
- Pacherie, E., & Dokic, J. (2006). From mirror neurons to joint actions. *Cognitive Systems Research*, 7(2-3), 101-112.
- Paternotte, C. (2020). Joint action: why so minimal? In In A. Fiebich (ed.), *Minimal Cooperation and Shared Agency* (pp. 41-58). Dordrecht: Springer.
- Paulus, M. (2012). Action mirroring and action understanding: An ideomotor and attentional account. *Psychological Research*, 76(6), 760-767.
- Paulus, M. (2016). The development of action planning in a joint action context. *Developmental Psychology*, 52(7), 1052.
- Paulus, M., Hunnius, S., Van Elk, M., & Bekkering, H. (2012). How learning to shake a rattle affects 8-month-old infants' perception of the rattle's sound: Electrophysiological evidence for action-effect binding in infancy. *Developmental Cognitive Neuroscience*, 2(1), 90-96.

- Paulus, M., Hunnius, S., van Wijngaarden, C., Vrins, S., van Rooij, I., & Bekkering, H. (2011). The role of frequency information and teleological reasoning in infants' and adults' action prediction. *Developmental Psychology*, 47(4), 976–983.
- Perner, J. (1991). *Understanding the representational mind*. MIT Press.
- Perner, J., & Esken, F. (2015). Evolution of human cooperation in *Homo heidelbergensis*: Teleology versus mentalism. *Developmental Review*, 38, 69-88.
- Perner, J., & Roessler, J. (2010). Teleology and causal understanding in childrens' theory of mind. In Aguilar, J. H. & Buckareff, A. A.(Eds.), *Causing human action: New perspectives on the causal theory of action* (pp. 199-228). Cambridge, MA: MIT Press.
- Perner, J., & Roessler, J. (2012). From infants' to children's appreciation of belief. *Trends in Cognitive Sciences*, 16(10), 519-525.
- Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep?. *Science*, 308(5719), 214-216.
- Perner, J., Brandl, J., & Garnham, A. (2003). What is a perspective problem? Developmental issues in understanding belief and dual identity. *Facta Philosophica*, 5: 355–378.
- Perner, J., Priewasser, B., & Roessler, J. (2018). The practical other: teleology and its development. *Interdisciplinary Science Reviews*, 43(2), 99-114.
- Perner, J., Stummer, S., Sprung, M., & Doherty, M. (2002). Theory of mind finds its Piagetian perspective: Why alternative naming comes with understanding belief. *Cognitive Development*, 17(3-4), 1451-1472.
- Peterson, C., & Slaughter, V. (2003). Opening windows into the mind: Mothers' preferences for mental state explanations and children's theory of mind. *Cognitive Development*, 18(3), 399-429.
- Pezzulo, G. (2011a). Shared representations as coordination tools for interaction. *Review of Philosophy and Psychology*, 2(2), 303-333.
- Pezzulo, G. (2011b). Grounding procedural and declarative knowledge in sensorimotor anticipation. *Mind & Language*, 26(1), 78-114.

- Pezzulo, G. (2013). Studying mirror mechanisms within generative and predictive architectures for joint action. *Cortex*, 49 (10): 2968-2969.
- Pezzulo, G., & Dindo, H. (2011). What should I do next? Using shared representations to solve interaction problems. *Experimental Brain Research*, 211(3-4), 613-630.
- Pezzulo, G., & Dindo, H. (2013). Intentional strategies that make co-actors more predictable: the case of signaling. *Behavioral and Brain Sciences*, 36(4), 371.
- Pezzulo, G., Barca, L., Maisto, D., & Donnarumma, F. (2020). Social epistemic actions. *Behavioral and Brain Sciences*, 43.
- Pezzulo, G., Donnarumma, F., & Dindo, H. (2013). Human sensorimotor communication: A theory of signaling in online social interactions. *PloS One*, 8(11).
- Pezzulo, G., Donnarumma, F., Dindo, H., D'Ausilio, A., Konvalinka, I., & Castelfranchi, C. (2019). The body talks: Sensorimotor communication and its brain and kinematic signatures. *Physics of Life Reviews*, 28, 1-21.
- Pezzulo, G., Donnarumma, F., Iodice, P., Maisto, D., & Stoianov, I. (2017). Model-based approaches to active perception and control. *Entropy*, 19(6), 266.
- Pezzulo, G., Iodice, P., Donnarumma, F., Dindo, H., & Knoblich, G. (2017). Avoiding accidents at the champagne reception: A study of joint lifting and balancing. *Psychological Science*, 28(3), 338-345.
- Pezzulo, G., Iodice, P., Ferraina, S., & Kessler, K. (2013). Shared action spaces: a basis function framework for social re-calibration of sensorimotor representations supporting joint action. *Frontiers in Human Neuroscience*, 7, 800.
- Pezzulo, G., Rigoli, F., & Friston, K. J. (2018). Hierarchical active inference: A theory of motivated control. *Trends in Cognitive Sciences*, 22(4), 294-306.
- Phillips, A. T., Wellman, H. M., & Spelke, E. S. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*, 85(1), 53-78.
- Piaget, J., & Inhelder, B. (1956). *The child's conception of space* (F. J. Langdon & J. L. Lunzer, Trans.). Atlantic Highlands, NJ: Humanities Press. Reprinted in H., E., Gruber, & J., J.,

