ATTENDING AND KNOWING TOGETHER
A NEW LOOK AT JOINT ATTENTION AND COMMON KNOWLEDGE AND THEIR ROLE IN COORDINATION

Barbora Siposova

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2018

Attending and knowing together

A new look at joint attention and common knowledge and their role in coordination

Barbora Siposova

University of St Andrews

This thesis is submitted in partial fulfilment for the degree of

Doctor of Philosophy (PhD)

at the University of St Andrews

July 2018

Candidate's declaration

I, Barbora Siposova, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 33,300 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2014.

I, Barbora Siposova, received assistance in the writing of this thesis in respect of grammar, which was provided by Christina Lowe.

I confirm that no funding was received for this work.

Date 10/7/18 Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date 10/7/18 Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Barbora Siposova, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date 10/7/18 Signature of candidate

Date 10/7/18 Signature of supervisor

Underpinning Research Data

Candidate's declaration

I, Barbora Siposova, understand that by declaring that I have original research data or digital outputs, I should make every effort in meeting the University's and research funders' requirements on the deposit and sharing of research data or research digital outputs.

Date 10/7/18 Signature of candidate

Permission for publication of underpinning research data or digital outputs

We understand that for any original research data or digital outputs which are deposited, we are giving permission for them to be made available for use in accordance with the requirements of the University and research funders, for the time being in force.

We also understand that the title and the description will be published, and that the underpinning research data or digital outputs will be electronically accessible for use in accordance with the license specified at the point of deposit, unless exempt by award of an embargo as requested below.

The following is an agreed request by candidate and supervisor regarding the publication of underpinning research data or digital outputs:

No embargo on underpinning research data or digital outputs.

Date 10/7/18 Signature of candidate

Date 10/7/18 Signature of supervisor

Abstract

Joint attention—the ability to coordinate attention to an object or event—is a key feature of human social cognition. Without joint attention, we would not be the cooperating species we are now. Most developmental studies focus on when joint attention emerges, rather than what joint attention actually is, and what consequences it has for different types of interactions. To address this gap, Part I of the thesis provides a new look at joint attention, presenting a systematic framework of four attention levels (from monitoring, to common, mutual, and shared attention) with corresponding distinctions for the knowledge states associated with each level. Cognitive, behavioral, and phenomenological aspects of the different levels are discussed, as well as the functions and consequences the levels have in terms of what kinds of obligations they can support.

Part II of the thesis follows on from this in investigating the role of joint attention in facilitating coordination. In two studies, 5- to 7-year-old children played a Stag Hunt coordination game in which they needed to decide whether to cooperate or play individually. During the decision-making phase, the children's partner either shared attention with them—she made ostensive, communicative eye contact—or looked non-communicatively at them. In Study 1, results showed that communicative looks produced an expectation of cooperation in children. In Study 2, children normatively protested when their partner did not cooperate, thus showing an understanding of the communicative look as a commitment to cooperate. This is the first experimental evidence, in adults or children, that in the right context, communicative, but not non-communicative, looks can signal a commitment. Thus the thesis highlights the special importance of the shared attention level in supporting cooperation.

General acknowledgements

This thesis was only possible because I was surrounded and supported by many inspirational people. I would like to express my sincere gratitude to all of them.

First, I would like to express my gratitude to my supervisors, Malinda Carpenter and Michael Tomasello, for providing their generous support, guiding me wisely through my PhD studies, and sharing their ideas and knowledge with me over the years. I am grateful to Malinda Carpenter for continuous encouragement, insightful feedback on all my writings, and countless exciting discussions. I am grateful to Michael Tomasello for making the developmental and comparative psychology department at the MPI an incredibly cooperative place. I am honored that I could be part of it. I would also like to thank Amanda Seed for her guidance through studies at St Andrews University.

A special thank you goes to all my child lab colleagues, especially to Susanne Mauritz, Elena Rossi, and Katharina Haberl for their invaluable help with the experimental studies. I'd also like to thank Theo Toppe, Anna-Marie Seidel, Nathalie Dosch, and Rahel Doehring for help with data collection; Christina Juengst, Silvan Hornstein, and Felix Wenger for help with coding; Marike Schreiber for her great study illustrations, Roger Mundry for his expert statistical advice, Petra Jahn for her reliable IT support, and Christina Lowe for her help and patience with my English grammar skills. Importantly, a very special thank you goes to the children for their participation. I would like to thank to Nathan Caruana, Sebastian Grueneisen, John Michael, Maria Ploetner, Gideon Salter, Theo Toppe, Wouter Wolf, Dan Zahavi, and especially Cathal O'Madagain and Margherita Isella for their constructive comments on the ideas presented in the thesis.

My thanks go to my dear colleagues and friends – our discussions and time spent together provided me with many insights and made my PhD studies a highly rewarding experience: Susanne, Patricia, Becky, Diotima, Sophie, Sebastian, Jan, Chris, Gregor, Ulrike, Avi, Jeremy, Lialin, Julia, Maria and Andreas. I also gratefully acknowledge the Minerva group members—Maria, Lars, Antonia, Harriet, and Katharina—and the Social Cognition Meeting group members for their useful feedback. Thank you to my dear friends: Anne, Maria, and Vivien, for making Leipzig my home; and Paula, Marcela, Marina, Kristina, and Ada for their care and warm-hearted friendship.

Last but not the least, I would like to thank to my brother and my parents who always believe in me.

To Marek for his endless encouragement, lighthearted approach to life, patience, and love.

Funding

This work was supported by the Max Planck Institute for Evolutionary Anthropology.

Research Data

Research data underpinning this thesis are available at https://doi.org/10.7910/DVN/Y3LADG, UNF:6:nZf4dwUPENyp2zi/Rt/rlQ==

Preface

"Momentarily he caught O'Brien's eye. O'Brien had stood up. He had taken off his spectacles and was in the act of resettling them on his nose with his characteristic gesture. But there was a fraction of a second when their eyes met, and for as long as it took to happen Winston knew—yes, he KNEW!—that O'Brien was thinking the same thing as himself. An unmistakable message had passed. It was as though their two minds had opened and the thoughts were flowing from one into the other through their eyes. 'I am with you,' O'Brien seemed to be saying to him. 'I know precisely what you are feeling. I know all about your contempt, your hatred, your disgust. But don't worry, I am on your side!' And then the flash of intelligence was gone, and O'Brien's face was as inscrutable as everybody else's."
- George Orwell, *1984, p. 22*

"Only connect!"
- E.M. Forster, *Howards End, p.134*

1. General introduction

The ability to experience the world together with others is a key feature of human life. It allows us to shift from solitary, private, and individual existence to shared, communal, and interconnected existence. We engage with each other in a large variety of ways: we share our perceptions, emotions, and thoughts. When we grasp, attune to, influence and share mental states, we engage in shared intentionality; we can experience the commonality of our mental states about something with others and focus on things together (Bratman, 1999; Searle, 1995; Tomasello & Carpenter, 2007; Tuomela, 1995). We could not have conversations, collaborate with co-workers, teach others, cook with friends, or play with children without shared intentionality. It enables us to successfully coordinate our actions with others in diverse settings. Shared intentionality allows us to coordinate concrete actions in small-scale activities, to coordinate our decisions and even to coordinate on a more abstract level and towards goals that we could never achieve alone. Humans have also succeeded in many large-scale projects which have involved extensive coordination: they have travelled to the moon, created the internet and mapped the human genome. The capacity for shared intentionality and coordination in humans exceeds that of other species (Tomasello, Carpenter, Call, Behne, & Moll, 2005) to such an extent that humans have received the label of being a hypersocial and hypercooperative species (Richerson & Boyd, 2004). The central question is, what are the cognitive processes supporting this capacity? Which cognitive mechanisms facilitate the "being and acting together" often referred to as "We mode" (also called collective intentionality) (e.g., Gilbert, 1989; Searle, 1995; Tollefsen, 2002)? Although no single cognitive skill could be named as an answer to these questions, joint attention—the ability to coordinate attention to an object or event of mutual interest— plays a crucial role (Tomasello & Carpenter, 2007). Therefore, this thesis focusses on joint attention, without which we would not be the social, sharing and cooperating species we are now.

Empirical research on joint attention started in the 1970s in the field of developmental psychology. Bruner and his colleagues published descriptions of infants' ability to follow, direct, and share attention with their caregivers (e.g., Bruner, 1974; Scaife & Bruner, 1975). Since then, a large body of research has demonstrated the importance of joint attention for social and communicative

development. Infants begin to actively share attention with others from around 9 to 12 months of age: they engage in pointing, showing and gaze alternation about objects with their caregivers (Carpenter, Nagell, & Tomasello, 1998). It has been shown that joint attention scaffolds the learning of new words (Tomasello & Farrar, 1986) and grammar (Rollins & Snow, 1998). Infants' joint attention skills predict their later mental state language (Kristen, Sodian, Thoermer, & Perst, 2011), false-belief understanding (Sodian & Kristen-Antonow, 2015) and social competence (VanHecke et al., 2007) at preschool age. The foundational role of joint attention skills in infants' development is especially visible when these skills are lacking or impoverished, as in children with autism (Sigman, Mundy, Sherman, & Ungerer, 1986).

Over the years, interest in joint attention has branched out far beyond developmental psychology; joint attention has been studied in comparative psychology, social robotics, philosophy, and education sciences. Several recent studies have investigated joint attention in adults and focus on consequences of joint attention. It has been shown that when adult social partners simultaneously attend to the same object or event, their experience of the co-attended thing changes. For example, social partners pay more attention to the things in their environment when they believe that others attend to those things as well (Shteynberg & Apfelbaum, 2013; Shteynberg et al., 2014). Similarly, their experiences become intensified when they attend to the same thing with someone else (Boothby, Clark, & Bargh, 2014; Boothby, Smith, Clark, & Bargh, 2016, 2017). These effects are especially evident when social partners are psychologically close to each other (Boothby et al., 2016, 2017; Shteynberg et al., 2014). Interestingly, another study has reported that joint attention affects social closeness: attending to the same object as opposed to a different object results in increased perceived closeness between adult participants (Wolf, Launay, & Dunbar, 2015). Thus, although more research is needed, these findings suggest that joint attention situations lead to increased cognitive processing and strengthen social bonding between social partners.

Attending to the world around us jointly with others is rewarding, but it brings important advantages for us as well. In social interactions, joint attention not only allows us to coordinate our attention to things in the environment that can be directly perceived, but also focus together on mental states such as thoughts (Mundy, Sullivan, & Mastergeorge, 2009; O'Madagain & Tomasello, submitted). As a joint

focus enables us to coordinate our thoughts, behaviors and different roles when acting towards a common goal, it has been suggested that joint attention is crucial for our cooperation with others (Brinck & Gardenfors, 2003; Tollefsen, 2005; Tomasello, 1999). Surprisingly, however, few empirical studies have directly investigated the precise role of joint attention in facilitating successful coordination and cooperation.

One reason is that although there is broad consensus about joint attention being one of the most important skills in social cognition, different scholars focus on different aspects of it, so there is still little agreement about what exactly joint attention is (Carpenter & Liebal, 2011). In the next section I first briefly distinguish between two main approaches in studying joint attention: one which focuses on behaviors and another which focuses on the knowledge that social partners engaged in joint attention have. Then I turn to the question of how social partners achieve jointness in joint attention. This discussion will show that further empirical research would benefit greatly from an organized framework of different joint attention levels.

1.1 The need for an organized framework

Although it seems that almost every scholar who has been studying joint attention has his or her own unique way of defining it (see Part I for details), it is possible to categorize the approaches to joint attention broadly into two groups. Scholars in the first group have tended to focus on the behavioral aspects of coordinating attention (e.g., Bakeman & Adamson, 1984; Baron-Cohen, 1989; Butterworth, 1995; Camaioni, 1993; Leavens & Racine, 2009; Mundy & Newell, 2007). Scholars in the second group have highlighted what kind of knowledge partners have about each other when they engage in joint attention. They have argued that to be in joint attention, partners must know together that they are attending to the same thing, that is, they have to have common knowledge of their mutual attending (e.g., Carpenter & Liebal, 2011; Hobson, 2005; Tomasello, 1995). Following on from this, a related question is how partners reach this mutual understanding. Again, the answers vary greatly and involve a focus either on some type of recursive mind-reading (e.g., Bach & Hamish, 1979; Harder & Kock, 1976; Tomasello, 2008, 2011), intersubjective engagement (Gómez, 2005; Hobson, 2005; Eilan, submitted; Reddy, 2011), or intentional communication (Carpenter & Liebal, 2011).

This lack of agreement is visible in the terminology used in this area too, as several other terms are used in the literature as synonyms for the term joint attention or describe phenomena that have significant overlap with joint attention, for example co-attention, shared attention, shared experiences, and joint engagement. This lack of agreement in both areas hinders progress in further understanding joint attention (Carpenter & Liebal, 2011). First, terminological confusion complicates dialogue among different disciplines. Second, different definitions lead to using different ways of measuring joint attention in experiments, which makes comparison of results across experiments difficult. For example, most developmental studies with infants operationalize joint attention in terms of initiating or responding to attention sharing communicative behaviors such as pointing, showing, and gaze alternation. In contrast, studies in social psychology tend to focus on situations in which adult participants co-attend to the same thing, and know they are co-attending, but do not engage in any communicative behaviors towards each other. For instance, a situation in which participants were simply looking at the same side of a computer screen while performing a reaction time task has been called joint attention (Wolf et al., 2015).

Here I will argue that joint attention is a different process when partners interact with each other (as in typical infant studies) and when partners co-attend without any interaction (as in typical adult studies). That is, recently, much attention has been given to the idea that our social-cognitive processes differ significantly when we observe others (adopting a third-person perspective) compared to when we are in direct engagement with others (adopting a second-person perspective) (e.g., DeJaegher, DiPaolo, & Gallagher, 2010; Heal, 2005; Reddy, 1996; Schilbach, 2010; Zahavi, 2008). In the same vein, several scholars have argued that a second-person relation and engagement are crucial for joint attention (Hobson, 2005; Gómez, 2005; Eilan, submitted; Reddy, 2011). Overall, this body of work directly relates to the question of how social partners achieve jointness and common knowledge in joint attention. At this point in our understanding of this topic, it thus would be very useful to flesh out these ideas further in a way that integrates the question of what joint attention is with the question of how the common knowledge in it is achieved. Therefore, the purpose of this thesis is to take a new look at joint attention and common knowledge. The first part of the thesis offers a new, organized framework of different joint attention levels with corresponding levels of knowledge states. The

framework emphasizes a distinction between a second-person, engaged perspective and a third-person, observational perspective and defines four joint attention levels (monitoring, common, mutual and shared attention).

All of the different levels of joint attention are useful for interpreting and guiding our behavior in social contexts. Correctly identifying where others are focusing their attention in the environment facilitates an understanding of their attitudes, intentions, emotions, and beliefs. However, the proposed attention levels fulfil distinct functions and facilitate different kinds of social interactions. For example, one level—shared attention—supports coordination much better than all the other levels for several reasons.

First, as I will explain in more detail in Part I, in shared attention partners are engaged in a second-person relation to each other and actively share attention by intentionally communicating about the object of their attention. Communicative exchange allows social partners to coordinate their mental states about the object of attention. This leads to more effective coordination compared to a situation in which partners know that they attend to the same object, but have less information about their attitudes, goals, or intentions towards that object. Second, in shared attention partners generate a specific type of common knowledge: shared knowledge. This is relevant for successful coordination because shared knowledge increases partners' certainty about their attention focus (Carpenter & Liebal, 2011). Third, communicative exchange creates a public space in which commitments can come into existence (Gilbert, 2007). Commitments stabilize successful coordination between partners (Michael & Pacherie, 2015). Under certain circumstances, attention sharing can become normatively binding and give rise to commitments. To summarize, attention sharing supports coordination in several ways. It allows partners to align their mental states about the co-attended object and therefore create stronger expectations about their partner's future behaviors. Attention sharing is also public, which allows partners to create commitments that facilitate successful coordination. Therefore, the rest of the introduction examines the role of attention sharing in coordination.

1.2 Attention sharing as a coordination tool

The goal of this section is to discuss how attention sharing facilitates successful coordination between partners. First, I focus on the context in which

attention sharing could have evolved. Then I briefly distinguish between two different types of coordination: fine-grained coordination of actions in time and space and strategic coordination. Next, I closely examine strategic coordination, explain the role of strategic coordination in coordination dilemmas, and highlight the importance of common knowledge as a crucial way of solving these dilemmas. Following that, I introduce coordination games as experimental methods for studying coordination dilemmas. There I discuss available experimental findings showing how adults and young children use attention sharing to solve the dilemmas. That section points to the need for further experimental studies investigating nonverbal ways of coordinating. Finally, in the last section I turn to attention sharing via communicative eye contact and argue that communicative eye contact is a powerful coordination tool, as it can create common knowledge and even possibly a commitment between social partners.

1.2.1 The context for the evolution of attention sharing

To understand the role of attention sharing in coordination, it is useful first to consider the environment in which the ability to share attention might have evolved. Humans are thought to have lived in small groups for a significant amount of time in their evolutionary history (Brewer, 2007). Although the exact time period and the specific reasons why attention sharing evolved are not known, the interdependence hypothesis argues that this skill probably evolved together with other social-cognitive and communicative skills supporting collaboration (Tomasello, Melis, Tennie, Wyman, & Herrmann, 2012). According to this theory, at some point in evolution, changes in the environment caused collaboration between social partners to become necessary for survival. It is argued that collaboration started in the context of collaborative foraging, when social partners had to coordinate with others to obtain food. Once collaboration was needed for survival, group members became interdependent, that is the success of one individual depended on the success of others. In other words, group members were interested in each other's welfare since their partner's welfare had direct beneficial consequences for them (Roberts, 2005). Following this line of argument, this type of mutualistic environment created a strong motivation to be a particularly helpful partner (Tomasello, 2014).

This is the context in which the ability to share attention evolved. When humans started helping each other and working together more regularly, the ability

to share attention to relevant features of the environment became crucial, because it enabled them to reach their goals much more effectively. Successful coordination increased success in foraging activities, and therefore partners with superior skills in coordinating were preferred by others as future partners in collaboration. This pattern resulted in selective pressure for the evolution of cognitive skills that supported coordination, and attention sharing was one of them (Tomasello, 2008). Thus it is probably the case that the ability to share attention evolved within the context of mutualistic collaboration when social partners had to coordinate concrete actions with others in daily activities. It seems plausible that this ability initially evolved in interpersonal interactions within the context of dyadic or small groups in which partners shared attention to concrete, salient things in their environment (for example, to a predator or prey); and only later a more complex version of the skill evolved which enabled humans to share attention within larger groups and to abstract things or even mental states (O'Madagain & Tomasello, submitted).

1.2.2 Two types of coordination

There are at least two different angles from which to look at coordination. The first angle focuses on coordination as action coordination between partners; the second on strategic coordination. Taking the first approach, several scholars have been asking how exactly social partners adjust their body movements to each other when they perform joint action, such as carrying a heavy object, dancing, or rowing a boat together. The main focus is on 1) studying low-level perceptual, cognitive and motor processes involved in temporal and spatial action coordination and 2) investigating how each of the partners forms representations of the other partner's action goals, predicts the other partner's action, and includes this information into their own actions (Sebanz & Knoblich, 2009; Vesper et al., 2017; Vesper, Schmitz, Safra, Sebanz, & Knoblich, 2016). Although attention sharing is not directly investigated in this line of research, it is acknowledged as one of the mechanisms that supports action coordination. Specifically, attention sharing makes action coordination more efficient by creating a shared perceptual space between partners and allowing them to direct their attention to the same things in the environment (Sebanz, Bekkering, & Knoblich, 2006). This claim is supported by evidence from a study in which adult participants' task was to assemble a Lego model together (Clark & Krych, 2004). During the task, one partner verbally instructed the other one about

how to build the model. When they both could see the building space, participants were twice as fast at completing the task compared to a situation in which they could not visually share the building space. Importantly, when they shared the perceptual space, they spontaneously used attention sharing behaviors such as pointing, showing blocks, nodding, and eye contact during the task, which helped them to coordinate.

The second approach to coordination focuses on higher-level cognitive processes that take place before we engage in joint actions with others. Before we act, we consider relevant information for our actions. Often, we need to decide how to act based on our expectations of what others will do, and more than one solution exists. Taking this approach, scholars have been investigating strategic decision-making in "coordination dilemma" situations.

1.2.3 Coordination dilemmas

In coordination dilemmas (sometimes referred to as "coordination problems" or "social dilemmas"), two or more social partners have a shared interest, each partner's outcome depends not only on his or her own decision but also on the decision of the other, and at least two solutions are possible (Lewis, 1969). To illustrate the nature of coordination dilemmas, consider the notable example by Chwe (Chwe, 2001). You and your companion are going to have a drink after work. You enter a bus to go to the pub, but the bus is crowded and you end up being separated from each other. At a bus stop which is before your agreed destination, a colleague of yours sees you and shouts from the sidewalk that both of you should join him there for a drink. You are interested in spending the evening with your companion, but joining another colleague would be an option as well, as long as you both go. Before the bus starts going again, each of you has to make a decision about whether to get off the bus or not. You search for your companion but cannot see her because of the crowd. How do you decide? It is likely in this situation that you would wonder if your companion heard the offer to join the colleague. Suppose that you see her looking at your colleague just for a second, so you know that she heard the suggestion. But does she know that you heard it? Even if she does, does she know that you know that she knows that you know? It is not enough that each of you knows about the knowledge of the other separately. To successfully solve the dilemma, you

have to know together about the colleague's invitation; in other words you need to have common knowledge about it (Clark, 1996; Chwe, 2001; Rubinstein, 1989).

Coordination dilemmas can be modelled and studied under controlled conditions using economic games. Economic games are widely used in game theory and behavioral economics. The most commonly used game is the Prisoner's Dilemma.[1] The Prisoner's Dilemma is a "conflict game": the individual interests of each partner conflict with the mutual benefits. Each partner would be better off if the other one is worse off. In contrast to conflict games, coordination dilemmas are modelled as "coordination games" in which the partners have mutual interests and a common goal they want to reach. In the bus example given above, your and your companion's shared goal is to spend the evening in each other's company. Given the mutualistic context in which human coordination probably have evolved (Tomasello et al., 2012), as well as the structure of many social situations nowadays that humans often encounter, it has been argued that coordination dilemmas, and not conflict dilemmas, should be the main focus of scientific investigation (Skyrms, 2004).

A prominent game used for studying coordination in mutualistic contexts is the Stag Hunt game (Rousseau, 1754; Skyrms, 2004). The structure of the game is illustrated by a hunting parable (Rousseau, 1754; Skyrms, 2004): Two hunters decide either to each hunt a hare (a certain but small prize) individually or to hunt a stag (a risky but big prize) together, if it is available. Hares appear in the forest often, stags in contrast appear only rarely. However, if just one partner decides to cooperate and hunt the stag and the other decides to hunt a hare (for instance, because he does not know the stag is available), the cooperating partner loses the chance to get anything, because she was wasting her effort in hunting the stag (see Figure 1).

[1] The Prisoners Dilemma is illustrated by a story in which two suspects are questioned by the police separately. If one suspect claims that the other committed the crime, and the other remains silent, the first suspect goes free and the other is sentenced to a long term in prison. If both remain silent, they both get a short-term sentence. If both claim the other committed the crime, they both have to stay in prison for a moderate time (Axelrod & Hamilton, 1981).

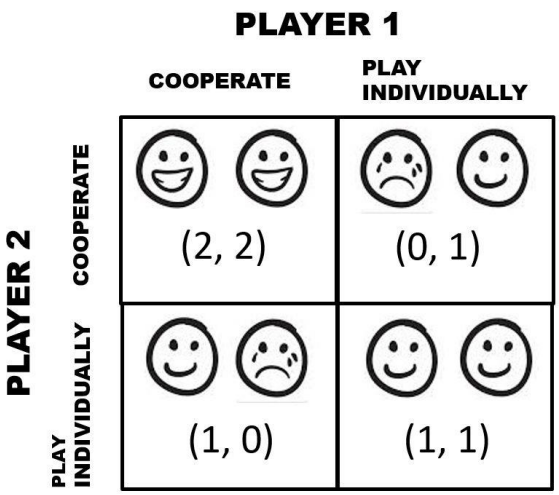

Figure 1. Example of a Stag Hunt game pay-off structure. Number two represents a stag, number one represents a hare.

In a single-trial game, both partners must individually decide how to play and can make only one final decision. Both partners know that only when they both decide to cooperate, they get a high-value prize. There are two solutions in a Stag Hunt dilemma in which both get at least something: Either both partners decide to cooperate for a high-value prize (pay-off dominant solution, but risky) or each partner plays individually for a low-value prize (safe, risk dominant solution, but with a low-value prize). In contrast to the bus example above, in which coordinating getting off the bus vs. staying on it were equally valued, in the Stag Hunt game the two solutions are not equally valued. Coordinating on a collaborative option is better than coordinating on an individual option. However, the underlying structure is the same: Both partners get the most benefit when they coordinate their decisions. Again, the partners can solve this coordination dilemma if they have common knowledge, in this case, common knowledge that the collaborative option has become available (Clark, 1996; Chwe, 2001; Rubinstein, 1989).

To test the prediction that common knowledge is needed for solving the Stag Hunt dilemma, in a recent study adult participants played a modified version of a Stag Hunt game (Thomas, DeScioli, Haque, & Pinker, 2014). Participants were allocated to one of four conditions and presented with written vignettes introducing the dilemma and the payoff structure of the game. Importantly, participants in each condition were given different information about the knowledge of their partner.

Before deciding between an individual or a collaborative option, all participants read that the collaborative option had become available. Participants in the first condition had no information about their partner's knowledge. In the second condition, participants read that their partner knew about the collaborative option as well. In the third condition, participants learned that their partner not only knew about the collaborative option but also that they themselves knew about the collaborative option as well. Finally, in the last condition, participants were told that the information about the collaborative option was publicly announced via a loudspeaker (so everyone knew that everyone knew about it). The results showed that the majority of the participants in the last condition chose the risky collaborative option and their collaboration rates were significantly higher than those in the other conditions. This study thus demonstrated that publicly-broadcasted announcements are powerful cues for common knowledge and that participants recognized common knowledge as a distinct category, which in turn helped them to coordinate in deciding on a collaborative solution. Importantly, these results also provide clear evidence that adding a recursive step to one's knowledge (i.e. you know that your partner knows that you know) does not equal the state of knowing it together.

This study provided an important starting point for empirical investigations into the role of common knowledge in coordination. However, note that as participants were given written instructions, they were provided with the relevant information for making their decisions. Therefore, their decision-making was arguably less challenging than more natural situations in which we need to coordinate our decisions with others daily. Often, we need to extract the relevant information from the changing environment ourselves, consider how it affects our partner's knowledge, and actively create common knowledge with our partners. So far, very few studies have addressed the question of how we produce, understand, and use common knowledge for strategic coordination during interpersonal interactions. As common knowledge is crucial part of attention sharing (e.g., Carpenter & Liebal, 2011; Tomasello, 1995), the next section discusses the available empirical evidence that both adults and children use attention sharing behaviors in strategic coordination.

