SPONTANEOUS VISUOSPATIAL PERSPECTIVE-TAKING IN HUMANS

by

Martin Freundlieb

Submitted to

Central European University

Department of Cognitive Science

In partial fulfillment of the requirements for the degree of Doctor of Philosophy

Supervisor: Natalie Sebanz

Secondary Supervisor: Ágnes M. Kovács

Declaration of Authorship

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

The present thesis includes work that appears in the following papers:

Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2016). When Do Humans Spontaneously Adopt Another's Visuospatial Perspective? *Journal of Experimental Psychology: Human Perception and Performance,* Vol 42(3), 401-412. http://dx.doi.org/10.1037/xhp0000153

Freundlieb, M., Sebanz, N., & Kovács, Á. M. (2017). Out of your sight, out of my mind: knowledge about another person's visual access modulates spontaneous visuospatial perspective-taking. *Journal of Experimental Psychology: Human Perception and Performance*.

Freundlieb, M., Kovács, Á. M., & Sebanz, N. (*under revision*). Reading your mind while you are reading – Evidence for spontaneous visuospatial perspective-taking during a semantic categorization task. *Psychological Science*.

M. Freundlieb

Martin Freundlieb

ABSTRACT

Perspective-taking is one of the fundamental building blocks enabling humans to successfully understand and relate to others in a large variety of social interactions. Yet, there are many open questions about whether, when and how instances of visuospatial perspective-taking occur during social interactions. This dissertation investigates the phenomenon of spontaneous visuospatial perspective-taking in humans. Chapter 1 discusses visuospatial perspective-taking in the wider context of social cognition abilities. The study presented in Chapter 2 explored the underlying factors as well as boundary conditions that characterize the spontaneous adoption of another person's visuospatial perspective (VSP). The results showed that participants spontaneously adopted a differing VSP, given there was an intentionally acting agent alongside of them. Chapter 3 investigated whether knowledge about another's visual access systematically modulates spontaneous VSP-taking. In two experiments we found that knowledge about another person's visual access indeed modulated the spontaneous integration of another person's VSP into one's own action planning. Specifically, our findings showed that participants only adopted the other person's VSP if he had unhindered visual access to the stimuli but regardless of whether or not he performed the same task or a different task. Finally, the study presented in Chapter 4 probed whether spontaneous VSP-taking also occurs in mental space where another person's perspective matters for mental activities rather than for physical actions. In three experiments participants reliably adopted the VSP of a confederate in the context of a semantic categorization task that involved reading words. Taken together, these studies show that we spontaneously take into account how somebody else perceives the environment, even in situations where we are not asked to do so, and we are likely not aware of doing so. This suggests that humans are endowed with a basic sensitivity to their conspecifics' viewpoints.

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisors Natalie and Ági. I feel so lucky that throughout the last five years I have been enjoying the continuous support and supervision not from one, but actually two *main* supervisors. And although I deeply admire each of your research individually, it was truly inspiring to see how perfectly the two of you were working together – in your common effort to supervise me. You are a great team! Thanks for letting me be a part of it in the past five years.

My next thanks goes out to my parents, without whom none of this would have been possible. But I am not only thankful for the many opportunities that you have been giving me; I also appreciate the way the two of you accompany me through my life: unceasingly encouraging, at all times reassuring, open for discussion and, finally, always up for some good laughter.

The same holds true both for my siblings and for my age-old friends (Marlte!); thank you for your companionship and your friendship that has kept me smiling for so much longer than the last five years. But I also cherish the new ties that I got to build throughout my stay in Budapest; Holger, Hannes, Denis, Luke, Clément, Ian, Stefan, Vish, Oktávio, Tamás and Judit – you have all reminded me in your individual ways (vapenaish) that this period of my life has been so much more than just going to school. Speaking of school; I very much appreciate the continuous support I received from CEU in general and from Réka, Eszter, Veronika, Anna Mária, Fruzsi, Dávid and Matteo in particular; thank you!