- Voneche (Eds.). (1977). *The essential Piaget: An interpretive reference and guide*, pp. 572-624. New York: Basic Books.
- Pomiechowska, B., & Csibra, G. (2017). Motor activation during action perception depends on action interpretation. *Neuropsychologia*, 105, 84-91.
- Poulin-Dubois, D., Rakoczy, H., Burnside, K., Crivello, C., Dörrenberg, S., Edwards, K., Krist, H., Kulke, L., Liszkowski, U., Low, J., Perner, J., Powell, L., Priewasser, B., Rafetseder, E., Ruffman, T. (2018). Do infants understand false beliefs? We don't know yet—A commentary on Baillargeon, Buttelmann and Southgate's commentary. *Cognitive Development*, 48, 302-315.
- Povinelli, D. J. (1995). The unduplicated self. In P. Rochat (Ed.), *The self in infancy: Theory and research* (pp. 161–192). Amsterdam: North-Holland/Elsevier Science.
- Povinelli, D. J. (2001). The self: Elevated in consciousness and extended in time. In C. Moore & K. Lemmon (Eds.), *The self in time: Developmental perspectives* (pp. 75–95). Mahaw, NJ: Lawrence Erlbaum Associates.
- Powell, L. J., Hobbs, K., Bardis, A., Carey, S., & Saxe, R. (2018). Replications of implicit theory of mind tasks with varying representational demands. *Cognitive Development*, 46, 40-50.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and Brain Sciences*, 1(4), 515-526.
- Press, C. (2011). Action observation and robotic agents: learning and anthropomorphism. *Neuroscience & Biobehavioral Reviews*, 35(6), 1410-1418.
- Priewasser, B., Rafetseder, E., Gargitter, C., & Perner, J. (2018). Helping as an early indicator of a theory of mind: Mentalism or Teleology?. *Cognitive Development*, 46, 69-78.
- Prinz, W. (1990). A common coding approach to perception and action. In O., Neumann, & W., Prinz, (Eds.), *Relationships between perception and action* (pp. 167-201). Berlin, Germany: Springer-Verlag.
- Puffe, L., Dittrich, K., & Klauer, K. C. (2017). The influence of the Japanese waving cat on the joint spatial compatibility effect: a replication and extension of Dolk, Hommel, Prinz, and Liepelt (2013). *PloS One*, 12(9).
- Quinton, A. (1975). Social Objects, *Proceedings of the Aristotelian Society*, 75: 1–27.

- Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*(2), 230-236.
- Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate–inanimate distinction. *Psychological Bulletin*, *127*(2), 209.
- Rakoczy, H. (2006). Pretend play and the development of collective intentionality. *Cognitive Systems Research*, *7*(2-3), 113-127.
- Rakoczy, H. (2007). Play, games, and the development of collective intentionality. *New Directions for Child and Adolescent Development*, *2007*(115), 53-67.
- Rakoczy, H. (2018). Development of Collective Intentionality. *The Routledge Handbook of Collective Intentionality*, *1*, 407-419.
- Ray, M., de Grosbois, J., & Welsh, T. N. (2017). Index of difficulty and side of space are accommodated during the selection and planning of a joint action. *Human Movement Science*, *54*, 197-209.
- Reddy, V. (2003). On being the object of attention: implications for self–other consciousness. *Trends in Cognitive Sciences*, *7*(9), 397-402.
- Reddy, V. (2005). Before the ‘third element’: understanding attention to self. *Joint attention: Communication and Other Minds*, 85-109.
- Reed, K., Peshkin, M., Hartmann, M. J., Grabowecky, M., Patton, J., & Vishton, P. M. (2006). Haptically linked dyads: are two motor-control systems better than one?. *Psychological Science*, *17*, 365-366.
- Reid, C. R., Lutz, M. J., Powell, S., Kao, A. B., Couzin, I. D., & Garnier, S. (2015). Army ants dynamically adjust living bridges in response to a cost–benefit trade-off. *Proceedings of the National Academy of Sciences*, *112*(49), 15113-15118.
- Reid, V. M., & Striano, T. (2005). Adult gaze influences infant attention and object processing: implications for cognitive neuroscience. *European Journal of Neuroscience*, *21*(6), 1763-1766.
- Reid, V. M., Striano, T., Kaufman, J., & Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *Neuroreport*, *15*(16), 2553-2555.

- Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, 26(6), 867-891.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169-192.
- Rizzolatti, G., & Sinigaglia, C. (2007). Mirror neurons and motor intentionality. *Functional Neurology*, 22(4), 205.
- Rizzolatti, G., & Sinigaglia, C. (2008). *Mirrors in the brain: How our minds share actions and emotions*. USA, Oxford University Press.
- Rizzolatti, G., Cattaneo, L., Fabbri-Destro, M., & Rozzi, S. (2014). Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. *Physiological Reviews*, 94(2), 655-706.
- Robson, S. J., & Kuhlmeier, V. A. (2013). Selectivity promotes 9-month-old infants to encode the goals of others. In 6<sup>th</sup> Biennial Meeting of the Society for Research in Child Development, Seattle, WA.
- Rochat, P. (2001). Social contingency detection and infant development. *Bulletin of the Menninger Clinic*, 65(3: Special issue), 347-360.
- Rochat, P. (2003). Five levels of self-awareness as they unfold early in life. *Consciousness and Cognition*, 12(4), 717-731.
- Rochat, P. (2018). The ontogeny of human self-consciousness. *Current Directions in Psychological Science*, 27(5), 345-350.
- Rochat, P., & Goubet, N. (1995). Development of sitting and reaching in 5-to 6-month-old infants. *Infant Behavior and Development*, 18(1), 53-68.
- Rochat, P., & Goubet, N. (2000). Connaissance implicite du corps au début de la vie. *Enfance*, 53(3), 275-285.
- Rochat, P., & Hespos, S. J. (1997). Differential rooting response by neonates: Evidence for an early sense of self. *Infant and Child Development*, 6(3-4), 105-112.

- Rochat, P., & Passos-Ferreira, C. (2009). From imitation to reciprocation and mutual recognition. In J., Pineda (Ed.). *Mirror neuron systems: The role of mirroring processes in social cognition*, (pp. 191-212). San Diego, CA: Humana Press.
- Rochat, P., & Striano, T. (1999). Social-cognitive development in the first year. In P. Rochat. (ed.). *Early social cognition: Understanding others in the first months of life*, (pp.3-34). Erlbaum.
- Rochat, P., & Zahavi, D. (2011). The uncanny mirror: A re-framing of mirror self-experience. *Consciousness and Cognition*, 20(2), 204-213.
- Rochat, P., Broesch, T., & Jayne, K. (2012). Social awareness and early self-recognition. *Consciousness and Cognition*, 21(3), 1491-1497.
- Roessler, J., & Perner, J. (2013). Teleology: Belief as perspective. *Understanding Other Minds: Perspectives from Developmental Social Neuroscience*, 35-51.
- Rosenbaum, D. A., Chapman, K. M., Weigelt, M., Weiss, D. J., & van der Wel, R. (2012). Cognition, action, and object manipulation. *Psychological Bulletin*, 138(5), 924.
- Rosenbaum, D.A. (2009). *Human motor control*. San Diego, Academic Press.
- Rueschemeyer, S. A., Gardner, T., & Stoner, C. (2015). The Social N400 effect: how the presence of other listeners affects language comprehension. *Psychonomic Bulletin & Review*, 22(1), 128-134.
- Ruffman, T., & Perner, J. (2005). Do infants really understand false belief?: Response to Leslie. *Trends in Cognitive Sciences*, 9(10), 462-463.
- Ruffman, T., Perner, J., Naito, M., Parkin, L., & Clements, W. A. (1998). Older (but not younger) siblings facilitate false belief understanding. *Developmental Psychology*, 34(1), 161.
- Ruissen, M. I., & de Bruijn, E. R. (2016). Competitive game play attenuates self-other integration during joint task performance. *Frontiers in Psychology*, 7, 274.
- Ruissen, M. I., & de Bruijn, E. R. (2016). Competitive game play attenuates self-other integration during joint task performance. *Frontiers in Psychology*, 7, 274.