1.2.4 Attention sharing in coordination dilemmas

To establish attention to the relevant information in coordination dilemmas, for example to share attention to the fact that a collaborative option is now available, the obvious solution is to use verbal communication. With verbal communication speakers can precisely guide the other's attention to specific things or events in the environment (Talmy, 2000) and by verbal exchange partners can create common knowledge between them. A study by Duguid and colleagues (Duguid, Wyman, Bullinger, Herfurth-Majstorovic, & Tomasello, 2014) demonstrated that verbal communication is a common strategy for solving coordination dilemmas even for preschoolers. In their study, pairs of 4-year-old children played a modified version of a Stag Hunt game. As they could not see each other's decisions due to a barrier between them, they were not sure whether their partner saw the collaborative option when it appeared. When children needed to decide whether to choose the risky collaborative option, they spontaneously shared attention to the fact that the collaborative option had become available by verbally communicating about it. Spontaneous use of verbal communication as a way to solve coordination dilemmas has been found with adult participants as well (Brosnan, Wilson, & Beran, 2012).

Another effective way in which partners can share attention when coordinating is through the use of gestural communication. It has been shown that small groups of adults in a social foraging experiment spontaneously used pointing and waving to indicate to their partners where to put their collective foraging efforts (King et al., 2010). Although they used verbal communication as well, it was the use of gestures that best predicted how successful they were in discovering the most profitable foraging location. Overall, participants who were allowed to communicate were more successful in collecting resources than participants in a condition in which communication was discouraged. Although King and colleagues did not discuss common knowledge as a reason for their participants' coordination success, it appears likely that that the common knowledge (about the most profitable location for foraging) that participants created by means of verbal and nonverbal communication resulted in an expectation that all members would put their efforts into that place. This study thus points to the potential importance of nonverbal communication for coordination success and illustrates the fact that attention sharing behaviors support not only dyadic but also group coordination.

It is important to note that for verbal or gestural communication to be a successful coordination tool, it is necessary that the communication is a bi-directional process: Both partners need to know that what one said or gestured the other understood (Clark & Brennan, 1991). Coming back to the bus example above, it would not be enough if in a crowded bus you shout, “Let’s get off at this bus stop” or point toward your colleague outside, as you would not know whether your partner could hear you or see your pointing gesture. Even 1-year-old infants know this and adapt their gestural communication according to their partner’s focus of attention (Liszkowski, Albrecht, Carpenter, & Tomasello, 2008). What is needed is that both of you are mutually aware of your colleague outside. Although existing theoretical accounts do not agree on how exactly mutual understanding between two partners is reached, it has been proposed that eye contact is an especially powerful nonverbal generator of common knowledge, as it provides instantaneous, direct, and bidirectional information about both partners’ focus of attention (Carpenter & Liebal, 2011; Chwe, 2001; Thomas et al., 2014). Although verbal exchange has the advantage that partners can create common knowledge even when they are separated in time and space, eye contact has the advantage over verbal communication (without eye contact) that partners do not need to check whether they are paying attention to each other. When the eyes of social partners meet, both know for certain where their attention is focused (Gómez, 1996). We often use a combination of verbal communication, gestures and eye contact to create common knowledge. However, in situations requiring instantaneous decisions in which verbal communication is either not possible or too time consuming, we might rely on eye contact only.

Eye contact is an intriguing behavior both from an evolutionary and a developmental perspective. Compared to most primates, the human eye region has subtle differences that make it highly salient. First, the sclera of the human eyeball is white; second, there is a contrast in color between human facial skin, the sclera, and the iris; third, human eyes are large in proportion to our body size; and fourth, humans eyebrows are salient (Kobayashi & Kohshima, 1997). It has been suggested that these features evolved as an adaptation to the increased need to coordinate with others using eye contact in collaborative interactions (Tomasello, Hare, Lehmann, & Call, 2007). From the developmental perspective, humans show great interest in the eye region basically from birth, with a special sensitivity to eye signals already

developing in early infancy (Farroni, Csibra, Simion, & Johnson, 2002). Importantly, infants' developing understanding of eye contact and gaze enables them—together with other nonverbal behaviors, such as pointing—to share attention and thus to gain basic competence in establishing common knowledge with their partners (Carpenter, 2009).

Much attention has already been given to eye contact as a coordination tool in face-to-face conversations: Eye contact is used to initiate and regulate the timing of interactions (Argyle, Ingham, Alkema, & McCallin, 1973), to attribute emotions and mental states to each other (Baron-Cohen & Cross, 1992; Baron-Cohen, Wheelwright, & Jolliffe, 1997), and to express intimacy (Argyle & Dean, 1965). Surprisingly, however, little is known about eye contact as a coordination tool in the context of strategic cooperation. To my knowledge, only one previous study has investigated this.

Wyman, Rakoczy, and Tomasello, (2013) studied children's ability to create common knowledge about the availability of the cooperative option using eye contact. In this study, 4-year-old children were discouraged from communicating verbally while playing a modified version of a Stag Hunt game with an experimenter. Crucially, when the collaborative option became available, before children had to choose whether to play individually or collaboratively, the experimenter made one of two different types of eye contact with children. In the control condition, when the collaborative option appeared, the experimenter monitored it visually—it was clear that she saw it—but did not look at children. Thus, in this condition, children themselves knew that the collaborative option was available and they also knew that their partner could see it as well, as she was looking directly at it. In the experimental condition, in contrast, the experimenter alternated gaze between the cooperative option and children's eyes. During this communicative eye contact, she opened her eye widely, raised her eyebrows and smiled at children. Therefore, in this condition, children not only knew that both they themselves and the experimenter saw the collaborative option, but also knew that both they and the experimenter both knew that they both saw it. Results showed that more children decided for the collaborative option in the experimental than in the control condition. This suggests that the communicative eye contact and smile from the experimenter potentially led to common knowledge about the availability of the collaborative option between them, which in turn facilitated their coordination. Note that although this study used a

different version of the Stag Hunt game than Duguid and colleagues (Duguid et al., 2014) and Thomas and colleagues (Thomas et al., 2014), all of these studies found evidence that when participants could gain some form of common knowledge, they overwhelmingly chose the collaborative option. The Wyman et al. (2013) study in addition provided evidence that eye contact can influence coordination decisions. However, due to the manipulation they used (i.e., communicative eye contact with a smile compared to no eye contact) it is unknown whether any type of eye contact can promote the common knowledge which makes coordination possible or whether the look needs to have specific characteristics, e.g., whether it needs to be an ostensive-communicative look.

It has been argued that to attend to and know something together with someone, it is necessary that partners share their attention through intentional communication (Carpenter & Liebal, 2011). According to this proposal there is an important qualitative difference between non-communicative eye contact (which could be even accidental or a result of checking where the other is looking) and ostensive-communicative eye contact that enables active sharing of attention. The communicativeness of eye contact could be conveyed by substantial (e.g., raising eyebrows, opening eyes widely) or subtle (e.g., lowering eyelids) movements of muscles in the eye region as well as by the timing and contingency of the eye contact in relation to things and events happening in the environment. In communicative eye contact both the bi-directionality of contact and intentional communication work together in promoting mutual understanding between partners. Eye contact serves as a platform for fast and efficient flow of information; both partners become senders and receivers of information at the same time so bidirectional attention contact can be established (Gómez, 1996). However, the communicative part is no less important. Only communicative eye contact is ostensive (Sperber & Wilson, 1986), signals communicative intention (Csibra & Gergely, 2006; Gómez, 1996, 2010) and allows partners to start experiencing the object of their attention truly together (Hobson & Hobson, 2007). Thus, communicative eye contact could generate common knowledge between partners and allows them to coordinate their decisions.

Interestingly, communicative eye contact might support successful coordination for a different reason as well: by creating a commitment to cooperate (Carpenter & Liebal, 2011). Bi-directional communication makes the content of the communication public (Gilbert, 2007; Taylor, 1985). A meta-analysis of conflict

games with adults demonstrated that when participants could communicate, successful coordination increased nearly 40% over baseline rates (Sally, 1995). Communication generally increases trust between social partners (Loomis, 1959), but it has been shown that the main reason why communication facilitates coordination in these contexts is that participants were exchanging promises to cooperate before the decision, and thus they created commitments (Bicchieri, 2006). Commitments serve the important function of reducing uncertainty in social interactions, for example, by generating specific expectations, by making existing expectations more salient, by activating reputation concerns and/or by making people's behavior mutually salient (Michael & Pacherie, 2015). Therefore, commitments help to stabilize both partners' behavior and coordination between them (Bratman, 1999).

Although there is a considerable amount of interest in verbal commitments (e.g., promising), it has been proposed that under certain conditions commitments might arise nonverbally as well (for details, see Part II). Communicative eye contact is a good candidate for signaling nonverbal commitments, because it is bidirectional, public and can communicate one's readiness to act in a certain way. For example, in Stag Hunt dilemma situations, both partners know that the appearance of the collaborative option is relevant. They also know together that if they have common knowledge about it, deciding for collaboration is in their mutual best interest. Because a partner who initiates communicative eye contact contingently upon the appearance of the collaborative option publicly signals that she knows about it (and therefore is prepared for collaboration), given the understanding that this is the best thing for both partners to do if possible, this behavior could be seen as a commitment to choose the collaborative option (Gilbert, 2007).

To summarize, the goal of the empirical investigation in Part II of this thesis is to explore the role of two types of eye contact (non-communicative vs. communicative eye contact) in solving a coordination dilemma. The goal is to advance our understanding of non-verbal ways of strategic decision-making in interpersonal situations and to contribute to discussions about the important but difficult question of how social partners achieve mutual understanding.

1.3 Focus of the thesis

This thesis has two parts: 1) A theoretical investigation into the distinct ways in which humans coordinate their attention and knowledge states with others, and 2) an empirical investigation into the role of sharing attention and communicative eye contact in particular in facilitating cooperative behavior.

More specifically, the first aim of the thesis is to address the problem that despite over 40 years of work and hundreds of publications on the topic of joint attention, there is still surprisingly little agreement in the field about what exactly joint attention is, and how the attending and knowing together state in it is achieved. Part of the problem, I propose, is that joint attention is not a single process, but rather it includes a cluster of different cognitive skills and processes, and different researchers focus on different aspects of it. A similar problem applies to common knowledge. The theoretical part of the thesis thus presents a new approach: I outline a typology of attention levels (from monitoring to common, mutual, and shared attention), along with corresponding levels of common knowledge and then consider cognitive, behavioral, and phenomenological aspects of the different levels. A key distinction in all of this is second-personal vs. third-personal relations. It is useful to distinguish these levels because they fulfil different functions and have different consequences in terms of what kinds of interactions and obligations they support. While the focus is mainly on joint attention and common knowledge, I also briefly discuss how these levels might apply to other 'joint' mental states such as joint goals.

The second aim of the thesis is to investigate closely the role of sharing attention in promoting coordination. Successful cooperation in the context of coordination dilemmas like the Stag Hunt game faces two main challenges. The first challenge is the uncertainty about one's partner's knowledge: "Does my partner know that the cooperative option is available?" It is not enough if each social partner knows individually that the cooperative option is available. Only if both partners know together (that is, if they have common knowledge) that the cooperative option is available can this challenge be overcome (Clark, 1996; Chwe, 2001; Rubinstein, 1989). The second challenge is about the uncertainty about one's partner's behavior: "Even if my partner does know about the cooperative option, will she cooperate?" To overcome this challenge, partners might create commitments to cooperate, as commitments stabilize cooperative behavior (Michael & Pacherie, 2015). However, there is an ongoing debate about how commitments can come about. Some scholars

propose that only certain verbal statements or conventional acts (e.g., saying “I promise”) will create promissory obligations, whereas others claim that the core of commitments is not rooted in social conventions, but instead in more general moral principles.

I hypothesized that communicative, but not non-communicative, eye contact can serve to create common knowledge about the availability of the cooperative option and thus promote expectations of cooperation between partners. I also hypothesized that communicative, but not non-communicative, eye contact might even be powerful enough to create a commitment between partners. This prediction was made because once shared attention is established, the fact that partners are paying attention to the same thing together is communicated, and therefore public, and, to some extent, undeniable. To test these hypotheses, I conducted two experimental studies with 5- to 7-year-old children. Although these hypotheses have not been tested in adults, testing young children enables to get at the origins of the ability to use attention sharing for coordination. In addition, testing young children provides an important methodological advantage, as studies can be designed as games, and allow to measure truly spontaneous behaviors.

2. Part I: A new look at joint attention and common knowledge

From infancy through adulthood, joint attention is one of the most important skills in human social cognition. In joint attention we focus on things together with others (Mundy et al., 2009). Thus, joint attention allows us to share experiences about the world with others, to coordinate our thoughts and behaviors, and to cooperate successfully with others (Tomasello, 1999; Trevarthen, 1979). It provides a foundation for early language development and social competence and facilitates social learning (Mundy, 2003). Individuals who lack the skills to follow and share attention with others have significant difficulties in relating to other people and sustaining relationships (Mundy & Newell, 2007).

Despite universal agreement about the importance of joint attention, however, there is still surprisingly little agreement on exactly what joint attention is and how it is achieved. Even now, more than 40 years after the first articles on joint attention were published (Bruner, 1974; Lewis, 1969; Scaife & Bruner, 1975), there is much variability in how different researchers define and use the term joint attention. For example, some focus mainly on behaviors and define joint attention simply as "looking where someone else is looking" (Butterworth, 1995, p. 29). Others go further and define it as "the intentional co-orientation of two or more organisms to the same locus" (Leavens & Racine, 2009, p. 241). Others bring in the idea of interest, defining joint attention as "the ability to coordinate attention toward a social partner and an object of mutual interest" (Bakeman & Adamson, 1984, p. 1278; see also Baron-Cohen, 1989; Camaioni, 1993; Mundy & Newell, 2007). Still others go further, considering what kind of awareness the individuals must have and arguing that to be in joint attention, the individuals must *know together* that they are attending to the same thing: They must have common knowledge of this (e.g., Carpenter & Liebal, 2011; R. P. Hobson, 2005; Tomasello, 1995). On the surface, these definitions may all sound similar – they all involve two individuals attending to the same thing – but actually, as we will see, there are important differences among them; differences which may account for many misunderstandings and unnecessary debates in the literature (Carpenter & Call, 2013).

A related long-standing debate concerns how the *jointness* and common knowledge in joint attention (and joint action) are achieved. There are definitional and terminological issues here as well: Some authors use the term "common knowledge" (Lewis, 1969) and others use the term "mutual knowledge" (Pinker,

2007; Schiffer, 1972) to describe this phenomenon, whereas others reserve "mutual knowledge" for the state in which you know X and I know X, but nothing else (Vanderschraaf & Sillari, 2014). Still others use "common knowledge" and "mutual knowledge" interchangeably (e.g., Clark, 1992, 1996). Regardless of which term is used, there is little agreement about how we can know something together with someone else. Some authors have proposed that individuals infer common knowledge through a finite number of recursive steps (e.g., three to five, Bach & Hamish, 1979; Harder & Kock, 1976). In some of his writings, Tomasello (e.g., 2008, 2011) takes a modified version of this approach, claiming that some type of recursive mind-reading is involved in joint attention, even in infants. However, others have pointed out that the processing demands for even just a few steps of recursive thinking are high, thus casting serious doubt on whether this can be how we – even as adults – achieve common knowledge instantaneously and effortlessly in most everyday circumstances (e.g., Clark & Marshall, 1981; Sperber & Wilson, 1986). Clark (1992) has suggested that we represent common knowledge as a 'single mental entity', e.g., "You and I mutually know X". Similarly, Pinker (2007) proposed that we employ a simplified recursive formula, such as Y: "You and I know X, and you and I know Y", where Y is both the entire formula and part of it.

More recently, there is growing momentum in the idea that there is a crucial difference between the social cognition process when we act as detached observers of others (adopting a third-person perspective) versus when we directly engage in interaction with them (adopting a second-person perspective). It has been claimed that the primary way of understanding others is not through observing but rather through interacting and experiencing with them; consequently, there are calls to increase scientific attention to second-person engagement (e.g., DeJaegher, DiPaolo, & Gallagher, 2010; Gallotti & Frith, 2013; Heal, 2005; Michael, 2011; Reddy & Morris, 2004; Reddy, 1996, 2018; Schilbach, 2010; Schilbach et al., 2013; Zahavi, 2008, 2011). Regarding joint attention more specifically, it has been suggested that a second-person relation between individuals (Eilan, submitted; Gómez, 2005) and intersubjective engagement (R. P. Hobson, 2005; Reddy, 2011) are necessary conditions for joint attention. These ideas have direct implications for how the jointness and common knowledge in joint attention (and joint action) could be achieved.

Therefore, here we develop these ideas further by taking a new look at joint attention and common knowledge and systematically distinguishing several closely related but clearly distinct phenomena within them. We begin by outlining a typology of different social attention levels, which are currently all referred to in the literature as joint attention. We call them 'social attention' because all levels include consideration of another person's attention; thus all levels depict some form of triadic relation between self, other, and an object of their attention. Within these levels, we propose corresponding distinctions for the social knowledge states that support these attention levels. We then introduce the idea that the social attention and knowledge levels exist on a 'scale of jointness' (i.e., in terms of how much the other is in mind, the certainty of jointness, and the felt connection with others), and describe several factors (e.g., goals, saliency, timing and contingency, behaviors) that can cause shifts in this scale. Next, we discuss the complexity of social attention and some of the different functions of these levels. Finally, we suggest directions for future studies: We briefly sketch how the levels map onto the available findings in human infants, animals, and social robots, and show how the typology can be applied to other joint mental states.

2.1 A typology of social attention and social knowledge

First, a few general notes. Although the literature on joint attention has primarily focused on the visual modality, note that each of these levels can apply to all types of external sensory inputs (visual, tactile, auditory, olfactory, and gustatory e.g., Bigelow, 2003); that is, not only can we look at something together, we can also, for example, listen to or smell something together. Moreover, objects of attention can be present, past, future, or imaginary events and states (including the lack of something); the self or the other (Reddy, 2003); and even mental states such as thoughts, beliefs, etc. (O'Madagain & Tomasello, submitted). In addition, joint attention can come about through both bottom-up processes (i.e., through automatic, reflexive shifts of attention to a salient stimulus) and top-down processes (i.e., in an active, goal-directed way; Brinck, 2004; Kaplan & Hafner, 2006). Finally, for the sake of simplicity, most of the time we talk about two individuals, but of course more individuals can engage in each level as well. We use examples involving humans, but our framework does not presuppose that only humans are able to engage in these levels (see below for more on this).

In what follows, we define four levels of social attention (monitoring, common, mutual, and shared). We categorize the levels according to the participants' perspective (i.e., detached, third-person perspective vs. engaged second-person perspective), the type of knowledge they achieve from being in these attentional states (individual vs. common knowledge of different types), the participants' dependency on each other (independence vs. interdependence), and the kind of experience the participants have (individual vs. co-created), and we discuss some of the behaviors involved in each level. Above we used the term common knowledge to refer generally to states that involve 'knowing together', as it is a well-established term. However, in what follows, we further distinguish common, mutual, and shared knowledge to refer to the specific knowledge states associated with the corresponding attention levels. We use the term *knowledge in common* as an overarching term to refer to different ways of knowing together in general.

A necessary pre-condition for each of the four social attention levels is that one is able to engage in *individual attention*. Individual attention simply means attending to something while engaging with the environment from a first-person perspective only. The attender is completely independent of others (no others need even be around), she has *individual knowledge* about the object of attention, and there is no intention or motivation to follow into or change or connect with someone else's attention state.

2.1.1 Monitoring attention

An individual is *monitoring attention* when she takes a third-person, observer's perspective on a second individual, and attends to what the other is attending to.

Example 1: On an airplane, Allison notices that Rob, who is sitting next to her, suddenly moves his head closer to the airplane window and looks out. Allison follows Rob's gaze and sees another airplane passing by and, at the same time, knows that Rob sees the airplane as well. Example 2: Annie is attending an anti-government demonstration. Suddenly, the sound of a loud siren fills the area. To protect her hearing, Annie covers her ears with her hands. A stranger, Alvy, stands in front of Annie with his back turned to her. Annie sees that Alvy also covers his ears when the siren starts. Annie is paying attention to the siren and, at the same time, focuses on the fact that Alvy is paying attention to the siren as well. Annie has *individual*

knowledge about the siren and about the fact that Alvy perceives the siren as well. Her attention level is *independent* of that of Alvy in the sense that she can engage in this level separately, without Alvy's active contribution or even awareness that she is there. Annie is having an *individual* experience. In this level, individuals do not have to be in the same level (as they do in the common, mutual, and shared levels below). Here, Annie is monitoring attention and Alvy is in individual attention. However, note that even when both individuals are monitoring each other's attention simultaneously, each just assesses the attention and knowledge state of the other individually, so both their experience and their knowledge remain individual. Often, monitoring attention results in an observable change in the monitor's behavior (e.g., turning one's gaze, head, and/or body orientation to look at what the other is looking at), but it is also possible to monitor someone's attention covertly, without producing any easily observable behaviors.

One can go beyond monitoring another's attention and additionally manipulate it. An individual is *manipulating attention* when she takes a third-person, observer's perspective on a second individual, monitors the other's attention, and then acts in such a way as to get the other to attend to something else. Like monitoring attention, manipulating attention involves *individual knowledge and experience* and can be *independent* of the other in the sense that we can manipulate others' attention without even being there (e.g., if we leave something out in a conspicuous place for them to see when they get home).

One can intentionally manipulate another's attention in several ways. First, one can increase the saliency of events such that they capture the other's attention. For example, if a guest at a dinner party does not want to ask directly for more wine, she can surreptitiously move her empty glass into the host's visual field so the host will likely notice it and refill it (Sperber & Wilson, 1986). This requires monitoring what the host can/will be able to see. Second, one can directly manipulate someone else's attention via a sort of unidirectional communication. The goal of unidirectional communication is only to convey information (e.g., there is something over there) from one individual to another, not to create a common focus of attention (in contrast to the bidirectional communication that takes place in the shared attention level below). Gestures such as pointing can be used to direct someone's attention in this way. For instance, when the siren starts, Annie turns to Alvy, points ambiguously

to the side while shouting, “The police are coming, look!” Alvy searches for the police. Annie uses the moment of distraction to steal Alvy’s bag and run away.

In contrast to the monitoring attention level, which involves individual knowledge and experience, some type of common knowledge about the focus of the individuals’ attention is an integral part of each of the next three (common, mutual, and shared) levels. That is, in each of the next levels, it is necessary that both individuals are in the same level in order to achieve these levels.

2.1.2 Common attention

Two individuals are in *common attention* when they each take a third-person, observer’s perspective on the other, more or less simultaneously attend to what the other is attending to, and infer not only that they are both attending to the same thing, but also that they are attending to each other’s attention to the thing.

Individuals can engage in common attention when 1) the object of attention is salient or public (so they can assume that they are attending to the same thing) and 2) the other’s attention is relevant for them (so they both have a reason to consider the other’s attention, for example they are in close physical proximity, or they have a previously-established joint goal, or they want to predict each other’s actions). Under these conditions, individuals could know that they are attending to each other’s attention without any contact or communication.

Example 1: Suddenly, the lights go off in the plane and Allison and Rob find themselves in darkness. Allison is attending to the fact that it is dark, to the fact that Rob must also be perceiving the darkness, and to the fact that each knows that they each know this, and Rob is engaging in the same attentional processes. Example 2: Alvy stands next to Annie, so they are aware of each other’s presence. Annie is attending to the siren, to the fact that Alvy is attending to the siren, and to the fact that each knows that they each know this since the siren is so loud that it is clear that everyone in the vicinity must be hearing it. Alvy is engaging in the same attentional processes. Annie and Alvy have *common knowledge* about the siren achieved via recursive assumptions, inferences, and perspective-taking, since there are no specific external behaviors (beyond monitoring attention) associated with this level. Their attention level is *dependent* on that of the other in the sense that, from a bird’s-eye perspective, both of them must engage in these attentional processes in order to

achieve this level. From each individual's perspective, however, each evaluates separately whether they are in common attention and each might be mistaken about it (e.g., one believes that they are in common attention but later finds out that they were not). Therefore, Annie's and Alvy's experience, when reasoning about each other's attention, is *individual*.

2.1.3 The question of jointness

This last point is key in separating the common attention level from the next two (mutual and shared) levels. According to one of the richest definitions of joint attention, individuals must *know together* that they are attending to the same thing (Tomasello, 1995). It might seem that already in common attention Annie and Alvy are attending together to the loud siren, because they have common knowledge about it. However, we would argue that they are not truly attending *with* each other yet, as they have not yet made 'attention contact' with each other (Gómez, 1996), actively *shared* the information in any way yet (Carpenter & Liebal, 2011), or in Taylor's (1985) terms, they have not yet put it out there *'between us'* as something to focus on together. Because in the common attention example Annie stands in a third-person relation to Alvy at this moment, she must represent and assess Alvy's knowledge recursively (he knows that I know that he knows...). With this classic recursive approach (or some simpler version of it; Clark, 1992; Pinker, 2007), they can each know that they each know about the siren and each other's knowledge of both it and their own knowledge, and thus can achieve a type of common knowledge. However, this level of attention and knowledge does not seem to be known truly together in any meaningful way. In contrast to the 'meeting of minds' that takes place in the classic 'joint attentional triangle', it is basically two individual perspectives that never meet in the middle: Each individual just assesses the attention and knowledge states of the other individually (Calabi, 2008; Carpenter & Liebal, 2011).

In third-person relations, fairly complex processes including recursive assumptions, inferences, and perspective-taking are used to achieve common knowledge. However, in second-person relations, engagement can allow for a different, more direct and non-inferential processing of the situation, which is not possible outside of this type of relation (Gómez, 2005; Reddy, 1996). Thus the proposal is that we should reserve the terms 'mutual' and 'shared' for interactions

that involve this second-personal relation – where it is really done *together* (Eilan, submitted; Gómez, 2005; Reddy, 1996, 2011; Zahavi, 2015).

In a second-person relation, one no longer has a detached, observer's attitude toward the other and his or her attention; the other is not perceived as 'he' or 'she' (Reddy & Morris, 2004). Instead, both individuals are in direct engagement with each other and treating each other as 'you', and both are senders and receivers of information at the same time (Argyle & Cook, 1976; Zahavi, 2015). Direct social interactions provide both partners with a wealth of cues about what is relevant and salient, and where each other's attention is focused, so there is no need to represent the recursive structure of the situation. Importantly, each partner becomes a "constituent part" of the other's experience (Campbell, 2005; Zahavi, 2015). That is, their perspectives and attention to the object of attention are colored by their mutual awareness of each other's attention, and the experience differs qualitatively from the situation in which the individuals attend to the same thing but are not relating as 'you' to each other (as in the monitoring and common attention levels above).

Although second-person engagement is typically described as being an extremely rich and in-depth intersubjective experience, actually it comes in degrees. Even within an interactional context, we can take different approaches towards the other (i.e., instrumental, detached, and objectified vs. engaged, involved, and attuned) – not all interactions involve mutual engagement (e.g., Buber, 1958; Fuchs, 2013, Reddy, 2018; Schilbach, 2013). We will thus use the term second-person relation to refer to situations in which individuals adopt an engaged approach towards each other in an interactional context. Both the intensity of interactive exchanges and participants' approach influence the richness of a second-person relation. A prolonged, face-to-face conversation between two close friends represents a rich version of this, but even a brief encounter between two strangers making eye contact can establish a second-person relation, although a very minimal one, as it already allows for bidirectional contact and signals openness for engagement (in contrast we skillfully avoid eye contact when no engagement is desired.)

This is the crucial difference that we would like to highlight as we think about the difference between the previous levels and the next two: The previous levels have a unidirectional and individualistic nature; the next levels, mutual and shared attention, are bidirectional and relational. The main difference between the next two

levels lies in the presence or absence of intentional bidirectional communicative exchange about the object of attention.