Finally, I want to finish these acknowledgments with what has - without a doubt - been most important for me during my PhD period in Budapest. And that is You, Anna. I am so thankful that I have met you here. This thesis is dedicated to you.

Table of Contents

List of Figures

Introduction

Perspective-taking is one of the fundamental building blocks enabling humans to successfully understand and relate to others in a large variety of social interactions. Whether we lead (or follow) a yoga class, guide an avatar through a complex map in a video game, or straightforwardly want to draw our interlocutor's attention to the fact that there is some spinach stuck between her[1] teeth – our daily life constantly confronts us with a plethora of perspectives that are virtually never perfectly aligned with our own visual perspective. On top of that, many situations do not allow us to contemplate the divergence of different perspectives but instead require us to instantaneously react and adjust our actions so that multiple perspectives are taken into account. Take the quarterback who is just about to be tackled by an opponent while aiming for the perfect pass to the receiver who might be more than 70 yards away once she will catch the football.

This ability to flexibly adopt another person's perspective, e.g. while jointly attending to an object, is crucial for implementing joint goals as well as for the successive coordination of actions (Bratman, 1992). Fortunately, we seem to be equipped with an incredibly sophisticated cognitive machinery that allows us to track and flexibly integrate varying perspectives on multiple levels. This is reflected, for example, in our fundamental comprehension that one and the same thing can be viewed or construed differently *depending on the chosen standpoint* – whether this requires an epistemic, conceptual, affective or visual perspective (see Perner et al., 2003).

Accumulating evidence indicates that humans readily compute whether somebody else can or cannot see a target object (referred to as *visual perspective-taking*, see Samson,

[1] For reasons of simplicity I will mostly use just one, namely the female pronoun, which is meant to include all genders in this dissertation.

Apperly, Braithwaite, Andrews, & Bodley Scott, 2010; Ramsey, Hansen, Apperly, & Samson, 2013; Baker, Levin, & Saylor, 2015). However, one could argue that human perspective-taking entails more than that. For instance, when you pluck up the courage to tell your friend that there is spinach between her teeth, you need to take into consideration that her left is your right side, and vice versa. This means that, besides representing *whether* another person can or cannot see a certain object (visual perspective-taking), we also need to be able to compute *how* objects are arranged and what they look like from another's point of view. This ability to take into account the location of objects relative to others, and infer what these objects look like from their perspective is called *visuospatial perspective-taking* (Kessler & Rutherford, 2010).

There is an on-going controversy about whether, when and how instances of visuospatial perspective-taking occur. For example, it is not known whether humans can adopt another's visuospatial perspective (VSP) by all means *spontaneously*, that is, without being (explicitly) prompted to do so (Tversky & Hard, 2009; Surtees, Butterfill, & Apperly, 2012). Furthermore, some proposals argue that judging relative locations from a particular perspective requires a transformation of one's own reference frame and is therefore assumed to be effortful, leading to more errors and increased response latencies (Kessler & Thomson, 2010; Surtees, Apperly, & Samson, 2013). But is the adoption of somebody else's VSP really always detrimental for task performance or are there specific circumstances in which VSP-taking can actually have a positive effect on performance? Moreover, what could be the mechanisms underlying VSP-taking and do we have evidence to claim that they are exclusively 'social in their nature' (in the sense that they are only triggered in the presence of an intentional agent; cf. Heyes, 2014)? Finally, is there a special link between adopting another's VSP and performing *physical* actions together (cf. Creem-Regehr, Gagnon, Geuss, & Stefanucci, 2013)? Or does VSP-taking also extend to mental space where spatial relations matter for mental activities rather than for physical actions? This work addresses these issues in a systematic manner and, in doing so, contributes to the larger question whether there are some uniquely human capacities that may explain the complex social structures of human societies.

Research Aims

This thesis investigates spontaneous visuospatial perspective-taking in humans. It begins with a theoretical discussion embedding the phenomenon of visuospatial perspective-taking in the wider context of social cognition abilities (Chapter 1). In three studies (cf. Chapter 2, 3, and 4), I will then investigate the underlying factors as well as boundary conditions that characterize the adoption of another person's visuospatial perspective (VSP) during social interactions. The aim of this thesis is to extend our prior knowledge on perspective-taking in multiple ways.