- Ruys, K. I., & Aarts, H. (2010). When competition merges people's behavior: Interdependency activates shared action representations. *Journal of Experimental Social Psychology*, 46(6), 1130-1133.
- Saby, J. N., Bouquet, C. A., & Marshall, P. J. (2014). Young children co-represent a partner's task: evidence for a joint Simon effect in five-year-olds. *Cognitive Development*, 32, 38-45.
- Saby, J. N., Marshall, P. J., & Meltzoff, A. N. (2012). Neural correlates of being imitated: An EEG study in preverbal infants. *Social Neuroscience*, 7(6), 650-661.
- Sacheli, L. M., Arcangeli, E., & Paulesu, E. (2018). Evidence for a dyadic motor plan in joint action. *Scientific Reports*, 8(1), 1-12.
- Sacheli, L. M., Tidoni, E., Pavone, E. F., Aglioti, S. M., & Candidi, M. (2013). Kinematics fingerprints of leader and follower role-taking during cooperative joint actions. *Experimental Brain Research*, 226(4), 473-486.
- Sakiyama, T. (2017). Ant Droplet Dynamics Evolve via Individual Decision-Making. *Scientific Reports*, 7(1), 1-8.
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1255.
- Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing?. *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 929.
- Sartori, L., Becchio, C., & Castiello, U. (2011). Cues to intention: the role of movement information. *Cognition*, 119(2), 242-252.
- Sartori, L., Becchio, C., Bara, B. G., & Castiello, U. (2009). Does the intention to communicate affect action kinematics?. *Consciousness and Cognition*, 18(3), 766-772.
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". *Neuroimage*, 19(4), 1835-1842.
- Saylor, M. M., & Ganea, P. (2007). Infants interpret ambiguous requests for absent objects. *Developmental Psychology*, 43(3), 696.

- Schmidt, M. F., & Tomasello, M. (2016). How chimpanzees cooperate: If dominance is artificially constrained. *Proceedings of the National Academy of Sciences*, 113(44), E6728-E6729.
- Schmidt, R. C., Carello, C., & Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of Experimental Psychology: Human Perception and Performance*, 16(2), 227.
- Scholl, B. J., & Leslie, A. M. (1999). Modularity, development and 'theory of mind'. *Mind & Language*, 14(1), 131-153.
- Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology*, 43(5), 1124.
- Schurz, M., Aichhorn, M., Martin, A., & Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. *Frontiers in Human Neuroscience*, 7, 712.
- Schuerk, T., Priewasser, B., Sodian, B., & Perner, J. (2018). The robustness and generalizability of findings on spontaneous false belief sensitivity: a replication attempt. *Royal Society Open Science*, 5(5), 172273.
- Scott, R. M. (2017). The developmental origins of false-belief understanding. *Current Directions in Psychological Science*, 26(1), 68-74.
- Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development*, 80(4), 1172-1196.
- Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences*, 21(4), 237-249.
- Scott, R. M., Baillargeon, R., Song, H. J., & Leslie, A. M. (2010). Attributing false beliefs about non-obvious properties at 18 months. *Cognitive Psychology*, 61(4), 366-395.
- Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*. New York, Cambridge University Press.
- Searle, J. (2010). *Making the Social World: The Structure of Human Civilization*. New York: Oxford University Press.