2.1.4 Mutual attention

Two individuals are in *mutual attention* when they are engaged in a second-person relation to each other, while they are more or less simultaneously attending to the same thing, such that both of them experience each other attending both to that thing and to their own attention to the thing and each other. Thus, mutual attention involves both individuals being aware (in a second-person relation to each other, but without intentional communication) that they are attending to the same thing with each other. Most prototypically, the individuals make eye contact about the thing. When their eyes meet, both know that they are attending to each other and 'attention contact' is established (Gómez, 1996). When the thing in the environment is salient and/or relevant for both individuals, eye contact enables them to gain reciprocal and reactive information about their attention to it. Another example of such bidirectional behavior is mutual touch (Botero, 2016) (e.g., squeezing hands).

Example 1: When the lights go off in the plane, Allison and Rob grasp each other's hand reflexively. Example 2: When the siren first starts, Annie looks around to see what is happening. So does Alvy, and their eyes meet. As their eyes meet, it is mutually manifested that they are both attending to the siren and that they both know that they both know this. Annie and Alvy have *mutual knowledge* about the siren achieved via experiencing it with each other. Their attention level is *dependent* on that of the other in the sense that both of them must engage in these attentional processes together in order to achieve this level. Their experience is qualitatively different from the individual experiences in the previous levels; they *co-create their experience* and it is colored by their direct mutual awareness of each other's attention. Even though the eye contact signals to both individuals that the contact was established, in mutual attention neither of the individuals intentionally communicates about the object of their attention.

2.1.5 Shared attention

Two individuals are in *shared attention* when on top of an interaction that meets the criteria for mutual attention, in addition they intentionally communicate with each other about the object of their attention. In contrast to mutual attention,

which apparently typically comes about incidentally (e.g., via incidental eye contact), the state of shared attention is achieved intentionally and can often be an end in itself (Gómez, Sarriá, & Tamarit, 1993).

Example 1: Rob sees the other airplane passing by in the distance. He points out the window for his son sitting next to him, waits until his son looks out at the airplane, and then they engage in eye contact and start talking about the plane. Example 2: When the siren sounds, as Annie starts looking around to see what is happening, so does Alvy, and their eyes meet. As their eyes meet, Annie raises her eyebrows questioningly, and Alvy responds with a concerned look. From the communication inherent in shared attention, Annie and Alvy gain several layers of *shared knowledge* about the siren (see below). Annie and Alvy's attention level is *dependent* on that of each other in the sense that both must engage in these attentional processes in order to achieve this level. Again, they *co-create their experience*, but the experience is different from mutual attention as the communicative exchange allows them to actively align psychological states such as attitudes, emotions, and goals about the object of attention with each other.

In the shared attention level, individuals follow and direct others' attention, but in contrast to the corresponding but third-person monitoring and manipulating attention above, the goal here is to create a bidirectionally-shared focus of attention (Gilbert, 2007). Often, in the top-down case, one partner intentionally directs another's attention to something they can focus on together, using behaviors that are intentionally communicative, checks that the other has perceived it, and the other partner confirms, communicatively, that attention is shared. Shared attention is thus characterized by behaviors in which individuals confirm to each other that they are attending to the same thing. The communication need not be verbal (although often it is, e.g., "Look!"); it could also, or in addition, simply involve 'communicative' or 'sharing' looks (e.g., Carpenter & Liebal, 2011; J. A. Hobson & Hobson, 2007), that is, looks that express some message. Other prototypical examples of nonverbal behaviors are declarative gestures such as pointing and showing. This confirmation is present in the bottom-up case as well. When a stimulus is salient, partners first automatically shift their attention to it, and then they confirm verbally (e.g., "That was loud!") or nonverbally (e.g., eye contact with nodding or smiling) that they are sharing attention (Brinck, 2004; Kaplan & Hafner, 2006).

Several layers of messages can be communicated in shared attention, which add to the layers of shared knowledge that can be gained. Not only do the partners have shared knowledge about the object of attention (the main referent or topic of their communication, e.g., we hear the siren), they also have shared knowledge about the communication itself (e.g., we intentionally communicated about our attention to the siren). In addition, partners also typically convey an attitude or comment about the object of attention (Bruner, 1975; Carpenter & Liebal, 2011; e.g., we are concerned about the situation), so they have shared knowledge about this as well. Sometimes they may even comment on the fact that attention is shared, or "share the experience of sharing" (R. P. Hobson & Hobson, 2008, p. 79), for example when an infant and her mother look at each other with a smile while playing, for no other reason than to express pleasure about their shared activity. The communication inherent in this level means that all of these things are out in the open, public, and, to some extent, undeniable (Carpenter & Liebal, 2011). All of these layers create more specific shared common ground about the object of attention, the partners' attitudes about it, and their shared experience, and contribute to building up the partners' conversational record (Lewis, 1969).

To summarize, here we have offered a new typology of different social attention and corresponding social knowledge levels, highlighting the difference between third-person and second-person relations. We suggest that we are not jointly attending to something with someone until we enter into a direct, reciprocal, and engaged second-person relation. The most prototypical case of this is the shared attention level; the minimal version is the mutual attention level. The levels have a cumulative structure in terms of the layers of knowledge involved. Using the siren example above:

i) In all four social levels, I know that I hear the siren and that you hear the siren.
ii) In the common, mutual, and shared levels, additionally, I know that each of us knows that each of us hears the siren and that we each know that we each know this about each other.
iii) In the mutual and shared levels, additionally, I know that we are both experiencing together our hearing the siren and both know this.
iv) In the shared level, additionally, we both know that we communicated to

each other about the siren and thus that we have confirmed to each other that we hear the siren together.

See Figure 2 for a schematic depiction of each level, and Table 1 for a summary of some of the characteristics of each level.

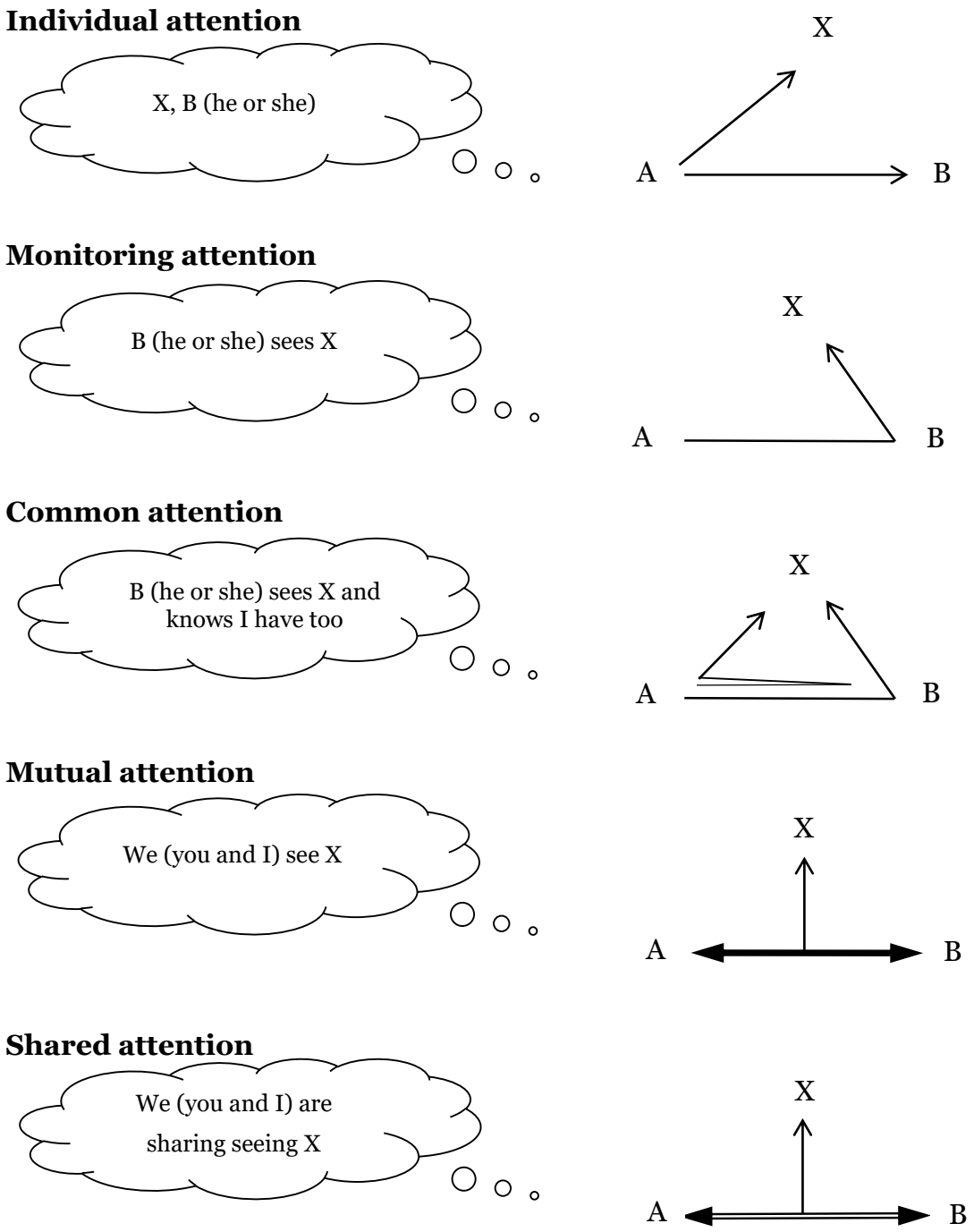

Figure 2. The figure is from the perspective of target individual A. X is the object of attention. Arrows represent what A is attending to. The solid two-way arrow represents non-communicative eye contact, and the special two-way arrow represents communicative eye contact (and/or other bidirectional communication). Thought bubbles represent what A has in mind (this figure depicts a visual example).

Table 1.

Characteristics of the different attention levels. The same characteristic across levels is marked by the same color.

	Individual	Monitoring	Common	Mutual	Shared
Type of perspective	First person	Third person	Third person	Second person	Second person
Description of the experience	I attend to X	He or she attends to X	He or she attends to X and knows I do too	We (you and I) attend to X	We (you and I) attend to X and are actively sharing this
Type of interaction	Dyadic (i.e., subject and object)	Triadic	Triadic	Triadic	Triadic
Knowledge gained	Individual knowledge	Individual knowledge	Common knowledge	Mutual knowledge	Shared knowledge
Does the second individual know that the first is in that level of attention?	Not necessarily	Not necessarily	Yes	Yes	Yes
"Intimate I+you we feeling"[2]	No	No	No	Yes	Yes, stronger
Commitments and obligations	No	No	Yes (but plausible deniability is possible)	Yes (but some plausible deniability is still possible)	Yes

[2] There are several types of "we feeling" (Zahavi, 2015). Here we distinguish a) the intimate/private/personal we feeling arising from a bidirectional, direct connection with others (this we feeling is subjectively strong) and b) the associative we feeling arising from association with others, arising because individuals have something in common (e.g., they both are attending to the same thing; this feeling comes from knowing rather than experiencing). It is interesting to note here that while in mutual and shared attention, almost by definition you know exactly who you are attending with, in common attention this does not need to be the case. For example, I can know that everyone (with normal hearing) within a certain radius must have heard the noise of a loud explosion, without knowing exactly who those people are.

2.2 A scale of jointness

The social attention levels exist on a scale of jointness, with monitoring attention (in which we have an individual experience of the other's attention) on the left side, representing no jointness, and shared attention (in which we truly attend to something together with the other) on the right side, representing the highest degree of jointness (see Figure 3). Aspects of jointness include (but are not limited to) the perceived psychological closeness between the individuals, how much the other is in mind, the connection felt (the 'we feeling'; Zahavi, 2015), and the certainty about the object of attention and/or that attention is joint (see below for more on this latter point). For the sake of simplicity we provided prototypical examples for each level above, but it is important to note that in reality there exist stronger and weaker examples in each level, that is, that there is a continuum within each level too. Consider the siren example in the mutual attention level. This situation could range from accidental eye contact (while Annie and Alvy are instinctively looking around), to more intentional eye contact (e.g., Annie and Alvy each look at each other to check if the other hears the siren too, and they happen to do this at the same time), to eye contact with mutual reactiveness (during the eye contact Annie and Alvy both reflexively cover their ears). These examples lie to the left, middle, and right side of the scale in the mutual attention level, respectively.

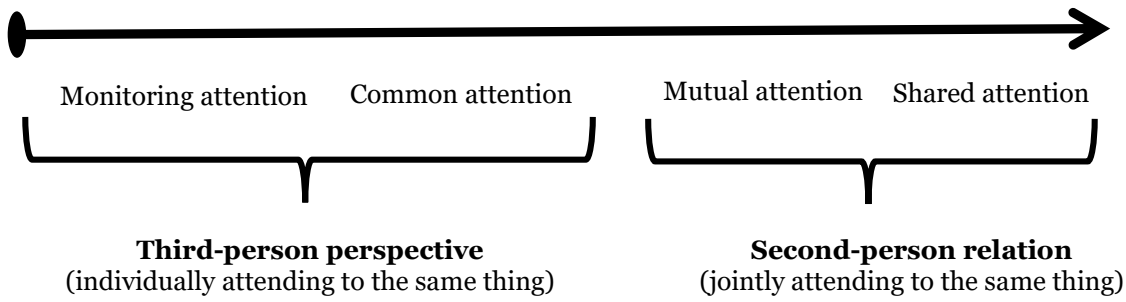

Figure 3. A scale of jointness.

2.2.1 Factors that can cause shifts in the scale of jointness

There are a variety of factors that can make it more likely that we move along the scale, both along the different levels of the scale and within each level:

- **Saliency:** Salient stimuli generally push situations to the right on the "jointness scale." It is easier for individuals to establish all of these types of attention to very salient things (e.g., an elephant standing in the lobby of one's office building) then to subtle things (e.g., a new vase standing in the lobby). However if a stimulus is too salient and attention-grabbing, we might focus solely on it and be stuck in individual attention (e.g., when a bomb explodes, we might at least initially just focus on ourselves and our own safety before thinking of anyone else).
- **Goals:** The existence and strength of a relevant shared goal between the individuals, and their interest in it, pushes situations to the right on the "jointness scale." If Annie and Alvy had recently spent long hours planning how to avoid being arrested at the anti-government demonstration, when the noise of the siren fills the area it is probable that they will be more on the right side of the scale compared to a situation in which they had not had a shared goal.
- **Type of common ground**[3]**:** Similarly, the strength and depth of the common ground the individuals have can push the situation to the right of the scale. Increasingly joint types of common ground generally range from no common ground, to common ground for rational beings, to cultural common ground, to common ground from co-presence, to personally-shared common ground (Clark, 1996; Clark & Marshall, 1981). Example: Robin and Max are waiting at a bus stop and suddenly a bomb explodes in front of them. Contrast the situation in which Max is the 2-year-old son of the woman standing next to Robin and the situation in which Max is an adult colleague who had just recently discussed with Robin a newspaper article saying that there is a security threat in their country (so, in addition

[3] Common ground suffers from terminological and conceptual confusion as well. In this paper we will use "common ground" as an umbrella term encompassing common knowledge, beliefs, and assumptions coming from different sources (e.g., co-presence, joint personal experiences, cultural background, etc.; Clark, 1996).

to all the other types, they have personally-shared common ground). In the first case, after looking communicatively to each other, because they share little contextually-relevant common ground, Robin and Max are on the left side of the scale in shared attention about the fact that the bomb exploded. In contrast, in the second case, they have quite specific shared common ground that is relevant to the explosion, so they are on the right side of the shared attention continuum. On top of their shared attention about the bomb exploding, they are sharing a more complex and detailed message such as, "Oh, no, it is here! We talked about this yesterday." Note that if we have very specific shared common ground (especially recently shared common ground) we can communicate much more complex messages with less obvious behaviors. For example, if two employees have just been complaining about their boss, the briefest, most neutral and surreptitious glance between them during a meeting with her can convey a complex message (e.g., "See what I mean?"), especially if the look occurs contingently to something relevant that their boss just said.

- **Contingency and timing:** As suggested above, strong contingency and/or a short time interval can push situations to the right on the scale (e.g., in the shared level, contingency and a short time interval between the appearance of the object of attention and the eye contact/communication, or between the formation or most recent discussion of the shared goal or common ground and the appearance of the object of attention). For example, in the siren example, consider the difference between cases in which Annie looks communicatively at Alvy immediately after the siren starts, 10 seconds later, or 10 minutes later. An immediate look would create a stronger feeling of jointness (lying more on the right side of the shared attention level) than a delayed look, and if they looked at each other 10 minutes later, it might not be clear what the other's look was about.
- **Perceptual space:** Perceptual space with clear borders (an office, a department's mailing list) or small perceptual spaces (a car) push situations to the right on the scale (especially with regard to the certainty that we are attending to the same thing with someone) compared to perceptual spaces

with vague borders (a university, Facebook) or huge perceptual spaces (a city, a concert hall).[4]

- **Behaviors:** Observable behavior, ranging from unintentional or reflexive behaviors to verbal or otherwise explicit communication, pushes situations to the right on the scale. For example, when someone sneezes, it might send a signal to others that he can smell the pepper spilled on the table, or when someone gasps or sighs while watching a movie, it might send a signal to others that he was attending to something in the movie. And of course in the shared attention level, sharedness becomes more clear depending on the explicitness and ostensiveness of the communication. To illustrate, consider the example with the siren. As Annie's and Alvy's eyes meet, if they look silently at each other with communicative looks after the siren filled the area they would be more to the left of the continuum in the shared attention level than if Annie looked to Alvy and said, "Oh, this is bad!" and Alvy agreed.
- **Individual differences:** People who tend to have others in mind more often or who have a higher motivation to share appear on the right side of the scale more often. One reason for this is simply because they initiate shared attention more often.
- **Relationship closeness:** People who are close to each other end up on the right side of the scale more often. There is a natural tendency to have close others in mind and share attention more with people who are close (family members, friends, partners).

2.2.2 Certainty

Two kinds of certainty are important when attending socially with someone: 1) certainty about the object of attention – how can we know with

[4] There is an interplay between the type of perceptual space and the type and saliency of the stimulus. For example, it is difficult to guess where the borders are for being able to perceive sound from an explosion, but it is easy to guess that the smell from freshly made cookies in a room will not reach the space outside the room when the door and windows are closed.

certainty that we are attending *to the same thing* as someone else? and 2) certainty that attention is joint – how can we know with certainty that we are attending to the same thing *with someone else*? Several ways of gathering information about other people's attentional focus exist in different situations and influence our certainty. Some require fewer assumptions than others, for example you can see your friend seeing X (gaze following), but you can also just hear (but not see) that your friend is co-present, or just assume that your friend is co-present, and therefore you can assume that your friend perceives X. Similarly, fewer assumptions are required when we can directly perceive X in the environment (e.g., we can see/smell/hear X) (perceptual attention) than when we must infer the relevant object of attention (conceptual attention, e.g., to absent referents, for instance, an infant pointing about a missing puzzle piece for her mother).

When the stimulus is very salient, the certainty that you are attending *to the same thing* as someone else does not necessarily change with the different social attention levels. We are sensitive to perceptual co-presence with other people, and expect similar perceptual or attentional abilities from them, unless we learn otherwise (e.g., we find out that someone is visually- or hearing-impaired; Clark, 1996). Therefore, when a loud siren fills the area and you can hear it, you know that everyone else in the vicinity must be hearing it too.

In contrast, being in the higher attention levels increases the certainty that we are attending to the same thing *with someone*. For example, imagine that you are sitting next to your friend in a train, and you see that, as soon as the woman sitting opposite you leaves for a moment, the man sitting next to her reaches into her purse and steals her wallet. Even before you and your friend engage in any behavior toward each other, you infer that the stealing is salient enough to grab the attention of your friend as well, and that your friend realizes the same about you, so you have *common knowledge* about the thief (common attention). However, your certainty that you are attending to this together increases considerably once you engage in eye contact after seeing it (mutual attention), and even more once you exchange concerned looks and/or point (shared attention).

Note that it is often impossible to be 100% certain that you are attending to the same thing with someone. Certainty is probabilistic and instead of "knowing" that you attend to the same thing, it is more precise to say that you "believe" this. That is, you each believe that you are attending to the same thing together, and that the other believes this with some probability too. How high the probability needs to be depends on the current purposes. So although we use terms like 'knowing together' and common, mutual, and shared knowledge, more specifically we mean belief with sufficiently high probability for those purposes (Clark, 1996; Monderer & Samet, 1989). To illustrate: If you consider confronting the thief together with your friend, certainty is important and thus you likely would not want to rely on common attention, but instead would engage in shared attention.

2.3 The complexity of social attention

Joint attention is often discussed in the literature as a discrete, momentary occurrence, in which individuals attend to a single object or event. However, many everyday social attention episodes are much more complex. First, most of the levels (i.e., all except mutual attention) can be quite extended in time. For example, a spy could monitor the attention of someone all day, and we can be in common or shared attention with other audience members at a concert or movie for hours.

Second, in any given situation, we can move from individual attention to any other level, or move among the different social levels, instantaneously. Sometimes one level might function as a precursor to another level (e.g., when you want to show a hot air balloon to your partner, you might first monitor your partner's attention, to determine where your partner's focus is, and then choose the best way to initiate attention sharing). Sometimes we can even engage in two or more levels simultaneously (e.g., if you and I are monitoring the attention of someone else together, or when I take a third-person perspective on you during a second-person interaction, thinking, "Why is she looking at me that way?"; Reddy, 2018). This can result in a number of hierarchically-embedded states of attention occurring at the same time, either with the same or different people. At a movie, for instance, I can simultaneously monitor the

attention of a character on the screen, be in common attention with other audience members that the character is looking at something, and turn to share attention to the thing the character is looking at with my partner. Similarly, the whole 'going to the movie' experience is a shared attention episode with my partner, and there are numerous hierarchically-embedded, smaller shared or mutual attention episodes within that experience, when we, for example, agree that the popcorn is very salty, or squeeze each other's hands reflexively at a scary moment, etc. Thus, although we mainly discuss the different social attention levels separately in order to highlight the important differences across them, we note that complex relations can also exist among and within them.

Third, the attention levels take place live, in real time. When we exchange letters with someone, although this is reciprocal communication, it is not shared attention (though real-time, immediate texting, in contrast, may be). However, we do create shared knowledge and common ground from these types of exchanges. We can also engage in virtual sharing in which we just imagine sharing attention with someone, for example, when I celebrate my country's team's goal while watching the World Cup abroad and imagine my dad cheering 'with' me in our home country. Still, direct, real-time experience of social attention has unique qualities that cannot be fully attained with experiences that are sequential or delayed in time or space.

2.4 Functions of the different levels

It is useful to distinguish these levels for several reasons. First, as mentioned above, it should help in eliminating unnecessary confusion and debate in discussions about whether, for example, different animals can engage in joint attention (see below for our take on this). In addition, as we will discuss here, the different social attention levels serve different functions.

Monitoring attention brings some advantages to individuals who are capable of it, over and above individual attention. First, it can be useful simply for gathering more information from the environment than one could on one's own, for instance when as a result of monitoring someone else's attention we discover something interesting or useful or dangerous that we might not have discovered on our own (e.g., Max is walking on a pier at a lake and he sees a

couple in front of him excitedly peering into the water. He approaches to find out what they are looking at and sees an alligator swimming there). Second, monitoring attention is useful in making informed predictions about others' future actions, so we can adjust our own actions accordingly. This is often useful in competitive situations, because then the information gathered can be used exploitatively to gain an advantage over others (Tomasello, 2008). But we can also monitor others' attention in prosocial situations (e.g., Pam is in a store and sees an elderly woman looking up, fixating her gaze frustratedly on something on a high shelf. Pam determines that the woman is looking at a box of tea, and so approaches and gets the box for the woman). Similarly, we can manipulate the attention of others to things in the environment either to reach our own goals (e.g., to get an object through others' help by making them attend to something we need, for example, our wineglass that needs refilling; or by distracting their attention deceptively) or to help others to reach their goals (e.g., by redirecting their attention to something relevant for them).

Common attention similarly helps us make informed predictions about others' future actions. However in common attention one can understand not only what someone else attends to/knows but also what the other attends to/knows about one's own attention/knowledge, and this allows for a considerably more complex appraisal of the situation. This can be especially helpful in solving 'coordination dilemmas' – situations in which it is in our interest to decide what to do based on our expectation of how our social partner is likely to behave, and our partner's decision likewise depends on what he or she expects from us (Clark, 1996; Chwe, 2001). For example, imagine that a group of protesters has agreed before going to a demonstration that, no matter where they are in the crowd, whenever they hear a siren, they will start shouting anti-police slogans. Although all the protesters want to participate as a group, no one wants to engage in this potentially dangerous act alone. When a siren sounds, the protesters' ability to engage in common attention with their fellow protesters to the sound of the siren may give them confidence to make the risky decision to start shouting. Being able to engage in common attention enables us to coordinate even in cases in which we are apart and contact or communication is not possible. However, in the common attention level we can

only coordinate attention towards salient or relevant things or events in the environment. If it is not clear that everyone can hear the siren, it would be more difficult or even impossible to engage in common attention with any certainty.

In contrast to the individualistic nature of these two levels, in which we must examine the situation and consider our partner's attentional state from the perspective of an outside observer, the direct nature of the attention contact in the mutual and especially shared attention levels, as well as the bidirectional, reciprocal evidence it provides about where each partner's attention is focused, makes communication, interaction, coordination, and collaboration much more efficient and fluent.

The reciprocal evidence of each other's attention that attention contact provides has further implications as well. Even the minimal attention contact present in the mutual attention level can create some feeling of obligation. For example, imagine that Jane slips on the wet floor and is having difficulty standing up again. As she looks around embarrassedly to see if anyone is seeing this, she accidentally makes eye contact with a man who was watching nearby. They end up in mutual attention about Jane's fall and her attempts to stand up, as this is a salient event. In this case, it would be difficult for the man to deny having seen Jane fall, because the same single event of making eye contact when seeing Jane's fall also removed any doubt in the man's mind that she saw him seeing her fall. Consequently, mutual attention can create some pressure to follow social norms in situations in which people need to decide whether to act in accordance with norms or not. That is, the man might feel more strongly that he should help Jane stand up when they have mutual knowledge about Jane's fall compared to having individual or even common knowledge about it.[5]

However this minimal feeling of obligation can turn into a much stronger feeling of commitment once attention is shared. That is, the bidirectional communication in shared attention makes public the fact that attention is shared. We can refer to the instance of sharing afterwards and it is difficult to plausibly deny both the fact that attention was shared and any

[5] We thank Sebastian Grueneisen (personal communication) for the examples highlighting how common, mutual, and shared knowledge heighten the pressure to follow social norms.

commitments or obligations following from this. This heightens substantially the pressure to follow norms in normative situations and facilitates social coordination above and beyond common and mutual attention.

Another helpful function of shared attention is that it enables us to create a shared focus even to subtle things in complex environments, which might not be perceived as salient and/or relevant for both of us at first. This is not only helpful in supporting conversations and joint action, but in learning and other contexts as well (e.g., for infant language acquisition).

Social attention is not only useful, but it can be a source of pleasure for us as well. This is especially pronounced for the shared attention level. It feels good to connect with others and to engage in shared attention (Schachter, 1959); we often engage in shared attention just for the sake of the sharing itself (Gómez et al., 1993). Even as 1-year-old infants, we are so motivated to share attention with others that we will actually turn away from interesting sights to share them with others – the sharing is just as pleasurable (or even more so) than the sight itself. We have a feeling of 'contact' with the other, and consequently we may feel a "we feeling" which causes changes in our experience (Campbell, 2005). In shared attention, we build our shared common ground. We know what objects and events we have shared, and can refer to them in the future. Sharing attention might also cause an increased sense of similarity (with all the benefits that that entails; Haun & Over, 2013), which may emerge because paying attention to the same things/events/ideas might lead to thinking that whoever is interested in the same things I am is probably like me. Similarly, both initiation of shared attention and responding to shared attention signal an interest in the other – an interest in exploring the world together. With shared attention we can also express trust, openness, and intimacy. There are individual differences in the ease with which we share attention with others. The closeness of our relationship with our partner may increase the probability of smooth, mistake-free episodes, and, in turn, successful episodes of shared attention likely increase bonding (see also Wolf, Launay, & Dunbar, 2015).