- Searle, John, 1990, Collective Intentions and Actions. In P., Cohen, J., Morgan, & M.E., Pollack (Eds.), *Intentions in Communication*, (pp. 401). Cambridge, Mass.: Bradford Books, MIT Press.
- Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, 1(2), 353-367.
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2), 70-76.
- Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: just like one's own?. *Cognition*, 88(3), B11-B21.
- Sebanz, N., Knoblich, G., Prinz, W., & Wascher, E. (2006). Twin peaks: An ERP study of action planning and control in coacting individuals. *Journal of Cognitive Neuroscience*, 18(5), 859-870.
- Sebanz, N., Rebbechi, D., Knoblich, G., Prinz, W., and Frith, C. D. (2007). It is really my turn? An event-related fMRI study of task sharing. *Social Neuroscience*, 2, 81-95.
- Senju, A. and Csibra, G. (2008) Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology*, 18, 668-671.
- Setoh, P., Scott, R. M., & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proceedings of the National Academy of Sciences*, 113(47), 13360-13365.
- Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology*, 81(1), 174.
- Simon, J. R., & Rudell, A. P. (1967). Auditory SR compatibility: the effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, 51(3), 300.
- Simon, J. R., & Wolf, J. D. (1963). Choice reaction time as a function of angular stimulus-response correspondence and age. *Ergonomics*, 6(1), 99-105.
- Sinigaglia, C., & Butterfill, S. (2015). Motor representation in goal ascription. Conceptual and Interactive Embodiment. *Foundations of Embodied Cognition*, 2, 149-164.

- 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.
- Skerry, A. E., Carey, S. E., & Spelke, E. S. (2013). First-person action experience reveals sensitivity to action efficiency in prereaching infants. *Proceedings of the National Academy of Sciences*, 110(46), 18728-18733.
- Sodian, B., Thoermer, C., & Metz, U. (2007). Now I see it but you don't: 14-month-olds can represent another person's visual perspective. *Developmental Science*, 10(2), 199-204.
- Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, 96(1), B1-B11.
- Song, H. J., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology*, 44(6), 1789.
- Southgate, V. (2013). Does infant behaviour provide support for the mirror neuron theory of action understanding. *Consciousness and Cognition*, 22(3), 1114-1121.
- Southgate, V. (2020). Are infants altercentric? The other and the self in early social cognition. *Psychological Review*, 127(4), 505–523.
- Southgate, V., & Begus, K. (2013). Motor activation during the prediction of nonexecutable actions in infants. *Psychological Science*, 24(6), 828-835.
- Southgate, V., & Verneti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition*, 130(1), 1-10.
- Southgate, V., Johnson, M. H., & Csibra, G. (2008). Infants attribute goals even to biomechanically impossible actions. *Cognition*, 107(3), 1059-1069.
- Southgate, V., Johnson, M. H., Karoui, I. E., & Csibra, G. (2010). Motor system activation reveals infants' on-line prediction of others' goals. *Psychological Science*, 21(3), 355-359.
- Southgate, V., Johnson, M. H., Osborne, T., & Csibra, G. (2009). Predictive motor activation during action observation in human infants. *Biology Letters*, 5(6), 769-772.
- Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18(7), 587-592.

- Stander, P. E. (1992). Cooperative hunting in lions: the role of the individual. *Behavioral Ecology and Sociobiology*, 29(6), 445-454.
- Steinbeis, N. (2016). The role of self–other distinction in understanding others' mental and emotional states: neurocognitive mechanisms in children and adults. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1686), 20150074.
- Stenzel, A., & Liepelt, R. (2016). Joint Simon effects for non-human co-actors. *Attention, Perception, & Psychophysics*, 78(1), 143-158.
- Stenzel, A., Chinellato, E., Bou, M. A. T., del Pobil, Á. P., Lappe, M., & Liepelt, R. (2012). When humanoid robots become human-like interaction partners: corepresentation of robotic actions. *Journal of Experimental Psychology: Human Perception and Performance*, 38(5), 1073.
- Stenzel, A., Dolk, T., Colzato, L. S., Sellaro, R., Hommel, B., & Liepelt, R. (2014). The joint Simon effect depends on perceived agency, but not intentionality, of the alternative action. *Frontiers in Human Neuroscience*, 8, 595.
- Suchak, M., Eppley, T. M., Campbell, M. W., Feldman, R. A., Quarles, L. F., & de Waal, F. B. (2016). How chimpanzees cooperate in a competitive world. *Proceedings of the National Academy of Sciences*, 113(36), 10215-10220.
- Surian, L., & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *British Journal of Developmental Psychology*, 30(1), 30-44.
- Surtees, A. D., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology*, 30(1), 75-86.
- Surtees, A., Apperly, I., & Samson, D. (2016a). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, 150, 43-52.
- Surtees, A., Samson, D., & Apperly, I. (2016b). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, 148, 97-105.
- Szameitat, A. J., Shen, S., & Sterr, A. (2007). Motor imagery of complex everyday movements. An fMRI study. *Neuroimage*, 34(2), 702-713.