Thus, shared attention is important in forming attachments and sustaining positive interpersonal relationships. On the other hand, skilled deceivers can

exploit this to create an illusion of openness and use shared attention behaviors for deception in competitive or deceptive situations (e.g., being sure to make eye contact when lying).

2.5 Future directions

The next important step in social attention research is to come up with operational definitions of these levels so it will be possible to determine which levels different types of individuals, such as young children, animals, and social robots, can achieve. As social attention states cannot typically be measured directly, it is challenging to find the best ways of investigating them.

One approach is to look for specific behaviors. The question is which behavior(s) should be considered as indicative of each of the levels. We have described some behaviors that typically accompany the levels above. However, these particular behaviors might differ across species and are neither necessary nor sufficient for identifying any given level. For example, turning one's head/gaze in the direction of someone else's could be a result either of low-level spatial cuing or monitoring or following someone else's attention. And a quick glance between partners could either be an accidental meeting of the eyes or sharing attention via communicative but neutral eye contact (e.g., in a situation, such as in a group meeting, in which the signaler does not wish anyone other than the recipient to be able to discern his meaning). It might thus be helpful to focus on behavioral sequences, instead of isolated behaviors. An example of this could be looking at partners' interaction history: for example, what preceded the quick glance between them, whether the glance was contingent on something salient for them, and what behaviors followed afterwards. However, quite often we do not have access to others' interaction history or future interactions. And although useful, investigation only of observable behaviors without considering knowledge states cannot provide conclusive evidence about which levels different types of individuals can achieve.

Another approach is to design studies in such a way that participants can only succeed on the task if they are engaging in a certain attention and knowledge level. For example, it could help to focus on individuals' behavior in situations such as coordination dilemmas, in which their responses are

predicted by their attention and knowledge states. In these situations, shared attention and knowledge predicts the highest success rate in solving the dilemma compared to lower attention and knowledge states (for an example of this approach see Thomas et al., 2014).

Still another approach is to experimentally manipulate characteristic features of the different levels in order to empirically test claims about, for example, what level of attention participants are in or the role of communication in achieving shared attention (see, e.g., Liszkowski, Carpenter, Henning, Striano, & Tomasello, 2004; Moll, Richter, Carpenter, & Tomasello, 2008; and Siposova, Tomasello, & Carpenter, 2018, for examples of this approach). Future studies could also systematically manipulate the type or degree of engagement and/or communication of one partner and measure the partners' ability to establish the mutual or shared attention level. Here, again, the success of establishing the levels could be evaluated by placing participants in situations in which their responses are predicted by the attention level they achieve, or by self-reports in human older children and adults.

Another important direction for future research is to determine which levels various types of individuals show. It is beyond the scope of this paper to review this in depth, but next we sketch briefly how the proposed levels map onto the available findings in human infants, animals, and social robots. Note that as previous studies were not designed to distinguish among these levels, at this point it is not always possible to make firm conclusions about which levels different individuals can achieve.

2.5.1 Which attention levels can different types of individuals achieve?

2.5.1.1 Human infants and children

A pre-condition for each of the four social attention levels is some understanding of attention, or at least perception. For example, minimally, it is necessary that individuals understand that when others are 'looking at something,' they are 'seeing something.' It has been proposed that already in their first months infants demonstrate an awareness of others' attention by

producing emotional reactions when experiencing others' attention to themselves (Reddy, 2003), and by decreasing smiling after interrupted eye contact at the age of 3 months (Hains & Muir, 1996). Even newborns look preferentially at faces with direct gaze (Farroni et al., 2002) and perceived gaze shifts trigger corresponding gaze shifts in 3-month-olds (Hood, Willen, & Driver, 1998). However, it is not clear how much infants understand about others' attention at these young ages. It becomes clearer around 12 months of age that infants do have some understanding of others as capable of attending to things, when they follow the gaze of others to things that they cannot see themselves, for example around visual barriers and even to things behind them (e.g., Deák et al., 2000; Moll & Tomasello, 2004).

Although we presented the levels from monitoring, to common, to mutual, to shared attention (representing the degree of jointness involved), interestingly, the latter two levels are seen earliest in development. Most researchers agree that joint attention develops between 9-12 months of age; however others have claimed that the beginnings of joint attention actually are seen well before this. For example, Rossmanith and colleagues have suggested that during a picture book activity, 3-month-old infants already coordinate their attention between the book and their caregiver, and that by 6 months they jointly communicate affectively about the activity (Rossmanith, Costall, Reichelt, López, & Reddy, 2014). Similarly, Striano and Bertin (2005) claimed that 5-month-olds engage in joint attention after observing that they alternate gaze between an object and a partner's face. However, without any other supporting evidence of mutual or shared knowledge, the behavior of gaze alternation alone could be evidence of any of the levels: alternating individual attention between the object and the partner's face, monitoring the partner's attention, or common, mutual, or shared attention. So far, the earliest supporting evidence of shared attention with shared knowledge from experimental studies is with 1-year-old infants. Starting from at least 14 months of age, infants distinguish what objects they have and have not shared communicatively with their partner (Moll et al., 2008; see also Tomasello & Haberl, 2003) and use this knowledge both when pointing for others and when interpreting others' pointing gestures (Ganea & Saylor, 2007; Liebal, Behne,

Carpenter, & Tomasello, 2009; Liebal, Carpenter, & Tomasello, 2010; Saylor & Ganea, 2007). These experimental findings, and experimental findings showing that infants point to share interest (Liszkowski et al., 2004), together with the many observational findings of spontaneous production of declarative gestures beginning between 9-12 months (e.g., Carpenter et al., 1998), provide strong evidence for the shared attention level. As previous studies have not distinguished between mutual and shared attention, whether these levels develop simultaneously is an open question for future research.

So far the earliest experimental evidence of monitoring attention is with 18-month-old infants. For example, 18-month-olds, but not 14-month-olds, could distinguish which objects an experimenter knew about when they observed, from a third-person perspective, the experimenter manipulating the objects (Moll & Tomasello, 2007b). At least from 18 months of age children also can learn novel words when they overhear an experimenter using them with someone else (Akhtar, Jipson, & Callanan, 2001; Floor & Akhtar, 2006). So far there is little if any evidence of young children manipulating attention in a third-person context (see Grosse, Scott-Phillips, & Tomasello, 2013 for the study that comes the closest to this). It might sound surprising that monitoring and manipulating attention may emerge relatively late in development, as many studies have found the parallel abilities of following and directing others' attention in younger infants, but almost all of these studies were conducted in a second-person context. Thus it is not yet known exactly when the skills to monitor and manipulate others' attention from a third-person perspective (that is, outside of a shared attention episode and without the active contribution of their partner) first emerge in infants.

Similarly, common attention has not yet been directly investigated in infants – again the vast majority of infant studies take place within a second-person context. Several findings suggest that adults engage in the common attention level, and that common attention has an impact on adults' experience of the thing they are attending to, their relationship with their partner, and their ability to coordinate in social dilemmas (Boothby et al., 2014, 2016; Thomas et al., 2014; Wolf et al., 2015). With young children, the closest study so far is a coordination study by Grueneisen, Wyman, and Tomasello (2015a).

This study investigated young children's ability to choose the same of several options as a partner without being able to communicate with or see one another. Five-year-old and older children, but not 3-year-olds, successfully chose the same option. To solve the task, each child had to attend to the fact that one option was more salient than the others, that the other child was likely to perceive the same option as salient too, and to the fact that each knew that they each knew this.

It thus seems that infants develop mutual and shared attention before they are able to understand others' attention via observing and inferring it solely from a third-person perspective without any direct contact, although the claim that monitoring and common attention might develop later needs to be confirmed with future research. Several findings support the idea that the second-person relation is the primary relation in which infants develop their earliest social attention skills (Moll & Tomasello, 2007a; Peacocke, 2005; Reddy & Morris, 2004; Vygotsky, 1978). Moll and colleagues have directly contrasted infants' ability to attend to others' attention in a second- vs. a third-person relation (Moll, Carpenter, & Tomasello, 2007; Moll & Tomasello, 2007b). They found that although 18-month-olds were able to do this in both situations, 14-month-olds recognized what their partner was attending to only when they were engaged with her in a second-person interaction, not when they were only observing her from a third-person perspective. Similarly, even within a second-person relation, infants show improved social attention skills in interactions with higher degrees of second-person engagement. For example, infants are much more likely to follow gaze shifts after their partner engages in eye contact (Farroni, Mansfield, Lai, & Johnson, 2003; Senju & Csibra, 2008) or points for them (Deák et al., 2000) compared to only seeing their partner's gaze shifts or head turn.

The crucial role of social attention in interactions is especially evident when social attention skills are impoverished. Children with autism show both quantitative and qualitative differences in many social attention skills compared to typically-developing children (for a review, see Bruinsma, Koegel, & Koegel, 2004). Based on a series of experimental studies, Hobson (1993, 2005, 2007, 2014) has argued that these atypical behaviors are a result of

deficits in second-person engagement, specifically, in difficulties with experiencing others' attitudes and emotions directly in interaction and thus identifying with one's partner's perspective. Further studies are needed to understand what children with autism know about others' attention. Especially useful would be experiments investigating their ability to achieve shared knowledge.

2.5.1.2 Nonhuman animals

Again, to be able to engage in social attention, a pre-requisite is to have some understanding of other's attention/perception. Numerous animal species show different reactions when others are looking towards vs. away from them, including lizards (Hennig, 1977), snakes (Burghardt & Greene, 1988), birds (Ristau, 1991), horses (Proops & McComb, 2009), dogs (Virányi, Topál, Gácsi, Miklósi, & Csányi, 2004), and many primates (for a review, see Emery, 2000). Many species are also able to follow the gaze (or at least head) direction of others, for example tortoises (Wilkinson, Mandl, Bugnyar, & Huber, 2010), goats (Kaminski, Riedel, Call, & Tomasello, 2005), dolphins (Pack & Herman, 2006), horses (Proops & McComb, 2009), dogs (Miklósi, Polgárdi, Topál, & Csányi, 1998), and again, many primates (for a review, see Emery, 2000). However, the degree to which some of these species understand others as attending to things remains an open question. Experiments testing animals' ability to follow gaze to locations outside their own visual field provide clearer evidence that they have some understanding of others as capable of attending to things. For example, it has been shown that ravens (Bugnyar, Stowe, & Heinrich, 2004), wolves (Range & Virányi, 2011), dogs (Met, Miklósi, & Lakatos, 2014), monkeys (Amici, Aureli, Visalberghi, & Call, 2009) and all ape species (Bräuer, Call, & Tomasello, 2005) follow gaze behind barriers.

Whether animals can engage in joint attention is a hotly-debated question. Most discussions focus on apes: Several researchers have argued that apes do engage in joint attention (e.g., Leavens & Bard, 2011; Leavens & Racine, 2009; Tanner & Byrne, 2010), whereas others have argued that they do not (e.g., Carpenter & Call, 2013; Tomasello, 2006). Evidence taken in support of the idea that apes can engage in joint attention includes findings that gorillas

(Tanner & Byrne, 2010) and infant chimpanzees (Bard, Bakeman, Boysen, & Leavens, 2014) alternate gaze and gesture in triadic situations. Leavens and colleagues have also claimed that apes point declaratively (e.g., Leavens, Hopkins, & Bard, 2005; Leavens & Racine, 2009). In contrast, other studies have found little evidence of joint attention in apes, either when operationalizing joint attention as gaze alternations or declarative gestures (e.g., Bard & Vauclair, 1984; Carpenter, Tomasello, & Savage-Rumbaugh, 1995; Rivas, 2005; Tomasello & Carpenter, 2005; Tomonaga et al., 2004).

However, as noted above for human infants, gaze alternation is not in itself diagnostic of any of the social attention levels. Furthermore, the few reported 'declarative' gestures in apes were either elicited by human caretakers (who told the apes to show them something) or were actually other, different types of gestures (e.g., informative, rather than declarative, gestures; Carpenter & Call, 2013). Although it is possible that informative and other types of gestures can involve joint attention, they need not do so. Further, experimental research is needed to determine what apes' motivation is when gesturing, and whether, when communicating, they create a shared focus of attention (with shared knowledge) or whether instead they engage in a more unidirectional type of communication: manipulating, rather than directing or sharing, others' attention.

Future studies, along the lines of experiments conducted with human infants (e.g., Moll et al., 2008), are also needed to investigate apes' ability to have mutual and shared knowledge with their partners (whether they be conspecifics or humans). For example, a recent study has claimed that apes likely have shared knowledge with their partner, because they modify their gestures according to the familiarity of the partner (Genty, Neumann, & Zuberbühler, 2015). However, further work would be needed to distinguish whether in this study apes were using individual knowledge about their partner's knowledge (e.g., he knows this gesture – monitoring and manipulating attention) or shared knowledge (e.g., we know this gesture together – sharing and directing attention).

Another useful approach is to conduct experiments in which participants can only succeed on the task if they are engaging in a certain attention and

knowledge level. This approach was used in a study by Duguid and colleagues (Duguid et al., 2014), and they found that unlike 4-year-old children, who spontaneously shared attention communicatively with their partner and thus solved a coordination problem, apes did not communicate before their decisions, and therefore were less successful.

Finally, with regard to joint attention, it is an open question whether and to what degree apes (and other animals) adopt instrumental vs. engaged approaches towards their partner during interactions (Buber, 1958; Fuchs, 2012; Reddy, 2018; Schilbach et al., 2013; compare with Gómez's, 2005, distinction between different notions of others as subjects). Answering this question would help us to distinguish whether, when animals gesture, they are directing others' attention to share it, or they are manipulating others' attention.

In contrast to human infants, with animals there is far more evidence of monitoring attention than shared attention. The experiments testing animals' understanding of others' knowledge provide the most direct evidence that these animals can monitor others' attention from a third-person perspective. Dogs, for example, know what others can and cannot see (e.g., Kaminski, Pitsch, & Tomasello, 2013), and both monkeys (Flombaum & Santos, 2005; Santos, Nissen, & Ferrugia, 2006) and apes (Bräuer, Call, & Tomasello, 2007; Kaminski, Call, & Tomasello, 2008; Melis, Call, & Tomasello, 2006) know what others can and cannot see and hear.

Common attention has not yet been studied in other animals. Future research could adapt previous studies investigating what apes know about others' knowledge (e.g., Hare, Call, Agnetta, & Tomasello, 2000; Kaminski et al., 2008) to include the additional step of testing what apes know about what others know about what they themselves know.

2.5.1.3 Social robots

Social attention abilities are critical in social robotics as well. Social robots have a rudimentary sense of what others can see and can track others' attention in simplified environments (Lemaignan, Garcia, Jacq, & Dillenbourg, 2016). So there is some limited evidence for skills of monitoring attention.

Robots' subtle gaze shifts towards objects have been shown to influence human participants' attention to objects even when they were unaware of the gaze shifts (Mutlu, Yamaoka, Kanda, Ishiguro, & Hagita, 2009). This finding provides some evidence that robots could be used to manipulate others' attention. The use of social gaze is widely studied in social robotics, as it a key component in regulating human-robot interactions. For example, robots can use direct gaze to initiate a conversation, use gaze aversion to facilitate turn-taking, and combine gaze with gestures and speech (for a review, see Admoni & Scassellati, 2017). Their repertoire can include intentionally communicative behaviors that are designed to evoke a feeling of shared attention in their human interaction partners, such as declarative pointing and vocalizations ("Look!"; Anzalone et al., 2014). Therefore, the behavioral prerequisites, at least, for the shared attention level seem satisfied. A main challenge for the development of future social robots is to work on the ability to acquire common, mutual, and shared knowledge, to interpret others' ambiguous verbal or subtle nonverbal cues based on common ground in more complex environments, and to track shared experiences and integrate them into future interactions with others (for a review, see Yang et al., 2018). Another challenge is to identify what types of cues social robots and avatars need to produce in order to create at least an illusion of a second-person, engaged relation that humans will react to naturally (see, e.g., Anzalone, Boucenna, Ivaldi, & Chetouani, 2015; Breazeal & Scassellati, 2002; Castro-González, Admoni, & Scassellati, 2016; Lee, Breazeal, & DeSteno, 2017; Leite et al., 2017; Salem, Kopp, Wachsmuth, Rohlfing, & Joublin, 2012).

2.5.2 Application of this typology to joint goals and other joint mental states

We focus on attention and knowledge states here; however, the same typology can be extended to other mental states that can be either individual or joint (such as emotions, goals, desires, and beliefs) as well. Here we briefly sketch an idea of how the levels could apply to them, using different types of goals as an example. To achieve an *individual goal*, one intentionally organizes one's own behavior to reach that goal. For example, while reading the

newspaper, Maria finds that she is strongly opposed to a law that has recently been passed. Her goal is to convince the president to veto the law. Maria stands in front of the president's residence, and holds up a sign protesting against the controversial law.

One can also *monitor* another's goal, that is, infer someone else's goal, adopt it as one's own, and intentionally organize one's behavior to reach it. For example, Maria overhears a colleague mentioning holding up a sign in front of the president's residence. She adopts this goal, knowing that her colleague has the same goal.

In contrast, two or more individuals can have the same goal and know this together. Again, we propose that it is useful to distinguish several types of such goals based on the ways the goals are created: via inference or via direct experience with or without communication. Consider this example of a *common goal*: Maria stands in front of the president's residence with her sign. She sees that there are already several other people facing her holding similar signs on the other side of the residence, but she cannot go there as the police have blocked the streets. Maria realizes that they have the same goal, and that each of them will assume that they have a common goal with her. In a third-person situation, individuals can have a common goal by making inferences about the situation and engaging in recursive reasoning. This can enable them to coordinate their actions towards their common goal, even in situations in which no direct contact is possible.

To illustrate a *mutual goal*: Maria approaches the president's residence with her sign, and she makes eye contact with the protesters already standing there. They quickly gaze away as they need to be vigilant about the police's action. This mutual contact directly manifests to each of them that they realize that they have a mutual goal to convince the president to veto the law. Two individuals can have a mutual goal when they become directly and reciprocally aware, via attention contact in a second-person relation but without intentional communication, of their same goal.

Once Maria talks to one of the protesters, "Let's do this, veto the law!", and he shouts back, "Yes!", they establish a *shared goal*. Two individuals can establish a shared goal when they are in a second-person relation to each other,

while they are engaged in bidirectional intentional communication about the goal.

In contrast to individual and monitored goals, common, mutual, and shared goals can give rise to additional, social obligations beyond one's own individual feelings of obligations to achieve the goal. In common goals these additional obligations come from the fact that each of the individuals knows about their knowledge of the goal and thus plausible deniability about the goal is lower. The minimal openness in the mutual goal level makes the obligation to achieve the goal feel stronger than in the common goal level, and the communication in the shared goal level goes beyond this to enable both individuals to acknowledge the shared goal and their shared knowledge of it, thus not only highlighting their individual obligation to achieve the goal, but also creating interpersonal, joint commitments to each other about the joint goal (Siposova et al., 2018). That is, as soon as it turns into any type of joint goal, by definition (Bratman, 1999; Gilbert, 1990) there is some joint commitment to it – and the feeling of commitment is on a continuum just like with social attention and social knowledge. This feeling of joint commitment supports the achievement of joint goals, as it reduces uncertainty and we put more effort into fulfilling joint commitments (Michael & Pacherie, 2015).

Extending the typology to other mental states such as emotions, goals, desires, and beliefs thus allows us to use consistent terminology across different mental states. It also accounts for both the observational (third person) and the experiential (second person) ways in which we understand and share the world with others.

2.5.3 Other possible levels

One final direction for future research and thinking is to consider whether, in addition to the levels described above, there might be further levels of social attention and knowledge, for example, a cultural or collective level in which we pay attention to the attention not only of our immediate social partners but also of the group as a whole.

2.6 Part I conclusion

We argue here that we will be able to address the question of what joint attention is and how the jointness in it is achieved more productively if we distinguish several levels of social attention. We offer a systematic framework of social attention and knowledge levels in which we distinguish between two types of relations: reasoning about others as an outside observer from a third-person perspective versus directly engaging with them within a second-person perspective. A key part of our proposal is the claim that we are not jointly attending to something with someone until we enter into a second-person relation with them. This usually involves communication (including simple communicative looks)—our shared attention level—but it does not always: Neutral, non-communicative eye contact (even accidental eye contact, if it is registered) can create a bidirectional connection (in our mutual attention level). This proposal contributes to both the psychological and the philosophical literature by suggesting that social attention is not a single process, but rather it consists of a cluster of different cognitive skills. This proposal also emphasizes the link between the different attention levels and their corresponding knowledge states and points to a need to study them together. We have pointed out that not only do the levels themselves lie on a scale of jointness, but there is a continuum of jointness within each of the social levels as well. This, as well as the fact that the levels can be nested hierarchically within a single episode, illustrates that social attention is far more complex than previously thought. These levels allow us to clearly see the gaps in our knowledge about social attention in different populations. It is also useful to distinguish these levels because of the different consequences they have in terms of what kinds of interactions and obligations they can support. First, monitoring attention itself is already an advanced social-cognitive skill, which gives advantages to the individuals capable of it, for example in making informed predictions about others' future actions. Second, the common, mutual, and shared levels can support obligations and even commitments, each level with varying degrees of strength. Additionally, each of these levels presents a distinct way to solve social coordination dilemmas in different types of circumstances.

Social attention is an important topic in many disciplines, ranging from philosophy, to developmental, comparative, and social psychology, to psychopathology, to social robotics. To be able to compare new findings and exchange knowledge across these diverse disciplines effectively, we need to be sure that we are talking about the same psychological processes. With this in mind, we propose that carefully distinguishing between social attention and knowledge levels helps us to study this intriguing topic better.

Although all the attention levels have the potential to facilitate coordination and cooperation, as discussed above, the shared attention level has the special importance in this. It allows partners to align actively their mental states about the co-attended object, and therefore create specific expectations about their partner's future behaviors. The bidirectional communication in shared attention makes public the fact that attention is shared, which allows partners to refer to the instance of sharing afterwards. It is difficult to plausibly deny any obligations following from this. Thus, in the right context, attention sharing can lead not only to an expectation that one's partner will cooperate, but even to the perception of a commitment. Part II examines the role of attention sharing in coordination.

3. Part II: The role of attention sharing in coordination

Successful cooperation enables individuals to achieve greater goals than would be possible on their own. However, entering into cooperative interactions carries risks. The classic example is that if many people participate in a public demonstration, they can create change, but if only a few people participate, their effort may be wasted and they may be put in danger. As a potential demonstrator, before one takes the risk of showing up to participate, one needs to judge whether others will participate too. To reduce uncertainty about others' behavior in cooperative interactions, communication and commitments are particularly useful tools (Michael & Pacherie, 2015). It has been shown that exchanging verbal commitments substantially increases successful cooperation in social dilemmas among adults (for a meta-analysis, see Sally, 1995).

Typically, commitments arise verbally through speech acts of promising or making agreements. For example, Peter can commit himself to washing the dishes after dinner by promising his wife he will do so. He then has an obligation to wash the dishes and his wife has the right to protest if he does not. Although different definitions of commitments exist, most involve this general formulation: If one social partner intentionally communicates to another that he intends to do X, and the other acknowledges this, then they have knowledge in common about this interaction, and the first partner is committed to do X (e.g., Austin, 1975; Searle, 1969; Scanlon, 1998).

There is considerable debate among philosophers about how one key type of commitments, promises, function. The main point of contention is the explanation of how promissory obligations come into existence. Conventionalist theories argue that promising is a social practice involving convention, and that only certain verbal statements (e.g., "I promise to do X" or "I will do X") or conventional acts (e.g., nodding) under the right circumstances will create promissory obligations (Hume, 1739–1740/1969; Kolodny & Wallace, 2003; Rawls, 1955; Searle, 1969). In contrast, most contemporary accounts reject the idea that the core of promises is rooted in social convention (Gilbert, 2004; Owens, 2006; Scanlon, 1998; Shiffrin, 2008). For example, Scanlon (1998)

argues that whenever one individual intentionally leads another to expect that he will do X (and knows that the other wants to be assured of X), he is committed to do X, as the general moral principle not to mislead others is in place. Similarly, MacCormick (MacCormick & Raz, 1972) has stressed the role of reliance: If one individual has intentionally induced another to rely on him, then he is committed to follow through. This is especially evident when the other individual takes detrimental action for herself based on her expectations of the first individual's behavior. Somewhat differently, in a discussion of joint commitments, Gilbert (1990, 2004, 2014) claims that promissory obligations are not necessarily moral obligations, but are sui generis form of normativity. In her view, commitments do not need to be construed verbally; they are created by each individual expressing readiness to be jointly committed under conditions of common knowledge.

Therefore, although perhaps the easiest way to create promissory obligations is to state, "I promise to X," these latter accounts suggest that commitments and promises could arise even without any words at all. It has been suggested that it would be useful to move away from the binary distinction between full-fledged verbal, explicit promises vs. not promises, and elucidate the full spectrum of promissory obligations (Shiffrin, 2008). Part of the debate is about what type of communication of an intention to bind oneself to do X is necessary to form promissory obligations (Gilbert, 2014; Owens, 2006; Scanlon, 1998; Shiffrin, 2008). To inform the theory in this area, we investigated empirically whether it is possible to commit oneself nonverbally, as long as intentional communication is involved (Raz, 1977).

To our knowledge, these ideas have never been tested empirically. We hypothesized that a communicative look is an especially good candidate for a minimal, nonverbal signal that might be powerful enough not only to promote expectations of cooperation, but also to create a commitment. This type of eye contact is ostensive (Sperber & Wilson, 1986), bidirectional, public, and enables one to communicate a message under conditions of shared knowledge (Carpenter & Liebal, 2011), as long as the common ground is strong enough (Tomasello, 2008). Thus, within the context of a cooperative coordination

problem, here we test whether communicative, versus non-communicative, eye contact can signal a commitment to cooperate.

We designed a game based on the Stag Hunt dilemma, an ideal model for studying social dilemmas in mutualistic contexts (Rousseau, 1754/1984; Skyrms, 2004). As explained in the introduction, in the Stag Hunt parable, two hunters decide either to each hunt a hare (a certain but small prize) individually or to hunt a stag (a risky but big prize) together, if it is available. However, if just one partner decides to cooperate and hunt the stag and the other decides to hunt a hare (e.g., because he does not know the stag is available), the cooperating partner loses the chance to get anything. Thus, successful cooperation in this context faces two main challenges: First, to reduce uncertainty about the partner's knowledge (here, about the presence of the stag), and second, to reduce uncertainty about the partner's behavior (whether he will cooperate).

The first challenge is epistemological. It is not enough if each social partner knows individually that the cooperative option is available. Instead, successful coordination is facilitated by knowledge in common (Chwe, 2001; Clark, 1996), that is, each partner needs to know that the other knows about the cooperative option, that the other knows he knows, etc., ad infinitum (e.g., Lewis, 1969). However, as the processing demands for even just a few levels of such recursive reasoning are high, it is likely that we use simpler shortcuts such as communication for creating knowledge in common (Clark, 1996). It has been shown that adults (Brosnan, Wilson, & Beran, 2012) and 4-year-old children (Duguid et al., 2014) spontaneously use verbal communication to achieve knowledge in common and solve the Stag Hunt dilemma successfully. Some authors have proposed that knowledge in common can be created by nonverbal signals like eye contact (Carpenter & Liebal, 2011; Chwe, 2001; Gómez, 1996; Thomas et al., 2014); however, empirical evidence for this in children or adults is scare. As mentioned in the introduction, Wyman et al., (2013) studied the ability of a communicative look with a smile to create knowledge in common about the presence of the cooperative option in a Stag Hunt game. In this study, 4-year-old children were discouraged from communicating verbally while playing this game with an experimenter. In the control condition, when the

cooperative option appeared, the experimenter monitored it – it was clear that she saw it – but she did not look at children, whereas in the experimental condition, she alternated gaze ostensively between the cooperative option and children's eyes while smiling. More children decided to cooperate in the experimental than the control condition, suggesting that this minimal nonverbal behavior established shared knowledge about the availability of the cooperative option.

Wyman et al. (2013) thus provide the first hint that nonverbal signals can serve to create shared knowledge in a collaborative decision-making situation. However, many open questions remain about what exactly caused children to behave differently in the two conditions of that study. For example, did the eye contact alone generate knowledge in common or was it the communication inherent in the look and/or smile? It has been argued that one cannot truly share attention to something, or know something together with one's partner, without some form of communication – even if just a communicative look (Carpenter & Liebal, 2011; Siposova & Carpenter, submitted). Thus in the current studies, our first aim was to test the hypothesis that communicative looks (but not non-communicative looks) help establish shared knowledge about the cooperative option, and therefore lead children to expect cooperation from their partner and thus to decide to risk cooperation.

The second challenge for cooperation is to reduce uncertainty about one's partner's behavior, and it has been suggested that commitments and promises are a key way of stabilizing cooperative behavior (Michael & Pacherie, 2015). Children begin to engage in collaborative activities with complementary roles and joint goals around two years of age (Brownell, 2011), but a basic understanding of commitments and at least some of the resulting obligations develops somewhat later, by the age of three – not coincidentally, at around the same age that children begin to show an understanding of social norms regarding moral transgressions (e.g., Smetana & Braeges, 1990; Vaish, Missana, & Tomasello, 2011), fairness (Smith, Blake, & Harris, 2013), and simple game rules (Rakoczy, Warneken, & Tomasello, 2008). For example, after verbally making a joint commitment to cooperate, 3-year-olds understand some of the obligations that both they and their partner have to keep playing (Gräfenhain,

Behne, Carpenter, & Tomasello, 2009) and protest when their partner intentionally defects (Kachel, Svetlova, & Tomasello, 2017). Preschoolers also understand promissory obligations: They tend to keep their own promises and refer to the promise that was made when their partners do not keep their promises (Heyman, Fu, Lin, Qian, & Lee, 2015; Kanngiesser, Köymen, & Tomasello, 2017). To our knowledge, all studies with children and adults investigating commitments focus on commitments made verbally. Therefore, our second aim was to investigate whether young children can understand even minimal nonverbal communicative signals as commitments to cooperate.

3.1 Study 1

In Study 1 we investigated to what extent a communicative, versus a non-communicative, look can produce an expectation of collaboration. While playing a novel version of the Stag Hunt game, 5-year-old children needed to decide whether to risk cooperating or take the safe option and play individually. At the critical moment, children saw that the cooperative option was available, but – unlike in the Wyman et al. (2013) study – they were not sure if the experimenter could see it as well. Thus, children were not able to assume knowledge in common about the presence of the cooperative option. Immediately before children needed to make their decision, depending on the condition, the experimenter either silently shared attention with them—she made ostensive, communicative eye contact—or looked non-communicatively at them. We predicted that only communicative looks would establish knowledge in common and promote cooperative decisions. We also investigated whether attention sharing via the communicative look could be seen by children as a commitment on the experimenter's part to collaborate. To do this, in both conditions the experimenter ended up playing for the individual option, and we gave children the chance to protest as a sign that they understood that their partner had broken a commitment to collaborate.

3.1.1 Method

3.1.1.1 Participants

Seventy-two 5-year-olds (36 girls, mean age: 5;2; range: 4;11-5;6) were included in the final sample. Additional children began testing but did not complete the training due to language difficulties (2) or failing the pre-tests (11). Other children were tested but excluded from analyses for not seeing the manipulation (5), disobeying instructions and communicating their decision to the experimenter (13), apparatus failure (3), or experimenter error (10). We recognize that this drop-out rate is higher than usual, and we attribute this to the complexity of the procedure and children's difficulty at inhibiting communication with the experimenter during the response phase (for more justification about why excluding these children was necessary, see below). Children's parents had given consent for them to participate in developmental studies. They were tested individually in their kindergartens in Leipzig. The university's ethics committee approved the study.

3.1.1.2 Materials and design

The apparatus for the game consisted of four tubes mounted at a downward angle towards a platform with boxes on it (see Figure 4). Two white tubes led to two white paper boxes representing the small prizes (i.e., the 'hares', the safe, individual option). Two colorful tubes led to one colorful wooden box representing the big prize (i.e., the 'stag', the risky, collaborative option). Children's partner was played by a puppet (controlled by an experimenter [E1]), because previous studies have shown that children are more likely to protest about the misbehavior of a puppet than that of an adult (e.g., Rakoczy et al., 2008). Behind the boxes sat another experimenter (E2) who operated the platform with boxes. The white boxes were always available, and the colorful box only occasionally also appeared in the middle. Children and the puppet could win the prizes by putting a marble into one of their two tubes, such that it rolled down the tube and knocked down the prize. There were small pegs in the tubes to prevent the marbles from rolling down the tubes until E2 gave permission for them to be released. Paper pictures of birds (3x3

cm) were used as the small prizes and, during training, nice stickers (3x3 cm) were used as the big prizes. Before the test phase, these nice stickers were replaced with 'special' stickers (4x4 cm) as the big prizes. Between children and the puppet there was a low barrier, so neither could see what the other chose, but they could both see each other's face. Another small occluder partially blocked the puppet's view of the colorful box, thus leaving children uncertain about the puppet's knowledge regarding the appearance of the big prize.

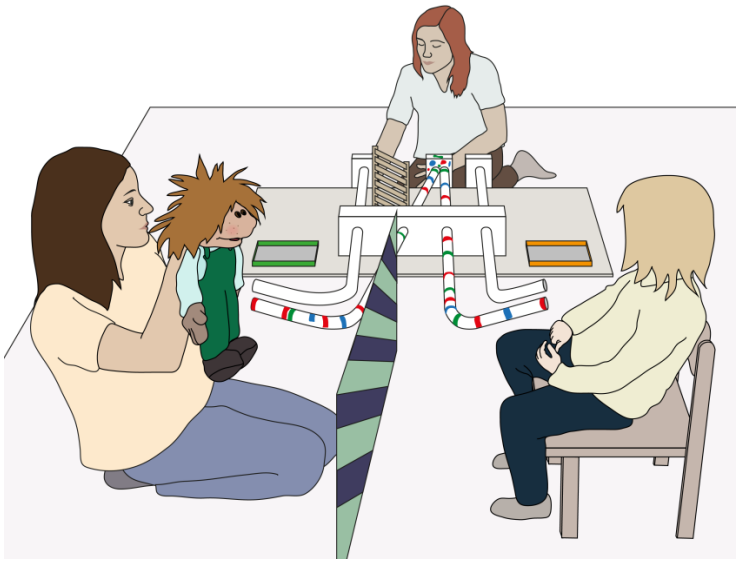

Figure 4. The apparatus with four tubes, two small prizes (the white boxes), one big prize (the colorful box), a partial occluder in front of the big prize on the puppet's side, and a barrier between the puppet and the child.

3.1.1.3 Procedure

Training: Children were picked up from their classroom by E1, who introduced herself and the puppet. As they entered the testing room, E2 introduced herself and a small stuffed animal – a cat from a popular children's cartoon, and familiarized the child with a smiley scale that would be used at the end of the procedure. Then children played a warm-up game with the puppet. The puppet performed three actions incorrectly (e.g., wrote with the wrong end of a pen) and children were encouraged by E2 to correct her.

Then E2 introduced the main game to children and the puppet. She showed children and the puppet that they could win prizes by putting a marble into one of two tubes to knock down one of two boxes. E2 introduced the partial occluder, and the puppet said that sometimes she could see the big prize and sometimes not but that she could always see the small prize. During training, children received four types of trials: In the trials in which the *puppet can see the big prize*, she said, "Look, the colorful box! Let's play together, [child's name], so we can get a colorful sticker." In the trials in which she *could not see the big prize*, she said aloud to herself, "I can just see the white box. So I'll play alone to knock down the white box," and she put her marble in the tube for the white box quickly so she could not change her decision if children corrected her. In the *only small prizes* trials, the puppet said, "Look, [child's name], only white boxes." In a single *child alone* trial, the puppet excused herself for a moment, and children played the fourth trial alone. This trial served as a first pre-test to ensure that children understood that it was not possible to get the big prize alone. To pass this pre-test, they needed to play for the small prize, even though the big prize appeared as well. If children decided to play for the big prize, E2 explained the rules of the game again, started training from the beginning, and children were given one additional chance to pass this pre-test. The training trials were presented in the following order: *puppet can see the big prize, puppet cannot see the big prize, only small prizes, child alone, only small prizes, puppet cannot see the big prize, puppet can see the big prize, only small prizes*. Of these, together the *puppet can* and *cannot see the big prize* trials served as the second pre-test, ensuring that children were motivated to play for the big prize when the puppet was likely to play for it but did not play for it when the puppet could not see it. If children did not play appropriately in any given trial, additional feedback was provided, and the trial was repeated once. To pass this pre-test, children had to play appropriately on the first or second attempt of each of these types of trials. Children who failed any of the pre-tests played one more cooperative trial, received a special sticker, and were thanked for playing so nicely and escorted back to class. For children who passed the pre-tests, before the last training trial, E2 showed children and the puppet a new, special sticker that would serve as the next big prize. Then E2

introduced a no-talking rule to prevent children from verbal communication about their decisions. E2 pretended that the cat wanted to sleep and said, "From now on, we have to be very quiet, and you can't talk to each other anymore so [cat's name] can sleep. This is a very important rule, okay? But don't worry; you can still see each other's faces so you will still be able to keep playing." Then a final small prize trial was conducted in order to make sure children would follow the no-talking rule.

Test phase: There was just one test trial. Children and the puppet watched while the platform with both the small and the big prizes on it appeared. Children could see the big prize but were not sure if the puppet could see it too. The puppet turned so she was facing children and, because E1's eyes were more expressive than the puppet's, E1 performed two looking sequences with her face next to and moving in tandem with that of the puppet. To be included in the analyses, children had to make eye contact with E1 at least once during the manipulation.

In the communicative look condition, E1 shared attention with children: she made eye contact with them with an ostensive-communicative look. She opened her eyes widely and raised her eyebrows (see Figure 5a). This look potentially communicated the message, "I can see the big prize, let's cooperate," thus establishing shared knowledge about the availability of the cooperative option. Then E1 looked back to the prizes. E2 gave a marble to each of them, and looked down. E1 repeated the communicative look, looked down to the openings of the tubes ambiguously (so children could not use her gaze direction to figure out how she would play), and the puppet put her marble into a tube, out of sight of children.

In the non-communicative look condition, each of the two looking sequences was split into two parts, with two shorter looks instead of one long one, since longer looks were judged as communicative by adults during informal piloting. After the prizes appeared, E1 (with a neutral, relaxed face) slowly turned her head in the direction of the child, made non-communicative eye contact with the child (i.e., neutrally, without raised eyebrows; see Figure 5b), and then continued on, looking past children's eyes and slightly to the side (approximately 30 cm from children's eyes). E1 then followed a similar path

back to looking at the prizes, making eye contact a second time while doing so. After taking the marble, E1 repeated the same gaze path, but after passing children's eyes at the end, E1 looked just slightly to the side (approximately 10 cm from children's eyes). Then E1 looked down and the puppet played exactly as in the other condition. We included the looks to the side to make the eye contact seem less communicative.[6]

Children then needed to decide whether to play individually or cooperatively. In both conditions the puppet played for the individual option. Then E2 informed children that they were allowed to talk again and they could remove the pegs to release the marbles. At this point, E1 went back to holding the puppet in front of her face so that anything children said next would be directed to the puppet. Children who played individually received the small prize, and went directly to the questioning phase (see below). Children who played for the big prize saw that their marble did not knock down the big prize and that the puppet, who had played individually, knocked down the small prize. The puppet and E2 then provided increasingly specific cues to elicit a response from children. First, E2 said, "Oh," in a slightly surprised way. After 10 s, E2 asked, "What happened?", alternating gaze between children and the puppet. After 10 more s, the puppet said, "You, [child's name], tried to knock down the colorful box, but I knocked down the white box." After 10 more s, the puppet asked, "And what do you say about this?" E2 gave children an additional 10 s to respond, and then the puppet excused herself and left the room.

[6] We noticed, belatedly, that, following a holiday break in testing, there were some errors in E1's looks: In the communicative look condition, she began smiling slightly and in the non-communicative look condition she sometimes neglected to look away to the side after looking at children's eyes. It was necessary to drop these children because different information was provided to them, and the non-communicative looks seemed more communicative. We assessed interrater reliability on this coding decision by having an independent, naive coder code ten cases that the authors determined to be experimenter errors and ten other, randomly-selected cases in matching conditions. There was good agreement (Cohen's κ=0.75), and the few disagreements between the naive coder and the authors about the drop outs were resolved by discussion.

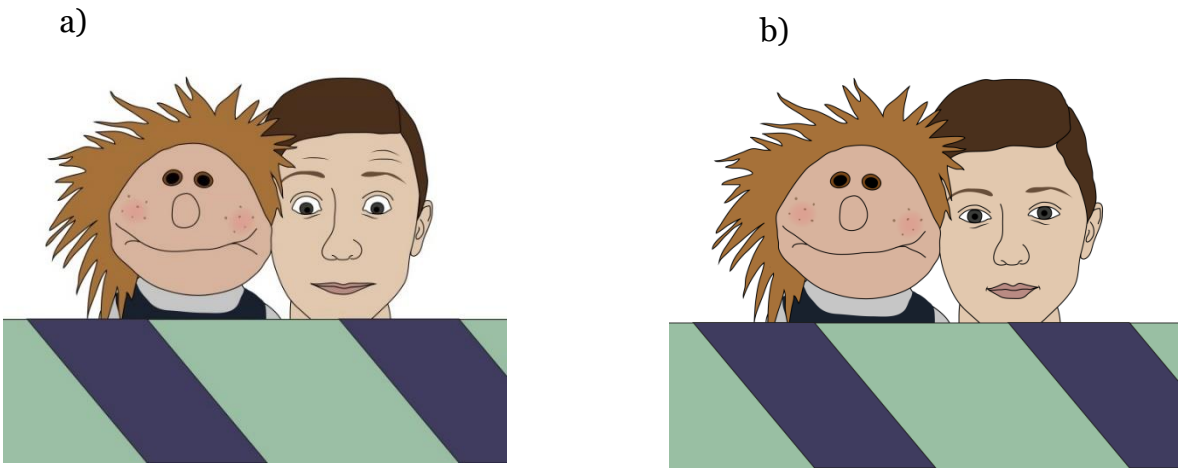

Figure 5. A depiction of a) the communicative look and b) the non-communicative look.

Questioning phase and debriefing: To gain more insight into children's thinking, we asked children several exploratory questions about, for example, their expectations and feelings for the puppet (see Result section 3.1.2.4 for these questions). Then E2 told children that they had played well and gave them the special sticker from the test trial. E1 apologized for playing incorrectly, and thanked children for playing.

3.1.1.4 Coding and reliability

The main coding was done from the video-recordings. To assess inter-rater reliability, an independent coder who was naive to the hypotheses coded a random sample of 25% of the children in each condition.

Cooperation vs. individual play: The main measure was whether children decided to cooperate. During the test trial, children's behavior was coded as either cooperation (if they put their marble into the colorful tube leading to the big prize) or individual play (if they put their marble into the white tube leading to the small prize). There was perfect inter-rater agreement on this measure.

Normative protest: We predicted that if children understood the communicative look as a commitment to cooperate, then they should protest

more in the communicative look condition than in the non-communicative look condition. However, it was important to ensure that if this were the case, it was not simply because children were less successful (and thus more upset) in the communicative look condition, i.e., because they decided to cooperate more in this condition and so did not receive a prize. Thus for this and the subsequent measures exploring children's responses to their partner's decision to play individually, all responses from the test phase were transcribed and coded only for the subset of children in both conditions who had decided to cooperate. In that way, children's decision, and, importantly, their experience of not receiving a prize, were identical. Normative protest was coded when children used normative vocabulary (e.g., "should," "must," "wrong") while referring to how the puppet played; for example, "You should knock down the colorful box!" There was perfect inter-rater agreement.

Children's verbal responses to their partner's decision to play individually: To further examine how strong children's expectation was that their partner would cooperate, for the subset of children who decided to cooperate, along with normative protests, verbal utterances during the test phase were additionally coded for the following exhaustive but non-exclusive responses: (a) Explicit expectation of cooperation: children said they expected that the puppet would play for the big prize, for example, "I thought that you knocked down the colorful one." (b) Negative evaluation of the puppet: Children expressed frustration directed to the puppet, for example, "I think this is a little stupid." (c) Other: Children did not produce a response falling into either of the above categories (e.g., stayed silent or just described what happened). There was perfect inter-rater agreement.

Children's nonverbal responses to their partner's decision to play individually: In addition to verbal statements indicating that children had expected the puppet to collaborate, for the same subset of children we also coded nonverbal signs of this: displacement behaviors, disappointment, and/or anger during the response period. Behaviors were mutually exclusive and the exact frequency for each behavior that occurred during the response period was recorded. We predicted that if children had stronger expectations about their partner's cooperation in the communicative look condition, they should show

more nonverbal signs that their expectations were violated in this condition. We used the Ethological Coding System for Interviews to code displacement behaviors (e.g., touching face, scratching, licking lips; Troisi, 2002). Displacement behaviors are self-directed activities that occur in situations characterized by stress and/or social tension, and which correlate with self-reported anxiety and negative affect (Troisi, 2002). They are thus a useful measure of violation of social expectations. We also coded for signs of disappointment (e.g., frown, disappointed voice, slumped body posture) and anger (e.g., clenched lips, annoyed voice or look, angry body posture, e.g., hands-on-hips gesture or raising fist at the puppet; adapted from Cole, Zahn-Waxler, & Smith, 1994). Reliability was excellent for the number of both displacement behaviors (r_s=0.92, with no difference between coders, Wilcoxon signed rank test: p=0.31) and expressions of disappointment (r_s=0.83, with no difference between coders, Wilcoxon signed rank test: p=1). No reliability coding for anger was conducted because only one such behavior was coded, and therefore anger was not included in the analysis. For statistical analyses, a composite score for nonverbal signs of violation of expectation of cooperation was calculated by summing the number of displacement and disappointment behaviors together (each behavior was equally-weighted).

Children's contribution to the mutual look in the communicative look condition: We also studied children's own contributions to the communicative look by investigating whether children contributed to the mutual gaze with a communicative behavior of their own. If so, by doing this, children would explicitly acknowledge their partner's commitment and/or even make a commitment back to her, thereby creating a joint commitment. We predicted that children who responded to their partner's communicative looks in this way would be even more likely to protest than children who did not. Children's reactions to the communicative look were coded as: raising their eyebrows, nodding (after receiving the look from E1), or no communicative behaviors. There was perfect inter-rater agreement on this measure. For the analyses, a binary outcome variable – whether or not each child produced a communicative behavior – was used.

Supplementary post-test questions: For exploratory purposes, to gain more information about children's behaviors, we asked children a series of questions at the end of the procedure. Reliability revealed excellent agreement for all measures (Cohen's κ's ranged from 0.90-1).

3.1.2 Results

3.1.2.1 Cooperation vs. individual play

Children in the communicative look condition overwhelmingly decided for cooperation: 78% of them played cooperatively (28 children cooperated, 8 played individually, binomial probability, p= 0.001). In contrast, in the non-communicative look condition, only 47% of children played cooperatively (17 children cooperated, 19 played individually). Children in the communicative look condition cooperated significantly more often than children in the non-communicative look condition (Fisher's exact test, p=0.01; odds ratio=3.83).

3.1.2.2 Protests and other responses to the partner's decision to play individually

Only one child in each condition normatively protested after finding out that the puppet did not cooperate. Similarly, only a few children verbally expressed expectations of cooperation and/or negative evaluations of the puppet (see Table 2), thus no analyses were performed on these measures. However, for nonverbal signs of violation of expectation, there was a significant effect of condition: Children showed more nonverbal signs of violation of expectation in the communicative look condition (M=3.75, range=0-10, SD=2.50) than in the non-communicative look condition (M=1.65, range=0-4, SD=1.54, Welch's t-test: t=3.29, df=40, p=0.002, n_1=28, n_2=17).

Table 2

Percentage of Children who Showed the Different Verbal Response Types During the Test Phase in Study 1.

Response types	Communicative look condition	Non-communicative look condition
Normative protests	4%	6%
Explicit expectations of cooperation	14%	0%
Negative evaluations of the puppet	14%	12%
Other	71%	88%

Note: Only the subset of children who played cooperatively are included (N=28 in the communicative look condition, N=17 in the non-communicative look condition).

3.1.2.3 Children's contribution to the mutual look in the communicative look condition

Twenty-five percent of children in the communicative look condition showed some communicative signs (8 children raised their eyebrows and 1 raised the eyebrows and nodded) during the mutual look. As only one child normatively protested, no analyses were performed to investigate whether children who responded with communicative behaviors to E1 were more likely to protest than children who did not. The one child who protested did raise his eyebrows during the look.

3.1.2.4 Supplementary results for the post-test questions

At the beginning of the interview phase, E2 asked the child, "How do you feel right now?" and she pointed to and asked about each choice on the 5-point smiley scale (ranging from 1 "Do you feel very sad?" to 5 "Do you feel very happy?"). With this question, we wanted to investigate differences in mood between conditions within the subset of children who cooperated. Stronger expectations in the communicative look condition could lead to stronger

disappointment when the puppet did not cooperate, which in turn could be reflected in children's mood. However, most children felt quite happy and there was no difference between conditions in children's mood (communicative look condition: M=4.14, SD=1.41; non-communicative look condition: M=4.47, SD=0.87; Welch's t-test: t=-0.49; df=38, p=0.63, n_1=28, n_2=17).

Then E2 continued, "Think back to the last round of the game when the boxes appeared, but before you put your marble into the tube. In that moment, what did you first think: In which tube would [puppet's name] put her marble?" We coded whether children expected: (a) Cooperation: Children mentioned, pointed at, or touched the colorful tube, or (b) Individual play: Children mentioned, pointed to, or touched the white tube, (c) Do not know: Children explicitly said they did not know or shrugged their shoulders, or (d) Other: Children said something that did not fit into any of these categories or stayed silent. (Note that the definitions for (c) and (d) are the same for all questions below.) In the communicative look condition, 61% of children expected cooperation, 33% expected individual play and 6% provided other answers. In the non-communicative look condition, 47% of children expected cooperation, 50% expected individual play and 3% answered that they did not know. To investigate whether children's responses differed between conditions, we analyzed whether in the communicative look condition more children answered that they expected cooperation. Three children did not answer that they expected either cooperation or individual play and were therefore excluded from this and the next analysis. There was no significant difference between conditions ($\chi^{2(1)}$=1.83, p=0.17). One reason for this might be that we asked the post-test questions after children had seen that the puppet had played individually. Therefore, some children might have adjusted their answers based on the puppet's actual behavior. We also looked at the whole sample to see whether children's decision in the test trial was driven by their expectations about what the puppet would do. We analyzed whether children who expected that the puppet would cooperate cooperated more themselves. Across conditions the results showed that children who expected the puppet to play cooperatively cooperated significantly more (74%) than children who expected the puppet to play individually (26%, $\chi^{2(1)}$=14.87, p<0.001). These results

support the idea that children based their decision to cooperate or not on their expectations about their partner's behavior.

To investigate the reasons why children played in a certain way, E2 asked the next two questions: "Why did you expect her to do that?" and "And why did you put your marble into the white/colorful tube in the last round?" Both were coded for indications that children thought that their partner would play either cooperatively or individually: (a) I thought she saw/wanted/would play for the colorful box: Children said they thought that the puppet had seen the colorful box or that she would like to cooperate or to play together. (b) I thought she did not see/did not want/would not play for the colorful box: Children said the puppet had not seen the colorful box or she would like to play individually or put her marble into the white tube, (c) Do not know, or (d) Other. Overall, the majority of children (79% for the first of these two questions and 67% for second one) could not answer these two questions or provided unrelated explanations, thus no further analysis on them was conducted.

Then, E2 asked, "How much do you like [puppet's name]?" and again showed and stepped through the smiley scale. The scores ranged from 1 ("Do you not like her at all?") to 5 ("Do you like her very much?") With this question, we wanted to examine differences in children's liking of the puppet within the subset of children who cooperated. Stronger expectations of cooperation in the communicative look condition might result in these children liking the puppet less than children in the non-communicative look condition. However, we did not find a significant difference between conditions (Welch's *t*-test: t=-0.30; df=33; p=0.77; n_1=28; n_2=16). Most children liked the puppet at least a bit (communicative look condition: M =4.14, SD=1.01; non-communicative look condition: M=4.25, SD=0.93.) One explanation for why we did not find a difference in children's mood (in the first question) and their liking of the puppet is that these questions were not asked immediately after children found out that the puppet did not cooperate, and that any initial differences might have diminished after a short period of time.

After these questions, E2 told children that they had played well and gave them the special sticker from the test trial. Then, E2 asked them again how

they felt (to make sure that no children felt bad before going back to class). Children generally felt either a little happy or very happy (M=4.65, SD=0.75).

Then, E1 (without the puppet) came back into the room, and after E2 made all the prizes appear on the platform, E1 enacted for children the communicative and the non-communicative looks in turn, in counterbalanced order. While doing so she asked, "When I look at you like this, what do you think this look means?" If children did not provide a relevant answer spontaneously (i.e., mentioning her ability, or lack of ability, to see the big prize box, or mentioning cooperation or individual play), she continued and asked, "Do you think I can see the colorful box when you see this look?" Then, she enacted the other look and asked the same questions. For both of these questions together, for each look separately, we coded for the following responses: (a) Yes/E1 can see the big prize: Children mentioned the colorful box or referred to the collaborative choice in some other way or answered yes to the second question, or (b) No/E1 cannot see the big prize: Children mentioned the white box or the individual choice in some other way or answered no to the second question. We found that after seeing the communicative look, 80% of children answered correctly that E1 could see the big prize, and after seeing the non-communicative look 68% of children answered correctly that E1 could not see the big prize (exact binomial tests revealed that these were both significantly better than chance, $p<0.001$ and $p=0.003$, respectively). This result further supports the idea that children recognized different meanings in these two types of eye contact and interpreted them correctly.

Finally, we were interested in investigating whether children evaluated E1's play differently in the different conditions. Therefore, at the very end, E1 asked, "[Child's name], do you think I played incorrectly in the last round? Children's responses were coded as: (a) Yes (also when children said things like "a little" or "a bit"), (b) No, (c) Do not know, or (d) Other. The majority of children in both conditions answered that they did not think that E1 played incorrectly (communicative condition: 64%; non-communicative look condition: 72%) and there was no difference between conditions ($\chi^2(3)=2.38$, $p=0.50$). It is perhaps puzzling that most children in the communicative look condition answered that they did not think that E1 played incorrectly. Children

clearly understood that the communicative look suggested that E1 would play cooperatively so they easily could have interpreted E1's individual play as incorrect. We think it likely that children did indeed believe that E1 played incorrectly, but were simply unwilling to tell her this to her face. Another possibility is that children interpreted E1's decision as unfair, but not necessarily incorrect.