- Takagi, A., Beckers, N., & Burdet, E. (2016). Motion plan changes predictably in dyadic reaching. *PloS One*, *11*(12), e0167314.
- Takagi, A., Usai, F., Ganesh, G., Sanguineti, V., & Burdet, E. (2018). Haptic communication between humans is tuned by the hard or soft mechanics of interaction. *PLoS Computational Biology*, *14*(3), e1005971.
- Téglás, E., Giroto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences*, *104*(48), 19156-19159.
- Tollefsen, D. (2004). Collective intentionality. In *Internet Encyclopedia of Philosophy*. <http://www.iep.utm.edu/coll-int/>.
- Tollefsen, D. (2005). Let's pretend! Children and joint action. *Philosophy of the Social Sciences*, *35*(1), 75-97.
- Tollefsen, D., & Dale, R. (2012). Naturalizing joint action: A process-based approach. *Philosophical Psychology*, *25*(3), 385-407.
- Tomasello, M. (1999). The human adaptation for culture. *Annual Review of Anthropology*, *28*(1), 509-529.
- Tomasello, M. (2000). Culture and cognitive development. *Current Directions in Psychological Science*, *9*(2), 37-40.
- Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, *10*(1), 121-125.
- Tomasello, M., & Rakoczy, H. (2003). What makes human cognition unique? From individual to shared to collective intentionality. *Mind & Language*, *18*(2), 121-147.
- Tomasello, M., (2009). *The Origins of Human Communication*, Cambridge Mass., MIT Press.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675-691.
- Tomasello, M., Moore, C., & Dunham, P. J. (1995). Joint Attention as Social Cognition. *Joint attention: Its origins and role in development*, (pp. 103-130). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Träuble, B., Marinović, V., & Pauen, S. (2010). Early theory of mind competencies: Do infants understand others' beliefs?. *Infancy*, 15(4), 434-444.
- Trevarthen, C. (1980). The foundations of intersubjectivity: Development of interpersonal and cooperative understanding in infants. In C., Trevarthen, D., R., Olson, J., Bruner, & G., A., Miller, (1980). *The social foundations of language and thought. Essays in Honor of Jerome Bruner*, (pp.316-342). New-York: Norton.
- Trevarthen, C., & Hubley, P. (1978). Secondary Intersubjectivity: confidence, confinding and acts of meaning in ten first years. In A., Lock (ed.). *Action, Gesture and Symbol: The Emergence of Language*. London: Academic Press.
- Tsai, C. C., & Brass, M. (2007). Does the human motor system simulate Pinocchio's actions? Coacting with a human hand versus a wooden hand in a dyadic interaction. *Psychological Science*, 18(12), 1058-1062.
- Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica*, 116(2), 185-203.
- Tuomela, R. (2002). *The Philosophy of Social Practices: A Collective Acceptance View*. Cambridge: Cambridge University Press.
- Tuomela, R. (2005). We-Intentions revisited, *Philosophical Studies*, 125: 327–69.
- Tuomela, R. (2007). *The Philosophy of Sociality: The Shared Point of View*. New York: Oxford University Press.
- Tuomela, R. (2013). *Social ontology: Collective intentionality and group agents*. Oxford University Press.
- Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, 110(1), 124-129.
- Umiltà, M. A., Intskirveli, I., Grammont, F., Rochat, M., Caruana, F., Jezzini, A., ... & Rizzolatti, G. (2008). When pliers become fingers in the monkey motor system. *Proceedings of the National Academy of Sciences*, 105(6), 2209-2213.
- Vaish, A., Carpenter, M., & Tomasello, M. (2010). Young children selectively avoid helping people with harmful intentions. *Child Development*, 81(6), 1661-1669.

- Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: the negativity bias in social-emotional development. *Psychological Bulletin*, 134(3), 383.
- van der Wel, R. P., Knoblich, G., & Sebanz, N. (2011). Let the force be with us: dyads exploit haptic coupling for coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 37(5), 1420.
- van Ulzen, N. R., Lamoth, C. J., Daffertshofer, A., Semin, G. R., & Beek, P. J. (2008). Characteristics of instructed and uninstructed interpersonal coordination while walking side-by-side. *Neuroscience Letters*, 432(2), 88-93.
- Vanderbilt, K. E., Liu, D., & Heyman, G. D. (2011). The development of distrust. *Child Development*, 82(5), 1372-1380.
- VanderBorgh, M., & Jaswal, V. K. (2009). Who knows best? Preschoolers sometimes prefer child informants over adult informants. *Infant and Child Development: An International Journal of Research and Practice*, 18(1), 61-71.
- Vanderwert, R. E., Fox, N. A., & Ferrari, P. F. (2013). The mirror mechanism and mu rhythm in social development. *Neuroscience Letters*, 540, 15-20.
- Venezia, M., Messinger, D. S., Thorp, D., & Mundy, P. (2004). The development of anticipatory smiling. *Infancy*, 6(3), 397-406.
- Verschoor, S. A., Spapé, M., Biro, S., & Hommel, B. (2013). From outcome prediction to action selection: developmental change in the role of action-effect bindings. *Developmental Science*, 16(6), 801-814.
- Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., Hristova, D., Karlinsky, A., McEllin, L., Nijssen, S. R. R., Schmitz, L., & Wahn, B. (2016). Joint action: mental representations, shared information and general mechanisms for coordinating with others. *Frontiers in Psychology*, 7.
- Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, 23(8-9), 998-1003.
- Vesper, C., Schmitz, L., Safra, L., Sebanz, N., & Knoblich, G. (2016). The role of shared visual information for joint action coordination. *Cognition*, 153, 118-123.

- Vesper, C., Soutschek, A., & Schubö, A. (2009). Motion coordination affects movement parameters in a joint pick-and-place task. *The Quarterly Journal of Experimental Psychology*, 62(12), 2418-2432.
- Vesper, C., van der Wel, R. P., Knoblich, G., & Sebanz, N. (2011). Making oneself predictable: reduced temporal variability facilitates joint action coordination. *Experimental Brain Research*, 211(3-4), 517-530.
- Vesper, C., van der Wel, R. P., Knoblich, G., & Sebanz, N. (2013). Are you ready to jump? Predictive mechanisms in interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 48.
- Vlainic, E., Liepelt, R., Colzato, L. S., Prinz, W., & Hommel, B. (2010). The virtual co-actor: the social Simon effect does not rely on online feedback from the other. *Frontiers in Psychology*, 1, 208.
- Wagner, L., & Carey, S. (2005). 12-month-old infants represent probable endings of motion events. *Infancy*, 7(1), 73-83.
- Wang, B., Hadi, N. S. A., & Low, J. (2015). Limits on efficient human mindreading: Convergence across Chinese adults and Semai children. *British Journal of Psychology*, 106(4), 724-740.
- Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy*, 11(3), 271-294.
- Warneken, F., & Tomasello, M. (2013). Parental presence and encouragement do not influence helping in young children. *Infancy*, 18(3), 345-368.
- Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Development*, 77(3), 640-663.
- 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.
- Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, 75(2), 523-541.

- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, 72(3), 655–684.
- Welsh, T. N. (2009). When  $1+1=1$ : The unification of independent actors revealed through joint Simon effects in crossed and uncrossed effector conditions. *Human Movement Science*, 28(6), 726-737.
- Welsh, T. N., Kiernan, D., Neyedli, H. F., Ray, M., Pratt, J., Potruff, A., & Weeks, D. J. (2013). Joint Simon effects in extrapersonal space. *Journal of Motor Behavior*, 45(1), 1-5.
- Wen, T., & Hsieh, S. (2015). Neuroimaging of the joint Simon effect with believed biological and non-biological co-actors. *Frontiers in Human Neuroscience*, 9, 483.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103-128.
- Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, 1(6), 209-216.
- Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience*, 3(11), 1212-1217.
- Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1431), 593-602.
- Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, 269(5232), 1880-1882.
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69(1), 1-34.
- Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior and Development*, 22(2), 145-160.
- Woodward, A. L. (2005). The infant origins of intentional understanding. *Advances in Child Development and Behaviour*, 33, 229-262.
- Woodward, A. L. (2009). Infants' grasp of others' intentions. *Current Directions in Psychological Science*, 18(1), 53-57.