3.1.3 Discussion of Study 1

At the crucial moment in a coordination game, when children needed to decide whether to play cooperatively or individually, the type of eye contact children's partner made with them influenced the inferences they made about their partner's likely behavior. The main finding was that children chose to cooperate more often when their partner looked at them with a communicative, compared to a non-communicative, look. When deciding whether to risk cooperating, children had to consider two important pieces of information: first, whether their partner had seen that the cooperative option was available and, second, whether she would then decide to cooperate. Therefore, children could interpret her ostensive-communicative look as a signal that she had seen the cooperative option and/or as a signal that she was going to cooperate. It is difficult to distinguish between these two interpretations because both implicitly include each other. For example, it was enough for the partner to signal that she had seen the cooperative option, because both children and the puppet knew together that cooperation was better for both partners and thus it was reasonable to assume cooperation would follow as long as both partners knew together that the cooperative option was available.

In addition, we found that children in the communicative look condition showed more nonverbal signs of violation of expectation when their partner played individually than children in the non-communicative look condition. This suggests that, even when children did cooperate in the non-communicative look condition, they had weaker expectations about the likelihood of their partner's cooperation.

Around half of the children cooperated even in the non-communicative condition. This is comparable to the cooperation rate in the Wyman et al.

(2013) study, in which 48% of children cooperated in a condition with no eye contact. Note that behavior in a coordination game is dependent on both the probability of getting the prizes and the values of both types of prizes. The big prize was appealing and children were highly motivated to play for it, thus they were likely to err on the side of false positives. The chance-level cooperation rates in this condition suggest that either children were not sure what the non-communicative look meant and simply guessed what to do or half the children assumed their partner would play individually and half cooperatively.

Counter to our expectations, we found almost no normative protests about the partner's decision to not cooperate. Although at first most of the children interpreted the communicative look correctly as a signal to cooperate, they then saw that the puppet played individually. This might have led them to reinterpret their initial understanding of the communicative look. That is, as children have a general expectation that adults are competent (e.g., Taylor, Cartwright, & Bowden, 1991), and an adult was controlling the puppet, they might have concluded that they had mistakenly interpreted the meaning of the look, and, instead of blaming the puppet/adult, may have blamed themselves for the coordination's failure.

In this study we clearly demonstrated that after seeing the communicative look, children had a strong expectation that their partner would cooperate, and therefore decided to cooperate themselves. However, because children rarely protested in this, here we could not demonstrate that children understood the communicative look as a commitment on the looker's part to cooperate. We should note here that only the child who protested normatively in the communicative look condition referred implicitly to a communicated commitment: "But you said [child mimics the communicative look]. For sure you picked the wrong hole." This suggests that some children might understand the communicative look as a commitment to cooperate. In order to investigate this further, we conducted Study 2.

3.2 Study 2

Some theorists would argue that a look cannot serve as a promissory commitment, since it is not a conventional act (Hume, 1739-1740/1969; Rawls,

1955; Searle, 1969). However, if, as others argue, promising is not rooted in social convention (Gilbert, 2004; Owens, 2006; Scanlon, 1998; Shiffrin, 2008), it is feasible that under certain circumstances a communicative look could be sufficient for establishing a commitment, as it involves the promiser intentionally communicating in such a way as to induce the recipient to rely on her (MacCormick & Raz, 1972; Scanlon, 1998), and allows individuals to express readiness to commit under conditions of shared knowledge (Gilbert, 1990, 2004, 2014). To inform this theoretical discussion, we attempted again to find evidence that commitments can be created nonverbally. We manipulated the type of look children's partner gave them at the critical decision-making point in a coordination game and we focused more precisely on children's protests when they learned that their partner did not cooperate.

We modified the procedure in several ways to make protesting more likely to occur. First, we tested slightly older children: 6- to 7-year-olds. During the transition between preschool and early school age, children become more sensitive to the contexts in which norms apply, and their explicit reasoning about social norms becomes more advanced (Riggs & Young, 2016; Smith, Blake, & Harris, 2013). Therefore, we reasoned that slightly older children might feel more confident about protesting when their partner did not cooperate. Second, to make it less likely that children would blame themselves for the breakdown in coordination, we made the game easier and got rid of the occluder (so children were no longer uncertain about their partner's knowledge). Third, for children's partner we did not use a puppet, but rather an adult experimenter. To reduce children's inhibition about protesting towards adults, in a training phase, in both conditions, we encouraged children to protest when she behaved incorrectly. Finally, we made the cooperative option both more appealing and always available, so children in both conditions had a reason to expect cooperation and to cooperate. We did this because the rate of normative protesting in studies with young children is generally low, so to be able to detect any differences between conditions, we needed more children to cooperate (in order to have a chance to protest) in both conditions. Thus our aim in this study was not to measure children's cooperation rates and replicate Study 1; rather, it was to see whether children who cooperated would protest

more following a communicative look than a non-communicative look. We conservatively focused on just one specific type of protest: normative protest referring, directly or indirectly, to the commitment (e.g., “You should have cooperated, because you told me you would”). We predicted that children in the communicative look condition would produce more of this specific type of protest than children in the non-communicative look condition. In addition, to zero in even more precisely on the role of the communicative look, children were only able to see the eye area of their partner during the critical decision-making phase: At this point, the two partners looked at each other through a tube.

3.2.1 Method

3.2.1.1 Participants

Seventy-two 6- to 7-year-olds (36 girls, mean age: 6;11; range: 6;4-7;3) were included in the final sample. We focused on children’s protests about their partner’s failure to cooperate; therefore, the final sample in both conditions was planned in advance to include only children who decided to cooperate and who thus had a reason to protest. Consequently, additional 15 children who played individually were excluded for this reason (8 in the non-communicative look and 7 in the communicative look condition). Additional children were tested but not included in the analyses for refusing to play the game (1), language comprehension difficulties (1), failing the second pre-test (2), and verbally or gesturally telling the experimenter how they would play during the critical decision-making phase (8). Participant recruitment was the same as in Study 1, except that children were tested in their schools.

3.2.1.2 Materials

The materials for the warm-up game included five pairs of different colored mats (30x20 cm) made of a thin, soft, rubbery material, which were placed on the ground, and a pack of cards (6x4 cm), each of which displayed two colors on the same side. The materials for the main game included a blue rectangular rug (170 x 140 x 170 cm) decorated with fish representing a lake,

and two low, yellow platforms decorated with palm trees representing the beach (see Figure 6). On the rug, there was a treasure box (20x26 cm) containing seven small containers. Inside each container were the big prizes: two different stickers of animals (4x4 cm). Between the two platforms, there was a tube (13 cm diameter) attached to a tripod with two spy cameras inside to film children's and their partner's face. Small paper pictures of a bird (3x3 cm) were the small prizes.

Figure 6. The setup of Study 2: the tube, the lake representing the cooperative option, and the beach representing the individual option.

3.2.1.3 Procedure

Warm up: The goal of the game was for children and E1 to jump to the same color mat on the floor. At the beginning of every trial, E2 picked a card from the pack and showed it to the child and E1. In every trial, E1 announced which of the two colors on the card they should jump on ("Let's jump on green"). The child jumped first and then E1 followed. After three correct trials, in the next trial, E1 jumped on a different color than the one she announced. E2 waited for the child's reaction, and then he corrected E1 and encouraged

children to do the same whenever E1 made a mistake: "That [E1's name]! Sometimes she does things wrong. Then you can tell her, 'That was wrong! You should do what you say you will do.'" Three more correct trials were followed by one incorrect trial and then two additional correct trials at the end. If in this second incorrect trial children did not spontaneously correct E1 after she jumped to the wrong mat, E2 corrected her, the game continued, and children had two additional chances if needed to correct her.

Training phase: E2 introduced the main game and explained to children and E1 that they could decide to play individually (i.e., sit on the platform representing the beach) and be certain of winning a bird picture (small prize). Alternatively, they could decide to play cooperatively (i.e., jump onto the rug representing the lake) and have the possibility of winning a colorful sticker (big prize), as long as both of them did this together.

E2 adjusted the height of the tube according to the height of the child. E2 showed and verbally explained to children and E1 that when deciding where to jump, both players had to stand on the starting positions and show a 'thumbs-up' sign to each other through the tube as a signal that they were ready. Then they had to hold the tube with both hands, look at each other through the tube, and, while looking at each other, E1 would verbally suggest where they should go. Finally, E1 would count to three and clap her hands as a signal for children to decide what to do. When children and E1 played individually, E2 gave each of them a bird picture. When they played cooperatively, they opened the treasure box together, took out one container, and split the two stickers between them. In the first trial, they both played individually and in the second, they both played cooperatively. In the third trial, E2 instructed E1 to play cooperatively and children to play individually in order to demonstrate that when only one player played cooperatively, that player does not win a prize. After three training trials, as a first pre-test to see if children understood the rules, E2 asked children, "What happens if you jump to the lake and [E1] doesn't?" and "What happens if you jump to the lake and [E1] jumps as well?" If children did not answer correctly (i.e., in the former case, that they would not receive any prize, and in the latter case, that they would both receive stickers), E2 explained the game from the beginning, children and E1 repeated the same three training

trials, and E2 asked these questions one more time. By this stage all children understood the rules and answered these questions correctly. After that, children played a series of individual and cooperative trials interspersed with three 'mismatch' trials in which E1 announced, "Let's jump to the lake, okay?" but after children jumped, E1 played individually, sitting on her platform. E2 waited to see if children would say something about this spontaneously, and if not then he encouraged them to protest to E1 by saying, "You can tell her: What you did was wrong! You should do what you said you would do." Then E2 corrected E1 himself. These mismatch trials in the main game, together with the mismatch trials in the warm-up game, served as a second pre-test and were designed to ensure that children were not too shy to protest to an adult experimenter about her behavior. One might worry that through this training children might have learned that they should protest whenever their partner does not cooperate, or whenever she does not do what she said she would do. However, it is important to note that the training was identical in both conditions, and during the test phase E1 did not say anything, so this training cannot explain our results – it was conducted simply to increase children's willingness to protest against the adult. To pass the second pre-test, children needed to correct E1 spontaneously at least once in the warm-up game (out of the four mismatch trials) and once in the main game (of the three mismatch trials).

Test: The test consisted of three phases, as follows.

Decision-making phase: After the training phase, E2 introduced a no-talking rule, telling children and E1 that from now on they were not allowed to talk while looking through the tube, although they could talk again after having jumped or sat on the beach. Then E2 moved away to write something down and E1 looked through the tube in such a way that children could see only her eye area. Although children could not see it, the corners of E1's mouth were turned slightly upward to express a generally positive attitude in both conditions. In the non-communicative look condition, E1 made neutral eye contact with children for a few seconds. In the communicative look condition, E1 made eye contact with children with an ostensive-communicative look: She opened her eyes wide and raised her eyebrows (see Figure 7). After the look, E1 returned to

an upright position, counted to three, clapped her hands, and waited for children to make their decision, all the while with her gaze still directed neutrally toward the tube. After children had decided, in both conditions E1 played individually. If children played individually, E2 gave them the small prize, and the test phase ended.

a) b) c)

Figure 7. A depiction of a) the non-communicative look, b) the communicative look, and c) the child's face position.

Spontaneous response phase: If children played cooperatively, they then had 30 s to talk spontaneously to E1. E1 sat silently on the platform and looked at children with a slightly positive expression.

Elicited response phase:

After 30 s, E2 returned and asked children five questions.

1. "What happened, [child's name]? Tell me."
2. "Did [E1] perhaps give you any sign that she would jump?" If children responded yes, E2 asked: "What sign did she give you?"
3. "Do you want to say something about that to [E1]? Tell her what you think about what she just did. I'm sure she would be glad to know." E1 nodded.

Then E1 excused herself, left the room, and E2 continued:

4. "Do you remember when you jumped to the lake and [E1] did not? What did you expect her to do?"

5. "Why did you expect that [E1] would jump to the lake/stay at the beach?"

Debriefing: Then E1 returned and apologized for playing incorrectly. They played two successful cooperative trials together, after which children were thanked and taken back to class.

3.2.1.4 Coding and reliability

Sessions were videotaped and coded from the video recordings. A second independent observer, who was naive to the hypotheses of the study, coded a random sample of 25% of children in each condition for reliability. There was perfect agreement on all main measures.

Cooperation vs. individual play: During the test trial, children's behavior was coded either as cooperation, if they jumped onto the rug representing the lake, or as individual play, if they sat on the platform representing the beach. This was coded because we only included children who cooperated.

Children's spontaneous protest after their partner did not cooperate: Children's verbal responses during the spontaneous response phase were transcribed, and we coded for protests that indicated that children understood E1's look as a commitment to cooperate. Promissory commitments can arise when there is intentional communication to one's partner about one's intention to perform some action (Raz, 1977). We did not expect these young children to refer explicitly to commitments; instead, we focused on protests that involved what E1 had communicated that she would do. We decided to focus on this specific type of protest because we had found that in Study 1 the child who protested normatively in the communicative look condition referred implicitly to a communicated commitment. This decision was made before the data collection began. Normative protest referring, directly or indirectly, to the commitment was coded when children used normative vocabulary (e.g., "should," "must," "wrong") and mentioned communication of any type (including verbal, gestural, with the eyes/looks, or with facial expressions, e.g., "You should do what you said!" or "[E1], that was wrong. You should pay

attention to your sign." This enabled us to focus on the commitment part of the protest, disentangling it from children's general disappointment about the fact that their partner did not cooperate (e.g., "You should have cooperated, because I wanted the sticker").

Children's protest across the entire response phase: Children's verbal responses across the entire response phase (i.e., the spontaneous plus the elicited response phase) were also transcribed and coded for normative protest referring, directly or indirectly, to the commitment. This was coded when the child used normative vocabulary and mentioned communication either together in one utterance (e.g., "One does not do that, that was wrong; you should do what you said") or separately across responses to different questions (e.g., if children said in the spontaneous response phase, "You did it wrong, I think," and later responded to one of the questions, "She did this (the child raised her eyebrows), so I did this (the child raised her eyebrows) and because of that I understood that she would stand over the treasure so with that I understood 'to the lake.'"

Children's contribution to the mutual look in the communicative look condition: We again investigated whether children who responded to their partner's communicative looks (thus explicitly acknowledging their partner's commitment and/or even making a joint commitment) were more likely to protest than children who did not. Children's reactions to the communicative look were coded as: raising their eyebrows, nodding (after receiving the look from E1), or no communicative behaviors. For the analyses, a binary outcome variable – whether or not each child produced a communicative behavior – was used.

Supplementary coding of children's responses: To gain more information about children's understanding of the critical test trial, we conducted more detailed coding of children's verbal responses, both in the spontaneous and the elicited response phases. Children's responses were transcribed and coded from the video recordings. All but two reliability measures revealed perfect agreement (range of Cohen's κ's=0.64-1). The two measures (Questions 1 and 6) with a Cohen's κ of 0.64 each had only 1

disagreement out of 18, and these two disagreements were resolved by a third coder who was blind to condition.

3.2.2 Results

3.2.2.1 Children's protest after their partner did not cooperate

At test, more children in the communicative look condition produced normative protests referring to commitment than did children in the non-communicative look condition, both in the spontaneous phase alone (marginally; Fisher's exact test, p=0.06; odds ratio=3.13) and when their responses across the entire response phase were considered (Fisher's exact test, p=0.01; odds ratio=4.38). In the communicative look condition, 39% of children protested spontaneously and 47% protested at least once across the entire response phase. In contrast, in the non-communicative look condition, 17% of children protested, both in the spontaneous part and across the entire response phase.

3.2.2.2 Children's contribution to the mutual look in the communicative look condition

Children who responded with communicative behaviors to E1 tended to be more likely to protest than children who did not (Fisher's exact test, p=0.08; odds ratio=4.20). Forty-seven percent of children in the communicative look condition showed some communicative behaviors (see Table 3). Of these children, 69% protested. In contrast, of the children who did not show any such behaviors, only 33% protested.

Table 3

Children's Communicative Behaviors During the Mutual Look in the Communicative Look Condition and Their Relation to Children's Protesting in Study 2.

	Normative protests referring to commitment	No normative protests referring to commitment
Communicative behaviors	**11**	**5**
Raised eyebrows	10	4
Nodding	1	0
Raised eyebrows and nodding	0	1
No communicative behaviors	**6**	**12**

Note: N=34. Due to a technical error, four videos were missing. E1 noticed the technical failure in two cases and noted live whether children showed any communicative behavior or not. These two cases are included in the table.

3.2.2.3 Supplementary coding of children's responses

We used the same coding for children's spontaneous responses and their responses to the first question, "What happened, [child's name]? Tell me" and the third question, "Do you want to say something about that to [E1's name]? Tell her what you think about what she just did. I'm sure she would be glad to know that." Children's verbal responses were coded for normative protests referring to commitment. For exploratory purposes, children who did not produce this specific type of protest were additionally given one of the following hierarchically-ordered codes: (a) Normative protest without reference to commitment: Children used normative vocabulary but did not refer to commitment (e.g., "You must come with me in the lake!" or "You did it wrong, I

think.”), (b) Communication only: The child mentioned communication of any type (including verbal, gestural, with the eyes/looks, or with facial expressions), but did not use normative language (e.g., “[E1’s name], you said that we would jump into the lake!”), (c) Interrupted normative protest: Children started to protest using normative vocabulary but did not finish the sentence (e.g., “You should…[and then the child stopped],” or “[E1’s name], you must do…”), (d) Other: The child was silent or did not produce a response falling into any of the above categories. The percentage of children whose responses fell into each of these categories is shown in Table 4. Most children in both conditions did not respond to these questions with any type of protest; nor did they mention communication.

For the second question, “Did she (E1) maybe give you any sign that she would jump?”/“What sign did she give you?”) and the fifth question, “Why did you expect that [E1] would jump to the lake/stay at the beach?”, children’s responses were categorized according to whether or not they mentioned communication (defined in the same way as above). Children in the communicative look condition mentioned communication in their answers to the second question significantly more often than children in the non-communicative look condition (28% vs. 3%, respectively; Fisher’s exact test, $p<0.01$; odds ratio=13.06). A similar difference was found in children’s answers to the fifth question: More children in the communicative look condition tended to mention communication there as well (19% vs. 3%, respectively; Fisher’s exact test, $p=0.06$; odds ratio=8.24).

Children’s responses to the fourth question, “Do you remember when you jumped to the lake and [E1] did not? What did you expect her to do?”, were coded as: (a) Cooperation: Children answered that they expected E1 to cooperate (e.g., to jump, to go to the lake, to jump together), (b) Individual play: Children answered that they expected E1 to play individually (e.g., to stay on the beach, not to jump), (c) Do not know, or (d) Other. The game was set up in such a way that children in both conditions were likely to expect cooperation from E1 and therefore to cooperate themselves. The final sample included only children who cooperated, however with this question we were interested in looking at children’s explicit verbal expectations in both conditions. In both

conditions, 75% of children answered that they expected cooperation from E1. This shows that children in the non-communicative look condition had expectations about E1's behavior as well.

At the very end of the procedure, after E1 apologized and played two successful cooperation trials with children, E2 asked the final question, "Are you [still] angry that [E1] did not always do what she said?" Children's responses were coded as: (a) Yes: Children answered positively or said that they were angry just a little bit, or (b) No: Children answered that they were not angry. The results showed that almost no children were angry at E1 at the end of the play session (97% in the communicative look condition and 92% in the non-communicative look condition said that they were not angry).

Table 4

Percentage of Children Who Showed Each Type of Verbal Response in Study 2.

Verbal response types during each response phase	Communicative look condition	Non-communicative look condition
Spontaneous response		
Normative protest referring to commitment	39%	17%
Normative protest without reference to commitment	8%	19%
Communication only	0%	3%
Interrupted normative protest	3%	6%
Other	50%	56%
"What happened?"		
Normative protest referring to commitment	0%	0%
Normative protest without reference to commitment	0%	3%
Communication only	11%	0%
Interrupted normative protest	0%	0%
Other	89%	97%
"Do you want to say something about that to [E1]?"		
Normative protest referring to commitment	3%	0%
Normative protest without reference to commitment	8%	8%
Communication only	3%	0%
Interrupted normative protest	0%	0%
Other	86%	92%
Total (across all phases) normative protest referring to commitment[7]	47%	17%

[7] For the communicative look condition, this includes two children who used normative vocabulary and referred to the commitment in different phases.

3.2.3 Discussion of Study 2

At the moment of decision-making, children's partner engaged in communicative or non-communicative eye contact with them and subsequently did not cooperate. Children who received a communicative look were more likely to protest about this, referring specifically to their partner's commitment, than children who received a non-communicative look. The vast majority of these children referred explicitly to "doing what you said [you would do]," but three children referred to nonverbal signs E1 had given (e.g., a certain kind of look). This suggests that children understood the communicative look as a binding statement from their partner in which she committed herself to choosing the cooperative option in the game.

It is interesting to note that approximately half of the children responded to E1's communicative look with looks (or other communicative behavior) of their own. One possible low-level interpretation of this behavior is that children automatically mimicked E1's communicative look. A second, more plausible interpretation is that children intentionally sent a communicative message back to E1. Two pieces of evidence support the latter interpretation. First, some children mentioned their own responses in their answers to E2's questions, (e.g., "She [E1] did this [the child raised her eyebrows], so I did this [the child raised her eyebrows] so I understood…"). Second, the results showed that children who produced these nonverbal signs themselves tended to be more likely to protest. It is likely that these signs served at least to acknowledge the partner's commitment to cooperate and, perhaps, in addition, to commit children themselves to cooperating too, thus forming a joint commitment. Either way, however, their partner did not fulfill her commitment, and thus children were entitled to protest.

3.3 Part II discussion

When deciding whether or not to cooperate, we are often faced with two challenges: uncertainty about our partner's knowledge and uncertainty about their future behavior. Here we investigated one key nonverbal way of overcoming these challenges. Children who received a communicative, as

compared to a non-communicative, look when deciding whether to cooperate both cooperated more (Study 1) and normatively protested more when their partner did not cooperate (Study 2). Thus, a communicative look can be powerful enough not only to promote expectations of cooperation but also to signal a commitment. This is the first empirical evidence, in adults or children, suggesting that nonverbal signals can be understood as commitments.

In Study 1, we focused on the uncertainty we may have about our partner's knowledge (in this case about the availability of the cooperative option). We demonstrated that the type of eye contact made influenced children's decisions to cooperate. Wyman et al. (2013) had previously shown that a communicative look increased young children's cooperation as well. However, our study goes beyond theirs in isolating precisely what it is that established knowledge in common about the cooperative option and led children to cooperate: the communicativeness of the look. Intentionally communicative eye contact not only transmits information from one social partner to another, but also provides an indication of where each partner's attention is focused, and thus removes any doubt about whether the partners are paying attention to the same thing and to each other (Carpenter & Liebal, 2011; Gómez, 1994). It thus creates a public space between the partners, enabling them to focus on something together (Gilbert, 2007).

Therefore, here we add a new type of information that children can use to solve the knowledge problem in coordination problems. Previous studies have shown that 5-year-old children are able to solve different coordination problems by using salience (Grueneisen et al., 2015a), majority information (Grueneisen, Wyman, & Tomasello, 2015b), and their partners' cultural knowledge (Goldvicht-Bacon & Diesendruck, 2016). Study 1 extends these findings by focusing on interpersonal signals that children can use in solving coordination problems.

In Study 2, we focused on the second challenge for successful cooperation involving uncertainty about one's partner's behavior. We found that after receiving a communicative look, children understood that their partner was committing herself to cooperate. Previous studies have all involved children's understanding of verbal promises and commitments (Gräfenhain et

al., 2009; Kachel et al., 2017; Kanngiesser et al., 2017). Here we showed that a nonverbal communicative signal can be understood as a commitment as well. Thus, along with contributing to the psychological literature by demonstrating that young children can normatively evaluate a nonverbal communicative signal as a binding statement, we also contribute to the theoretical debate in the philosophical literature, by showing that it is indeed possible to make a promissory commitment nonverbally. Even though a communicative look is not a conventionalized way of committing oneself, children still protested normatively and referred to a commitment (e.g., mentioning 'doing what you said you would do') after their partner looked at them communicatively but did not cooperate. This finding supports the more contemporary accounts of promissory obligations (Gilbert, 2004; Owens, 2006; Scanlon, 1998; Shiffrin, 2008) that argue that the core of promissory commitments might not be rooted in social convention (as Hume, 1739-1740/1969, Rawls, 1955, & Searle, 1969, have proposed).

One might wonder whether the communicative look signaled a promissory commitment in the strict sense, with normative obligations and rights involved, or rather only a more minimal 'sense of commitment' (Michael, Sebanz, & Knoblich, 2015). We argue for the former interpretation. All the children in Study 2 had decided to cooperate and in both conditions 75% of children answered to a post-test question that they had expected cooperation from their partner. This shows that children in the non-communicative look condition had explicit expectations about cooperation as well. However, after receiving a communicative look, around half of the children not only felt that they had the right to protest against their partner when she did not cooperate, but also often referred specifically in their protests to what she had said she would do. One might argue that children still did not understand their partner's nonverbal message as a binding statement committing her to cooperate. However, it does not seem plausible that intentional communication in this context can be perceived as an informative but non-binding statement at the same time (e.g., "I will cooperate, but I do not promise").

The communicative look could signal a commitment because children and their partner had the common ground that during the decision-making

phase they were deciding whether to cooperate. This common ground supported their interpretation of the look. The most relevant – and the only missing piece of – information when children saw the look was about whether E1 had seen the cooperative option (Study 1) or whether she was going to cooperate (Study 1 and 2). In fact, the difference in context and the most relevant information in the situation could explain why children apparently interpreted the communicative look differently in the two studies. In Study 1, children and their partner already had common ground about the payoff structure of the game: They knew together that the cooperative option had a higher payoff, but that it was risky, because if one partner decided for cooperation alone, she would lose the chance to gain anything. However, there was uncertainty created by the partial occluder about whether the partner could see the cooperative option or not. Children might thus have interpreted the communicative look as "I see it!", which then led them to expect cooperation but not necessarily to hold their partner normatively responsible for not cooperating. In contrast, in Study 2, children and their partner not only had common ground about the payoff structure but also about the availability of the cooperative option, so the most relevant information during the decision-making phase was about the decision to cooperate or not. Therefore, children likely interpreted the look as directly related to deciding whether to cooperate and the communicative look could signal a commitment. This might be one main reason why children protested only in Study 2, not in Study 1.

There are likely several other reasons why children protested in Study 2, and not in Study 1. First, in Study 2 there were several training rounds to ensure that children were not too shy to protest to an adult experimenter about her behavior. Thus, it is possible that in Study 2 children felt more confident about blaming their partner for the coordination's failure. Second, in Study 2 we tested older children (6- to 7-year-olds) than in Study 1 (5-year-olds). Thus, children's developing understanding of social norms (see, e.g., Smith et al., 2013) might have contributed to our findings. Even 3-year-olds protest when their partner intentionally performs a task incorrectly (Kachel et al., 2017) or violates simple game rules (Rakoczy et al., 2008), but in our studies children faced an arguably much more challenging task. In a Stag Hunt game, the fact

that the partner played individually is not in itself an incorrect or wrong action, so children could not refer to any rule of the game in their protests. Instead, what was wrong was that their partner intentionally made children believe that she would cooperate, but then did not follow through. As she did all this nonverbally, it might be that only older children could formulate this in their protests. The relation between children's developing understanding of norms and nonverbal commitments, as well as the question of whether even younger children also understand nonverbal commitments, are two important questions for future research. Even infants can infer some types of information from ostensive-communicative eye contact (Csibra & Gergely, 2006; Senju & Csibra, 2008), but we would only expect children to show the same pattern of protest results once they have reached an age at which they understand something about norms and commitments, around 3 years of age (Gräfenhain et al., 2009; Rakoczy et al., 2008).

It is interesting to note that looks do not have the specificity that language has, but this is not a problem here, because the certainty connected to a specific meaning just needs to be sufficient for the current purposes (Monderer & Samet, 1989). It is an open question in this regard whether sharing attention via nonverbal signals can fulfill all the same functions in creating expectations and signaling commitments as sharing attention via verbal signals, or whether under certain conditions (e.g., when the risk is too high, for example risking one's life), we might hesitate to rely on nonverbal signals. Future research should manipulate the risk-to-reward ratio to investigate this. Similarly, it would be interesting to compare the effectiveness of increasingly specific nonverbal signals (from just looking to smiling and/or winking and/or pointing and/or nodding) in promoting coordination success and creating commitments, and to study this cross-culturally to see if children in different cultures vary in their normative understanding of communicative signals with various degrees of explicitness.

In summary, these studies are the first to show that the subtle, nonverbal cue of a communicative look – but not a non-communicative look – can be a powerful tool for increasing coordination success in social dilemmas. In the context of a key coordination dilemma, communicative looks at the critical

decision-making phase produced an expectation of collaboration in 5-year-old children, and 6- to 7-year-old children apparently even understood the communicative looks as a commitment to cooperate. The ability to produce communicative looks, infer their meaning, and make commitments nonverbally might have been important steps in the evolution of action coordination in humans. These skills would have enabled successful coordination even before language evolved.

4. General discussion

Humans differ from other animals in the extent to which they shift from independent experience to interconnected experience of the world. We are not only sensitive to our social partner's focus of attention, but we are able to attend to things, events, and ideas together and to adapt our behavior accordingly. Coordination of attention and knowledge states with others is one of the foundations of the social and cooperative nature of humans (Tomasello & Carpenter, 2007).

The present thesis explored the ways in which we coordinate our attention with others and how this supports cooperation. First, in order to help resolve the problem of lack of consensus in the literature about what joint attention is, in the first part of the thesis I proposed a new framework of different social attention levels and corresponding knowledge state levels. More specifically, I systematically discussed cognitive, behavioral, and phenomenological aspects of the different levels. In monitoring attention, one takes a third-person's perspective and monitors the other's attention to something, for example by gaze following. In common attention, by taking a third-person perspective and making recursive inferences, each partner is aware that they are both attending to the same thing as well as to each other's attention to the thing. In mutual attention, partners engage in a second-person perspective. They both experience each other attending to the same thing and their attention to each other, for example by making eye contact about the thing. In shared attention, in addition to the conditions of mutual attention, partners actively share attention by communicating about the object of their attention. The framework demonstrates that joint attention is not a single cognitive skill. It also emphasizes the need to investigate attention and knowledge states together, as they are closely intertwined: when partners engage in these attention levels, they have qualitatively different knowledge states. The idea is that the social attention and knowledge state levels serve distinct functions and support different types of social interactions. Although all the levels have the potential to promote cooperation, the framework highlights the special importance of the shared attention level in this.

In the second part of the thesis I investigated the role of shared attention in supporting cooperation. In two empirical studies, young children played a coordination game in which they needed to decide whether to cooperate or play individually. The children's partner made either communicative eye contact or looked non-communicatively at them before their decision. In Study 1, results showed that communicative eye contact produced an expectation of collaboration in children. In Study 2, results showed that children normatively protested when their partner did not cooperate. These findings support the idea that sharing attention via communicative, but not non-communicative, eye contact can help social partners to achieve successful coordination. Communicative eye contact served as a powerful coordination tool for children facing a coordination dilemma. First, it established shared knowledge about the availability of the cooperative option, and therefore shared knowledge about the possibility of choosing the cooperative option. Second, it decreased uncertainty about their partner's future behavior as it helped to align their expectations, to create a commitment to cooperate, and to establish the right to complain if the other did not cooperate.

In the following section, first I summarize how the social attention and knowledge levels support cooperation and suggest that levels with a higher degree of jointness bring additional benefits for cooperation. Then I turn to new questions and propose directions for future research. Finally, I conclude with remarks about why it is beneficial to further investigate different types of social attention.

4.1 Social attention and knowledge states support cooperation

Coordination of attention and knowledge states with others enables better ways of cooperating. Action coordination is more effective, and it also allows a more complex type of coordination to emerge: strategic coordination. In coordination dilemmas, partners need to decide how to act based on their expectations of what others will do. They must overcome two types of uncertainty: first, uncertainty about their knowledge (e.g., knowledge about the presence of a collaborative option), and second, uncertainty about their partner's behavior (e.g., whether their partner will choose the collaborative

option). Shared knowledge helps to solve the former type of uncertainty, and it enables the formation of expectations and commitments, which help to solve the latter type.

Different attention and knowledge levels support the commitments and obligations, and the higher degree of jointness the levels have, the stronger the support they provide. This is the case for several reasons. First, higher degrees of jointness establish more certainty for social partners. For example, in the levels with lower degrees of jointness (monitoring and common attention) partners have to consider the other's attentional state, but in the levels with a higher degree of jointness (mutual and shared attention) the direct contact gives them immediate evidence that they are paying attention to each other (Carpenter & Liebal, 2011; Gómez, 1996). Second, in the level with lowest degree of jointness (monitoring attention), partners can only produce unilateral knowledge about the partner's focus of attention (e.g., she sees the collaborative option) and a unilateral expectation about the partner's behavior. However, the levels with a higher degree of jointness (common, mutual, and shared attention) can produce knowledge in common about their focus of attention (e.g., we see the collaborative option). In addition, the expectations about the partner's probable behavior could become knowledge in common between partners as well. This might cause partners to feel some sense of minimal commitment to fulfill these expectations (J. Michael, personal communication, January 15, 2018). The openness in the mutual and shared levels heightens the pressure to fulfill these expectations beyond the common level. Third, as it is difficult to plausibly deny that attention and knowledge are shared in the level with the highest degree of jointness, shared attention, in normative contexts can give rise to commitments and obligations.

Interestingly, the distinction between second-person and third-person ways of achieving knowledge in common may link to dual process theories of social cognition. These theories propose that there are two different kinds of mental processes: processes that are automatic, fast, efficient, and less cognitively demanding, and processes that are controlled, slower, and cognitively effortful (e.g., Apperly & Butterfill, 2009; Frith & Frith, 2008; Kahneman & Frederick, 2002). Empirical research is needed to clarify the

nature of processing in the second- versus third- person levels and to answer the question about their ontogenetic and phylogenetic primacy or co-development.

4.2 Future directions

The presented findings have opened up new questions for future research. One useful contribution of the theoretical part of the thesis is the scale of jointness idea and discussion of. It is interesting to note that several factors that influence different aspects of jointness differ significantly across different societies, for instance, the degree of interdependency between social partners, the frequency of face-to-face interactions, the focus of attention on oneself vs. others, the sensitivity to interpersonal and contextual cues, and the importance of indirect vs. direct communication styles. These factors might influence when people choose to engage in different levels, which specific behaviors they use, and what are the normative consequences of these levels. Future studies are needed to elucidate the specific roles of these factors in different societies. These studies would not only help us to generalize the findings presented in this thesis but also to improve our understanding of how individuals from different societies can cooperate more effectively.

Although the framework does not presuppose that social partners can engage in the different levels only when they are physically co-present, the examples used to illustrate the levels, as well as the two empirical studies, focused on the context in which the partners were in the same place at the same time. An interesting topic for future work would be to explore the role of shared attention in cooperation without physical co-presence, for example in virtual environments (e.g., using social media, messaging platforms, emails). We have started to rely on virtual environments to a greater extent than ever before. When partners are co-present, they can co-create their experience by bidirectional influence and quickly align their mental states about the co-attended object. In virtual environments, certainty about shared attention and knowledge might be decreased, because the contact between partners is restricted. For example, when using video call apps, it is harder to interpret the other's reactions (e.g., is the gaze away related to what I am saying or to some

distraction behind screen). In messaging apps, the contact is delayed in time and there is typically no simultaneous confirmation that both partners read the message. Therefore, it would be useful to test whether the social attention levels in virtual environments, compared to real environments, are equally effective in supporting communication and cooperation and which behaviors are especially beneficial to support smooth and mistake-free episodes of shared attention in virtual environments.

Another promising line of research would be to study the role of the social attention levels in facilitating prosocial behavior. It would be interesting to explore to what extent normative forces to behave prosocially are influenced by the different attention and knowledge levels. To illustrate: a young man is sitting in a tram and he sees an injured person enters the tram. He knows that he should offer his seat. It is likely that levels with higher degrees of jointness increase prosocial behavior. This might be the case because plausible deniability that the young man does not know that his help is needed differs between the levels. All levels with knowledge in common might facilitate young man's perceived obligation to get up over monitoring attention (with individual knowledge). In a recent study, we found that the probability of young children's helping in a similar situation was higher in common attention (children and an injured experimenter had common knowledge about the fact that children's help is needed) compared to monitoring attention (children knew individually that their help is needed) (Siposova, Grueneisen, Helming, Tomasello, & Carpenter, in prep.). Moreover, differences might exist even between the common, mutual and shared level. Future studies could manipulate the jointness about the need for help and measure whether it affects prosociality. For example, in all conditions the participant having a seat can see that an injured person needs a seat, what varies is which attention levels the participant engages in. The crucial measure is whether and to what extent the participant provides help.

4.3 Conclusion

Mutual understanding and efficient coordination are the building blocks of flourishing societies; they make the difference between failure and success in

dyads, groups, organizations, and nations. Therefore, it is important to investigate the underlying processes that make them possible and the conditions under which they can thrive. The ability to engage in different social attention levels and achieve corresponding knowledge states is crucial, as it allows us to focus on things together in different ways. Especially useful is the shared attention level, characterized by bidirectional and communicative co-creation of experience. By further studying social attention levels, we might advance our understanding of the key features that make us especially social, sharing and cooperating species.

References

Admoni, H., & Scassellati, B. (2017). Social eye gaze in human-robot interaction: A review. *Journal of Human-Robot Interaction*, *6*(1), 25–63. doi:10.5898/jhri.6.1.admoni

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development*, *72*(2), 416–430. doi:10.1111/1467-8624.00287

Amici, F., Aureli, F., Visalberghi, E., & Call, J. (2009). Spider monkeys (*Ateles geoffroyi*) and capuchin monkeys (*Cebus apella*) follow gaze around barriers: Evidence for perspective taking? *Journal of Comparative Psychology*, *123*(4), 368–374. doi:10.1037/a0017079

Anzalone, S. M., Boucenna, S., Ivaldi, S., & Chetouani, M. (2015). Evaluating the engagement with social robots. *International Journal of Social Robotics*, *7*(4), 465–478. doi:10.1007/s12369-015-0298-7

Anzalone, S. M., Tilmont, E., Boucenna, S., Xavier, J., Jouen, A.-L., Bodeau, N., … Cohen, D. (2014). How children with autism spectrum disorder behave and explore the 4-dimensional (spatial 3D+time) environment during a joint attention induction task with a robot. *Research in Autism Spectrum Disorders*, *8*(7), 814–826. doi:10.1016/j.rasd.2014.03.002

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–970. doi:10.1037/a0016923

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge, England: Cambridge University Press.

Argyle, M., & Dean, J. (1965). Eye-contact, distance and affiliation. *Sociometry*, *28*(3), 289–304. doi:10.2307/2786027

Argyle, M., Ingham, R., Alkema, F., & McCallin, M. (1973). The different functions of gaze. *Semiotica*, *7*(1), 19–32.

Austin, J. L. (1975). *How to do things with words*. Cambridge, MA: Harvard University Press.

Axelrod, R., & Hamilton, W. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396. doi:10.1126/science.7466396

Bach, K., & Hamish, R. M. (1979). *Linguistics communication and speech acts*. Cambridge, MA: MIT Press.

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, *55*(4), 1278–1289. doi:10.2307/1129997

Bard, K. A., Bakeman, R., Boysen, S. T., & Leavens, D. A. (2014). Emotional engagements predict and enhance social cognition in young chimpanzees. *Developmental Science*, *17*(5), 682–696. doi:10.1111/desc.12145

Bard, K. A., Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M., Costall, A., & Matsuzawa, T. (2005). Group differences in the mutual gaze of chimpanzees (*Pan troglodytes*). *Developmental Psychology*, *41*(4), 616–624. doi:10.1037/0012-1649.41.4.616

Bard, K. A., & Vauclair, J. (1984). The communicative context of object manipulation in ape and human adult-infant pairs. *Journal of Human Evolution*, *13*(2), 181–190. doi:10.1016/s0047-2484(84)80062-7

Baron-Cohen, S. (1989). Joint-attention deficits in autism: Towards a cognitive analysis. *Development and Psychopathology*, *1*(3), 185–189.

Baron-Cohen, S., & Cross, P. (1992). Reading the eyes: Evidence for the role of perception in the development of a theory of mind. *Mind & Language*, *7*(1-2), 172–186. doi:10.1111/j.1468-0017.1992.tb00203.x

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a "language of the eyes"? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition*, *4*(3), 311–331. doi:10.1080/713756761

Bicchieri, C. (2006). *The grammar of society*. Cambridge, England: Cambridge University Press. doi:10.1017/cbo9780511616037

Bigelow, A. E. (2003). The development of joint attention in blind infants. *Development and Psychopathology*, *15*(2), 259–275. doi:10.1017/s0954579403000142

Böckler, A., Knoblich, G., & Sebanz, N. (2011). Giving a helping hand: Effects of joint attention on mental rotation of body parts. *Experimental Brain Research*, *211*(3-4), 531–545. doi:10.1007/s00221-011-2625-z

Bohn, M., Call, J., & Tomasello, M. (2015). Communication about absent entities in great apes and human infants. *Cognition*, *145*, 63–72. doi:10.1016/j.cognition.2015.08.009

Boothby, E. J., Clark, M. S., & Bargh, J. A. (2014). Shared experiences are amplified. *Psychological Science*, *25*(12), 2209–2216. doi:10.1177/0956797614551162

Boothby, E. J., Smith, L. K., Clark, M. S., & Bargh, J. A. (2016). Psychological distance moderates the amplification of shared experience. *Personality and Social Psychology Bulletin*, *42*(10), 1431–1444. doi:10.1177/0146167216662869

Boothby, E. J., Smith, L. K., Clark, M. S., & Bargh, J. A. (2017). The world looks better together: How close others enhance our visual experiences. *Personal Relationships*, *24*(3), 694–714. doi:10.1111/pere.12201

Botero, M. (2016). Tactless scientists: Ignoring touch in the study of joint attention. *Philosophical Psychology*, *29*(8), 1200–1214. doi:10.1080/09515089.2016.1225293

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

Bräuer, J., Call, J., & Tomasello, M. (2005). All great ape species follow gaze to distant locations and around barriers. *Journal of Comparative Psychology*, *119*(2), 145–154. doi:10.1037/0735-7036.119.2.145

Bräuer, J., Call, J., & Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Animal Cognition*, *10*(4), 439–448. doi:10.1007/s10071-007-0088-1

Breazeal, C., & Scassellati, B. (2002). Robots that imitate humans. *Trends in Cognitive Sciences*, *6*(11), 481–487. doi:10.1016/s1364-6613(02)02016-8

Brewer, M. B. (2007). The importance of being we: Human nature and intergroup relations. *American Psychologist*, *62*(8), 728–738. doi:10.1037/0003-066x.62.8.728

Brinck, I. (2004). The pragmatics of imperative and declarative pointing. *Cognitive Science Quarterly*, *3*(4), 429–446.

Brinck, I., & Gardenfors, P. (2003). Co-operation and communication in apes and humans. *Mind and Language*, *18*(5), 484–501. doi:10.1111/1468-0017.00239

Brosnan, S. F., Wilson, B. J., & Beran, M. J. (2012). Old world monkeys are more similar to humans than new world monkeys when playing a coordination game. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*(1733), 1522–1530. doi:10.1098/rspb.2011.1781

Brownell, C. A. (2011). Early developments in joint action. *Review of Philosophy and Psychology*, *2*(2), 193–211. doi:10.1007/s13164-011-0056-1

Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation and*

Developmental Disabilities Research Reviews, *10*(3), 169–175. doi:10.1002/mrdd.20036

Bruner, J. S. (1974). From communication to language - A psychological perspective. *Cognition*, *3*(3), 255–287. doi:10.1016/0010-0277(74)90012-2

Buber, M. (1958). *I and Thou*. New York, NY: Charles Scribner's Sons.

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

Burghardt, G. M., & Greene, H. W. (1988). Predator simulation and duration of death feigning in neonate hognose snakes. *Animal Behaviour*, *36*(6), 1842–1844. doi:10.1016/s0003-3472(88)80127-1

Butterworth, G. (1995). Origins of minds in perception and action. In C. Moore, & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 29–40). Hillsdale, NJ: Erlbaum.

Calabi, C. (2008). Winks, sighs and smiles? Joint attention, common knowledge and ephemeral groups. In N. Psarros, H. B. Schmidt, & K. Schulte-Ostermann (Eds.), *Concepts of sharedness* (p. 41–58). Frankfurt, Germany: Ontos Verlag.

Camaioni, L. (1993). The development of intentional communication: A re-analysis. In J. Nadel, & L. Camaioni (Eds.), *New perspectives in early communicative development* (pp. 82–96). London, England: Routledge.

Campbell, J. (2005). Joint attention and common knowledge. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 287–297). Oxford, England: Clarendon Press.

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

Carpenter, M., & Call, J. (2013). How joint is the joint attention of apes and human infants? In J. Metcalfe, & H.S. Terrace (Eds.), *Agency and joint attention* (pp. 49–61). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780199988341.003.0003

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* (pp. 159–181). Cambridge, MA: MIT Press.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), serial no. 255.

Carpenter, M., Tomasello, M., & Savage-Rumbaugh, S. (1995). Joint attention and imitative learning in children, chimpanzees, and enculturated chimpanzees. *Social Development*, *4*(3), 217–237. doi:10.1111/j.1467-9507.1995.tb00063.x

Castro-González, Á., Admoni, H., & Scassellati, B. (2016). Effects of form and motion on judgments of social robots' animacy, likability, trustworthiness and unpleasantness. *International Journal of Human-Computer Studies*, *90*, 27–38. doi:10.1016/j.ijhcs.2016.02.004

Chwe, M. S.-Y. (2001). *Rational ritual: Culture, coordination, and common knowledge*. Princeton, NJ: Princeton University Press.

Clark, H. H. (1992). *Arenas of language use*. Chicago, IL: University of Chicago Press.

Clark, H. H. (1996). *Using language*. Cambridge, England: Cambridge University Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Ed.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association. doi:10.1037/10096-006

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*(1), 62–81. doi:10.1016/j.jml.2003.08.004

Clark, H. H., & Marshall, C. R. (1981). *Definite reference and mutual knowledge*. Cambridge, England: Cambridge University Press.

Cole, P. M., Zahn-Waxler, C., & Smith, K. D. (1994). Expressive control during a disappointment: Variations related to preschoolers' behavior problems. *Developmental Psychology*, *30*(6), 835–846. doi:10.1037/0012-1649.30.6.835

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata, & M. H. Johnson (Eds.), *Processes of change in brain and cognitive development. Attention and performance, XXI*. (pp. 249–274). Oxford, England: Oxford University Press.

Deák, G. O., Flom, R. A., & Pick, A. D. (2000). Effects of gesture and target on 12-and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental Psychology*, *36*(4), 511. doi:10.1037/0012-

1649.36.4.511

DeJaegher, H., DiPaolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*(10), 441–447. doi:10.1016/j.tics.2010.06.009

Duguid, S., Wyman, E., Bullinger, A. F., Herfurth-Majstorovic, K., & Tomasello, M. (2014). Coordination strategies of chimpanzees and human children in a Stag Hunt game. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*(1796). doi:10.1098/rspb.2014.1973

Eilan, N. (submitted). Joint attention and the second person. *Manuscript submitted for publication. Draft available online: http://www2.warwick.ac.uk/fac/soc/philosophy/people/eilan/*.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, *24*(6), 581–604. doi:10.1016/s0149-7634(00)00025-7

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, *99*(14), 9602–9605. doi:10.1073/pnas.152159999

Farroni, T., Mansfield, E. M., Lai, C., & Johnson, M. H. (2003). Infants perceiving and acting on the eyes: Tests of an evolutionary hypothesis. *Journal of Experimental Child Psychology*, *85*(3), 199–212. doi:10.1016/s0022-0965(03)00022-5

Floor, P., & Akhtar, N. (2006). Can 18-month-old infants learn words by listening in on conversations? *Infancy*, *9*(3), 327–339. doi:10.1207/s15327078in0903_4

Flombaum, J. I., & Santos, L. R. (2005). Rhesus monkeys attribute perceptions to others. *Current Biology*, *15*(5), 447–452. doi:10.1016/j.cub.2004.12.076

Frith, C. D., & Frith, U. (2008). Implicit and explicit processes in social cognition. *Neuron*, *60*(3), 503–510. doi:10.1016/j.neuron.2008.10.032

Fuchs, T. (2012). The phenomenology and development of social perspectives. *Phenomenology and the Cognitive Sciences*, *12*(4), 655–683. doi:10.1007/s11097-012-9267-x

Gallotti, M., & Frith, C. D. (2013). Social cognition in the we-mode. *Trends in Cognitive Sciences*, *17*(4), 160–165. doi:10.1016/j.tics.2013.02.002

Ganea, P. A., & Saylor, M. M. (2007). Infants' use of shared linguistic information to clarify ambiguous requests. *Child Development*, *78*(2),

493–502. doi:10.1111/j.1467-8624.2007.01011.x

Genty, E., Neumann, C., & Zuberbühler, K. (2015). Bonobos modify communication signals according to recipient familiarity. *Scientific Reports*, *5*(1). doi:10.1038/srep16442

Gilbert, M. (1989). *On social facts*. London, England: Routledge.

Gilbert, M. (1990). Walking together: A paradigmatic social phenomenon. *Midwest Studies In Philosophy*, *15*(1), 1–14. doi:10.1111/j.1475-4975.1990.tb00202.x

Gilbert, M. (2004). Scanlon on promissory obligation: The problem of promisees' rights. *Journal of Philosophy*, *101*(2), 83–109. doi:10.5840/jphil200410126

Gilbert, M. (2007). Mutual recognition, common knowledge, and joint attention. In T. Ronnow-Rasmussen, B. Petersson, J. Josefsson, & D. Egonsson (Eds.), *Hommage à Wlodek: Philosophical papers dedicated to Wlodek Rabinowicz* (pp. 1–21). Lund, Sweden: Lund University.

Goldvicht-Bacon, E., & Diesendruck, G. (2016). Children's capacity to use cultural focal points in coordination problems. *Cognition*, *149*, 95–103. doi:10.1016/j.cognition.2015.12.016

Gómez, J. C. (1994). Mutual awareness in primate communication: A Gricean approach. In S. Parker, R. Mitchell, & M. Boccia (Eds.), *Self-awareness in animals and humans: Developmental perspectives* (pp. 61–80). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511565526.007

Gómez, J. C. (1996). Ostensive behavior in great apes: The role of eye contact. In A. E. Russon, K. A. Bard, & S. T. Parker (Eds.), *Reaching into thought: The minds of the great apes* (pp. 131–151). Cambridge, England: Cambridge University Press.

Gómez, J. C. (2005). Joint attention and the notion of subject: Insights from apes, normal children and children with autism. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 65–84). Oxford, England: Clarendon Press.

Gómez, J. C. (2010). The emergence of eye contact as an intersubjective signal in an infant gorilla: implications for models of early social cognition. *Acción Psicológica*, *7*(2), 35–43.

Gómez, J. C., Sarriá, E., & Tamarit, J. (1993). The comparative study of early communication and theories of mind: Ontogeny, phylogeny, and pathology. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.),

Understanding other minds: Perspectives from autism (pp. 397–426). New York, NY: Oxford University Press.

Grosse, G., Scott-Phillips, T. C., & Tomasello, M. (2013). Three-year-olds hide their communicative intentions in appropriate contexts. *Developmental Psychology*, *49*(11), 2095–2101. doi:10.1037/a0032017

Gräfenhain, M., Behne, T., Carpenter, M., & Tomasello, M. (2009). Young children's understanding of joint commitments. *Developmental Psychology*, *45*(5), 1430–1443. doi:10.1037/a0016122

Grueneisen, S., Wyman, E., & Tomasello, M. (2015a). Children use salience to solve coordination problems. *Developmental Science*, *18*(3), 495–501. doi:10.1111/desc.12224

Grueneisen, S., Wyman, E., & Tomasello, M. (2015b). Conforming to coordinate: Children use majority information for peer coordination. *British Journal of Developmental Psychology*, *33*(1), 136–147. doi:10.1111/bjdp.12078

Hains, S. M. J., & Muir, D. W. (1996). Infant sensitivity to adult eye direction. *Child Development*, *67*(5), 1940–1951. doi:10.2307/1131602

Harder, P., & Kock, C. (1976). *The theory of presupposition failure*. Copenhagen, Denmark: Akademisk Forlag.

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*(4), 771–785. doi:10.1006/anbe.1999.1377

Haun, D. B. M., & Over, H. (2013). Like me: A homophily-based account of human culture. In P. J. Richerson, & M. H. Christiansen (Eds.), *Cultural evolution: Society, technology, language, and religion* (pp. 75–85). Cambridge, MA: MIT.

Heal, J. (2005). Joint attention and understanding the mind. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds*. Oxford, England: Clarendon Press.

Hecke, A. V. V., Mundy, P. C., Acra, C. F., Block, J. J., Delgado, C. E. F., Parlade, M. V., ... Pomares, Y. B. (2007). Infant joint attention, temperament, and social competence in preschool children. *Child Development*, *78*(1), 53–69. doi:10.1111/j.1467-8624.2007.00985.x

Hennig, C. W. (1977). Effects of simulated predation on tonic immobility in *Anolis carolinensis*: The role of eye contact. *Bulletin of the Psychonomic Society*, *9*(4), 239–242. doi:10.3758/bf03336987

Heyman, G. D., Fu, G., Lin, J., Qian, M. K., & Lee, K. (2015). Eliciting promises from children reduces cheating. *Journal of Experimental Child Psychology*, *139*, 242–248. doi:10.1016/j.jecp.2015.04.013

Hobaiter, C., Leavens, D. A., & Byrne, R. W. (2014). Deictic gesturing in wild chimpanzees (*Pan troglodytes*)? Some possible cases. *Journal of Comparative Psychology*, *128*(1), 82–87. doi:10.1037/a0033757

Hobson, J. A., & Hobson, R. P. (2007). Identification: The missing link between joint attention and imitation? *Development and Psychopathology*, *19*(02), 411–431. doi:10.1017/s0954579407070204

Hobson, R. P. (1993). The emotional origins of social understanding. *Philosophical Psychology*, *6*(3), 227–249. doi:10.1080/09515089308573090

Hobson, R. P. (2005). What puts the jointness into joint attention? In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 185–204). Oxford, England: Clarendon Press.