- Woodward, A., & Needham, A. (Eds.). (2008). *Learning and the infant mind*. Oxford University Press.
- Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, 105(13), 5012-5015.
- Yoo, K. H., Cannon, E. N., Thorpe, S. G., & Fox, N. A. (2016). Desynchronization in EEG during perception of means-end actions and relations with infants' grasping skill. *British Journal of Developmental Psychology*, 34(1), 24-37.
- Zani, G., Butterfill, S. A., & Low, J. (2020). Mindreading in the balance: adults' mediolateral leaning and anticipatory looking foretell others' action preparation in a false-belief interactive task. *Royal Society Open Science*, 7(1), 191167.
- Zwaan, R. A., Van Der Stoep, N., Guadalupe, T., & Bouwmeester, S. (2012). Language comprehension in the balance: the robustness of the action-compatibility effect (ACE). *PloS One*, 7(2).



## RÉSUMÉ

Bon nombre d'importantes réalisations de notre espèce sont le résultat d'actions conjointes. L'action conjointe implique que les co-agents forment des représentations partagées de leur environnement, de leurs objectifs et de leurs actions ; ainsi, ils intègrent des représentations de soi et des autres dans une représentation unitaire. Ma thèse porte sur le développement des capacités d'action conjointe et de représentations partagées dans l'ontogénie humaine.

En philosophie, les maximalistes défendent que l'action conjointe nécessite que les co-agents partagent leurs intentions (grâce à une théorie de l'esprit) alors que selon les minimalistes elle nécessite seulement que les individus partagent des buts. Je défends une position minimaliste considérant que la variété des actions conjointes conduit à différentes demandes de représentation et de partage. En psychologie du développement, l'émergence d'une théorie de l'esprit chez le jeune enfant est encore débattue. J'examine les implications de ce débat pour la capacité des jeunes enfants à s'engager dans une action conjointe et conclus qu'ils peuvent répondre à des demandes propres aux actions conjointes minimales grâce à des représentations de buts et des capacités de prise de perspective.

Enfin, je propose une matrice d'action conjointe combinant : (1) le type de représentation impliquée dans l'action conjointe en fonction de sa complexité et (2) le niveau de partage qu'impliquent de telles représentations en fonction de la stratégie de maximisation de performance mise en place par les co-agents. Les formes d'action conjointe observées à un âge donné sont ainsi fonction de la complexité de la tâche et de la stratégie qu'elle met en jeu et dépendent du degré de développement à cet âge de compétences cognitives telles que conscience de soi, prise de perspective et théorie de l'esprit.

## MOTS CLÉS

Développement de l'action conjointe, Capacités de prise de perspective, Théorie de l'esprit, Maximalisme, Minimalisme

## ABSTRACT

Many significant achievements of our species are the result of joint actions. Joint action requires co-agents to form shared representations of their environment, their goals and actions; to do so, they integrate self and other representations within a unitary representation. I propose to investigate how a capacity for joint action and shared representations develops in human ontogeny.

In philosophy, maximalists argue that joint action requires co-agents to share intentions (using theory of mind capacities) when minimalists contend that it requires only that individuals share goals (understanding others' behaviors as goal-directed). I defend a minimalist account considering that different forms of joint actions make different representational and sharing demands. In developmental psychology, early theory of mind abilities remain debated. I discuss the implications of this debate for young children's capacities to engage in joint action, concluding that young children may meet representational and sharing demands of minimal joint actions using goal representations and visual perspective-taking capacities (instead of intentions and theory of mind capacities).

I offer a joint action matrix combining: (1) the type of representation involved in joint action based on its complexity and (2) the level of sharedness implied by such representations based on the co-agents' strategy to maximize the joint performance. I argue that, before young children fully develop the mindreading and perspective-taking abilities required for more complex joint actions, they may already successfully engage in minimal forms of joint action – imposing less representational and sharing demands.

## KEYWORDS

Joint action Development, Perspective-taking abilities, Theory of Mind, Maximalism, Minimalism