Hobson, R. P. (2007). Communicative depth: Soundings from developmental psychopathology. *Infant Behavior and Development*, *30*(2), 267–277. doi:10.1016/j.infbeh.2007.02.006

Hobson, R. P. (2014). The coherence of autism. *Autism*, *18*(1), 6–16. doi:10.1177/1362361313497538

Hobson, R. P., & Hobson, J. A. (2008). Engaging, sharing, knowing. Some lessons from research in autism. In J. Zlatev, T. P. Racine, C. Sinha, & E. Itkonen (Eds.), *The shared mind. Perspectives on intersubjectivity (pp. 67–88)*. Amsterdam, Netherlands: John Benjamins.

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science*, *9*(2), 131–134. doi:10.1111/1467-9280.00024

Hume, D. (1969). *A treatise of human nature*. London, England: Penguin. (Original work published 1739–1740).

Kachel, U., Svetlova, M., & Tomasello, M. (2017). Three-year-olds' reactions to a partner's failure to perform her role in a joint commitment. *Child Development*. doi:10.1111/cdev.12816

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Ed.), *Heuristics and Biases* (pp. 49–81). Cambridge, England: Cambridge University Press. doi:10.1017/cbo9780511808098.004

Kaller, T. (2012). *Joint attention in wild chimpanzees and human infants: A comparative study*. (PhD thesis, University of York). York, England. Retrieved from http://etheses.whiterose.ac.uk/3182/

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, *109*(2), 224–234. doi:10.1016/j.cognition.2008.08.010

Kaminski, J., Pitsch, A., & Tomasello, M. (2013). Dogs steal in the dark. *Animal Cognition*, *16*(3), 385–394. doi:10.1007/s10071-012-0579-6

Kaminski, J., Riedel, J., Call, J., & Tomasello, M. (2005). Domestic goats, *Capra hircus*, follow gaze direction and use social cues in an object choice task. *Animal Behaviour*, *69*(1), 11–18. doi:10.1016/j.anbehav.2004.05.008

Kaplan, F., & Hafner, V. V. (2006). The challenges of joint attention. *Interaction Studies*, *7*(2), 135–169.

Kanngiesser, P., Köymen, B., & Tomasello, M. (2017). Young children mostly keep, and expect others to keep, their promises. *Journal of Experimental Child Psychology*, *159*, 140–158. doi:10.1016/j.jecp.2017.02.004

King, A. J., Narraway, C., Hodgson, L., Weatherill, A., Sommer, V., & Sumner, S. (2010). Performance of human groups in social foraging: the role of communication in consensus decision making. *Biology Letters*, *7*(2), 237–240. doi:10.1098/rsbl.2010.0808

Kobayashi, H., & Kohshima, S. (1997). Unique morphology of the human eye. *Nature*, *387*(6635), 767–768. doi:10.1038/42842

Kolodny, N., & Wallace, R. J. (2003). Promises and practices revisited. *Philosophy & Public Affairs*, *31*(2), 119–154. doi:10.1111/j.1088-4963.2003.00119.x

Kristen, S., Sodian, B., Thoermer, C., & Perst, H. (2011). Infants' joint attention skills predict toddlers' emerging mental state language. *Developmental Psychology*, *47*(5), 1207. doi:10.1037/a0024808

Leavens, D. A., & Bard, K. A. (2011). Environmental influences on joint attention in great apes: implications for human cognition. *Journal of Cognitive Education and Psychology*, *10*(1), 9–31. doi:10.1891/1945-8959.10.1.9

Leavens, D. A., Hopkins, W. D., & Bard, K. A. (2005). Understanding the point of chimpanzee pointing. *Current Directions in Psychological Science*, *14*(4), 185–189. doi:10.1111/j.0963-7214.2005.00361.x

Leavens, D. A., & Racine, T. (2009). Joint attention in apes and humans. Are humans unique? *Journal of Consciousness Studies*, *16*, 240–267.

Lee, J. J., Breazeal, C., & DeSteno, D. (2017). Role of speaker cues in attention inference. *Frontiers in Robotics and AI*, *4*. doi:10.3389/frobt.2017.00047

Leite, I., McCoy, M., Lohani, M., Ullman, D., Salomons, N., Stokes, C., ... Scassellati, B. (2017). Narratives with robots: The impact of interaction context and individual differences on story recall and emotional understanding. *Frontiers in Robotics and AI*, *4*. doi:10.3389/frobt.2017.00029

Lemaignan, S., Garcia, F., Jacq, A., & Dillenbourg, P. (2016). From real-time attention assessment to with-me-ness in human-robot interaction. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (pp. 157–164). IEEE Press. doi:10.1109/hri.2016.7451747

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

Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Developmental Science*, *12*(2), 264–271. doi:10.1111/j.1467-7687.2008.00758.x

Liebal, K., Carpenter, M., & Tomasello, M. (2010). Infants' use of shared experience in declarative pointing. *Infancy*, *15*(5), 545–556. doi:10.1111/j.1532-7078.2009.00028.x

Liszkowski, U., Albrecht, K., Carpenter, M. & Tomasello, M. (2008). Infants' visual and auditory communication when a partner is or is not visually attending. *Infant Behavior and Development*, *31*(2), 157–167. doi:10.1016/j.infbeh.2007.10.011

Liszkowski, U., Carpenter, M., Henning, A., Striano, T., & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science*, *7*(3), 297–307. doi:10.1111/j.1467-7687.2004.00349.x

Loomis, J. L. (1959). Communication, the development of trust, and cooperative behavior. *Human Relations*, *12*(4), 305–315.

Lucca, K., MacLean, E. L., & Hare, B. (2017). The development and flexibility of gaze alternations in bonobos and chimpanzees. *Developmental Science, 1–7*. doi:10.1111/desc.12598

MacCormick, N., & Raz, J. (1972). Voluntary obligations and normative powers. *Proceedings of the Aristotelian Society, 46*, 59–102. doi:10.1093/aristoteliansupp/46.1.59

Melis, A. P., Call, J., & Tomasello, M. (2006). Chimpanzees (*Pan troglodytes*) conceal visual and auditory information from others. *Journal of Comparative Psychology*, *120*(2), 154–162. doi:10.1037/0735-7036.120.2.154

Met, A., Miklósi, Á., & Lakatos, G. (2014). Gaze-following behind barriers in domestic dogs. *Animal Cognition*, *17*(6), 1401–1405. doi:10.1007/s10071-014-0754-z

Michael, J. (2011). Interactionism and mindreading. *Review of Philosophy and Psychology*, *2*(3), 559–578. doi:10.1007/s13164-011-0066-z

Michael, J., & Pacherie, E. (2015). On commitments and other uncertainty reduction tools in joint action. *Journal of Social Ontology*, *1*(1), 89–120.

Michael, J., Sebanz, N., & Knoblich, G. (2015). The sense of commitment: a minimal approach. *Frontiers in Psychology*, *6*, 1–11. doi:10.3389/fpsyg.2015.01968

Miklósi, Á., Polgárdi, R., Topál, J., & Csányi, V. (1998). Use of experimenter-given cues in dogs. *Animal Cognition*, *1*(2), 113–121. doi:10.1007/s100710050016

Moll, H., Carpenter, M., & Tomasello, M. (2007). Fourteen-month-olds know what others experience only in joint engagement. *Developmental Science*, *10*(6), 826–835. doi:10.1111/j.1467-7687.2007.00615.x

Moll, H., Richter, N., Carpenter, M., & Tomasello, M. (2008). Fourteen-month-olds know what "we" have shared in a special way. *Infancy*, *13*(1), 90–101. doi:10.1080/15250000701779402

Moll, H., & Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Developmental Science*, *7*(1), 1–9. doi:10.1111/j.1467-7687.2004.00315.x

Moll, H. & Tomasello, M. (2007a). Cooperation and human cognition: the Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 639–648. doi:10.1098/rstb.2006.2000

Moll, H. & Tomasello, M. (2007b). How 14- and 18-month-olds know what others have experienced. *Developmental Psychology*, *43*(2), 309–317. doi:10.1037/0012-1649.43.2.309

Monderer, D., & Samet, D. (1989). Approximating common knowledge with common beliefs. *Games and Economic Behavior*, *1*(2), 170–190.

Mundy, P. (2003). Annotation: The neural basis of social impairments in autism: the role of the dorsal medial-frontal cortex and anterior cingulate

system. *Journal of Child Psychology and Psychiatry*, *44*(6), 793–809. doi:10.1111/1469-7610.00165

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social Cognition. *Current Directions in Psychological Science*, *16*(5), 269–274. doi:10.1111/j.1467-8721.2007.00518.x

Mundy, P., Sullivan, L., & Mastergeorge, A. M. (2009). A parallel and distributed-processing model of joint attention, social cognition and autism. *Autism Research*, *2*(1), 2–21. doi:10.1002/aur.61

Mutlu, B., Yamaoka, F., Kanda, T., Ishiguro, H., & Hagita, N. (2009). Nonverbal leakage in robots: Communication of intentions through seemingly unintentional behavior. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (pp. 69–76). ACM Press. doi:10.1145/1514095.1514110

O'Madagain, C., & Tomasello, M. (submitted). Joint attention to mental content and the social origin of reasoning. *Manuscript submitted for publication*.

Owens, D. (2006). A simple theory of promising. *Philosophical Review*, *115*(1), 51–77. doi:10.1215/00318108-115-1-51

Pack, A. A., & Herman, L. M. (2006). Dolphin social cognition and joint attention: Our current understanding. *Aquatic Mammals*, *32*(4), 443–460. doi:10.1578/am.32.4.2006.443

Peacocke, C. (2005). Joint attention: Its nature, reflexivity, and relation to common knowledge. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 298–324). Oxford, England: Clarendon Press.

Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. New York, NY: Viking.

Proops, L., & McComb, K. (2009). Attributing attention: the use of human-given cues by domestic horses (*Equus caballus*). *Animal Cognition*, *13*(2), 197–205. doi:10.1007/s10071-009-0257-5

Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: young children's awareness of the normative structure of games. *Developmental Psychology*, *44*(3), 875. doi:10.1037/0012-1649.44.3.875

Range, F., & Virányi, Z. (2011). Development of gaze following abilities in wolves (*Canis Lupus*). *PLoS ONE*, *6*(2), 1–9. doi:10.1371/journal.pone.0016888

Rawls, J. (1955). Two concepts of rules. *The Philosophical Review*, *64*(1), 3–32. doi:10.2307/2182230

Raz, J. (1977). Promises and obligations. In P. M. S.Hacker & J. Raz, (Eds.), *Law, morality, and society: essays in honour of HLA Hart*. (pp. 210–228). Oxford, England: Clarendon Press.

Reddy, V. (1996). Omitting the second person in social understanding. *Behavioral and Brain Sciences*, *19*(01), 140. doi:10.1017/s0140525x00041996

Reddy, V. (2000). Coyness in early infancy. *Developmental Science*, *3*(2), 186–192. doi:10.1111/1467-7687.00112

Reddy, V. (2003). On being the object of attention: Implications for self-other consciousness. *Trends in Cognitive Sciences*, *7*(9), 397–402. doi:10.1016/s1364-6613(03)00191-8

Reddy, V. (2011). A gaze at grips with me. In Seemann, A. (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 137–159). Cambridge, MA: MIT Press.

Reddy, V. (2018). Why engagement? A second person take on social cognition. In A. Newen, L. de Bruin, & S. Gallagher (Eds.), *Oxford handbook of 4e cognition*. Oxford, England: Oxford University Press.

Reddy, V., & Morris, P. (2004). Participants don't need theories: Knowing minds in engagement. *Theory, & Psychology*, *14*(5), 647 – 665. doi:10.1177/0959354304046177

Richerson, P. J., & Boyd, R. (2004). *Not by genes alone*. Chicago, IL: University of Chicago Press. doi:10.7208/chicago/9780226712130.001.0001

Riggs, A. E., & Young, A. G. (2016). Developmental changes in children's normative reasoning across learning contexts and collaborative roles. *Developmental Psychology*, *52*(8), 1236–1246. doi:10.1037/dev0000119

Ristau, C. A. (1991). Before mindreading: Attention, purposes and deception in birds? In A. Whiten (Ed.), *Natural theories of mind. Evolution, development and simulation of mindreading* (pp. 209–222). Cambridge, MA: Basil Blackwell.

Rivas, E. (2005). Recent use of signs by chimpanzees (*Pan troglodytes*) in interactions with humans. *Journal of Comparative Psychology*, *119*(4), 404–417. doi:10.1037/0735-7036.119.4.404

Roberts, G. (2005). Cooperation through interdependence. *Animal Behaviour*, *70*(4), 901–908. doi:10.1016/j.anbehav.2005.02.006

Rollins, P. R., & Snow, C. E. (1998). Shared attention and grammatical development in typical children and children with autism. *Journal of Child Language*, *25*(3), 653–673. doi:10.1017/s0305000998003596

Rossmanith, N., Costall, A., Reichelt, A. F., López, B., & Reddy, V. (2014). Jointly structuring triadic spaces of meaning and action: book sharing from 3 months on. *Frontiers in Psychology*, *5*, 1–22. doi:10.3389/fpsyg.2014.01390

Rousseau, J. (1984). *A discourse on inequality*. New York, NY: Penguin Books. (Original work published 1754)

Rubinstein, A. (1989). The electronic mail game: Strategic behavior under "almost common knowledge." *The American Economic Review*, 385–391.

Salem, M., Kopp, S., Wachsmuth, I., Rohlfing, K., & Joublin, F. (2012). Generation and evaluation of communicative robot gesture. *International Journal of Social Robotics*, *4*(2), 201–217. doi:10.1007/s12369-011-0124-9

Sally, D. (1995). Conversation and cooperation in social dilemmas a meta-analysis of experiments from 1958 to 1992. *Rationality and Society*, *7*(1), 58–92.

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

Saylor, M. M., & Ganea, P. (2007). Infants interpret ambiguous requests for absent objects. *Developmental Psychology*, *43*(3), 696–704. doi:10.1037/0012-1649.43.3.696

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, England: Cambridge University Press.

Searle, J. R. (1995). *The construction of social reality*. New York, NY: Free Press.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. doi:10.1016/j.tics.2005.12.009

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, *1*(2), 353–367. doi:10.1111/j.1756-8765.2009.01024.x

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *18*(9), 668–671.

doi:10.1016/j.cub.2008.03.059

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*(5489), 265–266. doi:10.1038/253265a0

Scanlon, T. (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.

Schachter, S. (1959). *The psychology of affiliation*. Stanford, CA: Stanford University Press.

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

Schilbach, L. (2010). A second-person approach to other minds. *Nature Reviews Neuroscience*, *11*(6), 449. doi:10.1038/nrn2805-c1

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(4), 393–414.

Shiffrin, S. (2008). Promising, conventionalism, and intimate relationships. *Philosophical Review*, *117*(4), 481–524. doi:10.1215/00318108-2008-014

Shteynberg, G. & Apfelbaum, E. P. (2013). The Power of Shared Experience. *Social Psychological and Personality Science*, *4*(6), 738–744. doi:10.1177/1948550613479807

Shteynberg, G., Hirsh, J. B., Apfelbaum, E. P., Larsen, J. T., Galinsky, A. D., & Roese, N. J. (2014). Feeling more together: Group attention intensifies emotion. *Emotion*, *14*(6), 1102–1114. doi:10.1037/a0037697

Sigman, M., Mundy, P., Sherman, T. & Ungerer, J. (1986). Social interactions of autistic, mentally retarded and normal children and their caregivers. *Journal of Child Psychology and Psychiatry*, *27*(5), 647–656. doi:10.1111/j.1469-7610.1986.tb00189.x

Siposova, B., & Carpenter, M. (submitted). A new look at joint attention and common knowledge. *Manuscript submitted for publication*.

Siposova, B., Grueneisen, S., Helming, K. A., Tomasello, M., & Carpenter, M. (in prep.). Common knowledge increased helping in young children. *Manuscript in Preparation*.

Siposova, B., Tomasello, M., & Carpenter, M. (2018). Communicative eye contact signals a commitment to cooperate for young children. *Cognition*.

Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge, England: Cambridge University Press. doi:10.1017/CBO9781139165228

Smetana, J. G., & Braeges, J. L. (1990). The development of toddlers' moral and conventional judgments. *Merrill-Palmer Quarterly*, *36*(3), 329–346.

Smith, C. E., Blake, P. R., & Harris, P. L. (2013). I should but I won't: Why young children endorse norms of fair sharing but do not follow them. *PLoS ONE*, *8*(3), 1–11. doi:10.1371/journal.pone.0059510

Sodian, B., & Kristen-Antonow, S. (2015). Declarative joint attention as a foundation of theory of mind. *Developmental Psychology*, *51*(9), 1190–1200. doi:10.1037/dev0000039

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford, England: Basil Blackwell.

Striano, T., & Bertin, E. (2005). Coordinated affect with mothers and strangers: A longitudinal analysis of joint engagement between 5 and 9 months of age. *Cognition and Emotion*, *19*(5), 781–790. doi:10.1080/02699930541000002

Talmy, L. (2000). *Toward a cognitive semantics* (Vol. 2). Cambridge, MA: MIT press.

Tanner, J. E., & Byrne, R. W. (2010). Triadic and collaborative play by gorillas in social games with objects. *Animal Cognition*, *13*(4), 591–607. doi:10.1007/s10071-009-0308-y

Taylor, C. (1985). *Human agency and language*. Cambridge, England: Cambridge University Press.

Thomas, K. A., DeScioli, P., Haque, O. S., & Pinker, S. (2014). The psychology of coordination and common knowledge. *Journal of Personality and Social Psychology*, *107*(4), 657–676. doi:10.1037/a0037037

Tollefsen, D. P. (2002). Collective intentionality and the social sciences. *Philosophy of the Social Sciences*, *32*(1), 25–50. doi:10.1177/004839310203200102

Tollefsen, D. P. (2005). Let's pretend! Children and joint action. *Philosophy of the Social Sciences*, *35*(1), 75–97. doi:10.1177/0048393104271925

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore, & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Erlbaum.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Tomasello, M. (2011). Human culture in evolutionary perspective. In M. J. Gelfand, C. Y. Chiu, & Y. Y. Hong (Eds.), *Advances in culture and psychology: Vol. 1*. (pp. 5–51). New York, NY: Oxford University Press.

Tomasello, M. (2014). The ultra-social animal. *European Journal of Social Psychology*, *44*(3), 187–194. doi:10.1002/ejsp.2015

Tomasello, M. & Carpenter, M. (2005). The emergence of social cognition. *Monographs of the Society for Research in Child Development*, *70*, 1–122.

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, *10*(1), 121–125. doi:10.1111/j.1467-7687.2007.00573.x

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*(05). doi:10.1017/s0140525x05000129

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*(6), 1454–1463. doi:10.2307/1130423

Tomasello, M. & Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what is new for other persons. *Developmental Psychology*, *39*(5), 906–912. doi:10.1037/0012-1649.39.5.906

Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. *Journal of Human Evolution*, *52*(3), 314–320. doi:10.1016/j.jhevol.2006.10.001

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two key steps in the evolution of human cooperation. *Current Anthropology*, *53*(6), 673–692. doi:10.1086/668207

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech: The beginning of interpersonal communication* (pp. 321–347). New York, NY: Cambridge University Press.

Troisi, A. (2002). Displacement activities as a behavioral measure of stress in nonhuman primates and human subjects. *Stress*, *5*(1), 47–54. doi:10.1080/102538902900012378

Tuomela, R. (1995). *The importance of us*. Stanford, CA: Stanford University Press.

Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *British Journal of Developmental Psychology*, *29*(1), 124–130. doi:10.1348/026151010x532888

VanHecke, A. V., Mundy, P. C., Acra, C. F., Block, J. J., Delgado, C. E. F., Parlade, M. V., ... Pomares, Y. B. (2007). Infant joint attention, temperament, and social competence in preschool children. *Child Development*, *78*(1), 53–69. doi:10.1111/j.1467-8624.2007.00985.x

Vanderschraaf, P., & Sillari, G. (2014). Common knowledge. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Retrieved from https://plato.stanford.edu/archives/spr2014/entries/common-knowledge

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., ... Wahn, B. (2017). Joint action: Mental representations, shared information and general mechanisms for coordinating with others. *Frontiers in Psychology*, *07*. doi:10.3389/fpsyg.2016.02039

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. doi:10.1016/j.cognition.2016.05.002

Virányi, Z., Topál, J., Gácsi, M., Miklósi, Á. & Csányi, V., (2004). Dogs respond appropriately to cues of humans' attentional focus. *Behavioural Processes*, *66*(2), 161–172. doi:10.1016/j.beproc.2004.01.012

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children*, *23*(3), 34–41.

Wilkinson, A., Mandl, I., Bugnyar, T., & Huber, L. (2010). Gaze following in the red-footed tortoise (*Geochelone carbonaria*). *Animal Cognition*, *13*(5), 765–769. doi:10.1007/s10071-010-0320-2

Wolf, W., Launay, J., & Dunbar, R. I. M. (2015). Joint attention, shared goals, and social bonding. *British Journal of Psychology*, *107*(2), 322–337. doi:10.1111/bjop.12144

Wyman, E., Rakoczy, H., & Tomasello, M. (2013). Non-verbal communication enables children's coordination in a "Stag Hunt" game. *European Journal of Developmental Psychology*, *10*(5), 597–610. doi:10.1080/17405629.2012.726469

Yang, G.-Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., ... Wood, R. (2018). The grand challenges of science robotics. *Science Robotics*, *3*(14). doi:10.1126/scirobotics.aar7650

Zahavi, D. (2008). Simulation, projection and empathy. *Consciousness and Cognition*, *17*(2), 514–522. doi:10.1016/j.concog.2008.03.010

Zahavi, D. (2011). Empathy and direct social perception: A phenomenological proposal. *Review of Philosophy and Psychology*, *2*(3), 541–558. doi:10.1007/s13164-011-0070-3

Zahavi, D. (2015). Self and other: From pure ego to co-constituted we. *Continental Philosophy Review*, *48*(2), 143–160. doi:10.1007/s11007-015-9328-2

Zimmermann, F., Zemke, F., Call, J., & Gómez, J. C. (2009). Orangutans (*Pongo pygmaeus*) and bonobos (*Pan paniscus*) point to inform a human about the location of a tool. *Animal Cognition*, *12*(2), 347–358. doi:10.1007/s10071-008-0194-8

Ethics applications approvals

University of St Andrews

University Teaching and Research Ethics Committee
Sub-committee

31 October 2014

Ethics Reference No: *Please quote this ref on all correspondence*	PS11204
Project Title:	Inferences children make following different types of eye contact in a 'Stag Hunt' cooperative game
Researchers' Names:	Barbora Siposova, Professor Malinda Carpenter, Professor Michael Tomasello, Elena Rossi and Kristina Schilke
Supervisor:	Professor Malinda Carpenter

Thank you for submitting your application which was considered at the Psychology & Neuroscience School Ethics Committee meeting on the 7th October 2014. The following documents were reviewed:

1. Ethical Application Form 22/10/2014
2. Details of information and consent procedure at the MPI-EVA 22/10/2014
3. Consent Form 22/10/2014
4. Debriefing Form 22/10/2014
5. Letter to Parents 22/10/2014

The University Teaching and Research Ethics Committee (UTREC) approves this study from an ethical point of view. Please note that where approval is given by a School Ethics Committee that committee is part of UTREC and is delegated to act for UTREC.

Approval is given for three years. Projects, which have not commenced within two years of original approval, must be re-submitted to your School Ethics Committee.

You must inform your School Ethics Committee when the research has been completed. If you are unable to complete your research within the 3 three year validation period, you will be required to write to your School Ethics Committee and to UTREC (where approval was given by UTREC) to request an extension or you will need to re-apply.

Any serious adverse events or significant change which occurs in connection with this study and/or which may alter its ethical consideration, must be reported immediately to the School Ethics Committee, and an Ethical Amendment Form submitted where appropriate.

Approval is given on the understanding that the 'Guidelines for Ethical Research Practice' https://www.st-andrews.ac.uk/utrec/guidelines/ are adhered to.

Yours sincerely

Convenor of the School Ethics Committee

Ccs Prof M Carpenter (Supervisor)
School Ethics Committee

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

University of St Andrews
from first to foremost

600 YEARS
1413 – 2013

Project Title	Inferences children make following different types of eye contact in a 'Stag Hun' cooperative game
Researchers' Names	Barbora Siposova, Professor Malinda Carpenter, Professor Michael Tomasello, Elena Rossi, Anna Maire Seidel, Nathalie Dosch and Rahel Döhring-Jahn
Supervisor	Professor Malinda Carpenter
Department/Unit	School of Psychology & Neuroscience
Ethical Approval Code (Approval allocated to Original Application)	PS11204
Original Application Approval Date	22 October 2014
Amendment Application Approval	03 April 2015

Ethical Amendment Approval

Thank you for submitting your amendment application which was considered by the Psychology & Neuroscience School Ethics Committee on the 3rd April 2015. The following documents were reviewed:

1. Ethical Amendment Application Form 03/04/2015

The University Teaching and Research Ethics Committee (UTREC) approves this study from an ethical point of view. Please note that where approval is given by a School Ethics Committee that committee is part of UTREC and is delegated to act for UTREC.

Approval is given for three years from the original application only. Ethical Amendments do not extend this period but give permission to an amendment to the original approval research proposal only. If you are unable to complete your research within the original 3 three year validation period, you will be required to write to your School Ethics Committee and to UTREC (where approval was given by UTREC) to request an extension or you will need to re-apply. You must inform your School Ethics Committee when the research has been completed.

Any serious adverse events or significant change which occurs in connection with this study and/or which may alter its ethical consideration, must be reported immediately to the School Ethics Committee, and an Ethical Amendment Form submitted where appropriate.

Approval is given on the understanding that the 'Guidelines for Ethical Research Practice' (http://www.st-andrews.ac.uk/media/UTRECguidelines%20Feb%2008.pdf) are adhered to.

Yours sincerely

Convenor of the School Ethics Committee

Cc School Ethics Committee

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

University of St Andrews

University Teaching and Research Ethics Committee
Sub-committee

8 July 2015

Ethics Reference No: *Please quote this ref on all correspondence*	PS11669
Project Title:	The role of different types of eye contact in a cooperative game
Researchers' Names:	Barbora Siposova, Professor Malinda Carpenter and Professor Michael Tomasello
Supervisor:	Professor Malinda Carpenter

Thank you for submitting your application which was considered at the Psychology & Neuroscience School Ethics Committee meeting on the 30th June 2015. The following documents were reviewed:

1. Ethical Application Form 07/07/2015
2. Consent Form 07/07/2015
3. Debriefing Postcard 07/07/2015
4. Letter to Parents 07/07/2015

The University Teaching and Research Ethics Committee (UTREC) approves this study from an ethical point of view. Please note that where approval is given by a School Ethics Committee that committee is part of UTREC and is delegated to act for UTREC.

Approval is given for three years. Projects, which have not commenced within two years of original approval, must be re-submitted to your School Ethics Committee.

You must inform your School Ethics Committee when the research has been completed. If you are unable to complete your research within the 3 three year validation period, you will be required to write to your School Ethics Committee and to UTREC (where approval was given by UTREC) to request an extension or you will need to re-apply.

Any serious adverse events or significant change which occurs in connection with this study and/or which may alter its ethical consideration, must be reported immediately to the School Ethics Committee, and an Ethical Amendment Form submitted where appropriate.

Approval is given on the understanding that the 'Guidelines for Ethical Research Practice' https://www.st-andrews.ac.uk/utrec/guidelinespolicies/ are adhered to.

Yours sincerely

Convenor of the School Ethics Committee

Ccs Prof Malinda Carpenter (Supervisor)
School Ethics Committee

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071