First of all, I will explore whether humans are able to adopt another's VSP by all means spontaneously, i.e. without being prompted to do so (Chapter 2). Additionally, I will discuss whether VSP-taking only manifests in a deterioration of task performance or whether there are circumstances in which the adoption of another person's VSP can actually facilitate performance (Chapter 2 and 3). Directly addressing the on-going debate about alternative (lower-level) explanations for perspective-taking effects, I will then ask whether the evidence for VSP-taking in our studies is based on general attention-shifting mechanisms (cf. Heyes, 2014), or on knowledge about the visual access of the other agent (Chapter 3 and 4). Finally, I will investigate whether VSP-taking is restricted to situations involving action planning in the context of physical interactions with objects, or whether it also extends to mental space where spatial relations matter for cognitive processes rather than for physical actions (Chapter 4).

In a nutshell, I will thus address the following research questions:

- What are the building blocks alongside visual perspective-taking enabling humans to understand and relate to others? (Chapter 1)
- Do humans spontaneously adopt another's visuospatial perspective and if so, what are the boundary conditions that can lead to or restrict spontaneous visuospatial perspective-taking? (Chapter 2)
- Does knowledge about another's ability to see the object on which we have a different perspective affect VSP-taking? Or is VSP-taking based on lower level processes that are independent of encoding how an object is perceived by another person?
(Chapter 3)
- Does spontaneous VSP-taking also occur in mental space where another person's perspective matters for mental activities rather than physical actions? (Chapter 4)
- How can we describe spontaneous VSP-taking at a cognitive level? (Chapter 5)

Chapter 1
Building blocks for Understanding and Relating to Others

Imagine you are working as a caregiver in a nursing home. Since you started the job not long ago you are still spending most of your days getting to know the residents of the nursing home and, importantly, figuring out their individual needs. One afternoon you enter Gábor's room and immediately get the impression that something is bothering him and that he is already quite agitated. Gábor is a 90 year-old man with dementia that your colleagues already praised for being a particularly good-tempered resident. However, you also know from your colleagues that it is of the utmost importance for Gábor to watch the 9 o'clock news every day. As you enter his room you see how his eyes furiously wander around the bed in which he is lying. The bed is covered with books and newspaper articles and, as Gábor starts to moan, you spot the remote control on his duvet, partially covered by one of the newspapers. Moreover, it catches your attention that Gábor (or an absentminded colleague, for that matter) must have put on his watch upside-down. As a consequence, the watch appears to show 9 o'clock, while the actual time is only approaching 3:30 p.m. You quickly rush over to the bed, retrieve the supposedly lost remote control, call Gábor's attention to the fact that his watch is upside down and receive a big smile in return.

Even if this specific example comes across as a bit constructed, our daily lives provide us with myriads of varying situations in which we quickly need to figure out both the right context but also the specific mental states of other people in order to successfully interact with them. Looking into the details of these situations allows us to highlight the kind (and amount) of cognitive challenges and complexities underlying such every-day-like social interactions.

For example, the above story illustrated how, paired with the evidence for Gábor's uneasiness (and against the background of his preference for watching the news), we first followed his wandering eye gaze in order to arrive at the conclusion that he must be searching for something. Furthermore, we grasped how, from his restrained position in bed, he can simply not see the remote control lying under one of the newspaper articles. Finally, we noticed that his watch is upside down and that from Gábor's perspective, it therefore actually looks a bit like it is 9 o'clock in the evening rather than 3:30 in the afternoon.

Crucially, on the basis of all of these observations, it becomes a lot easier first to understand and relate to the specific condition Gábor finds himself in (i.e. stressed out and irritated), but then also to properly interact with him and help him in a tailored manner. In contrast, adequately helping Gábor only on the basis of the explicit information received by the colleagues would probably turn out to be much more difficult.

The ability to handle situations like these uniquely shapes the way in which we interact with others and hence, how we become what some argue to be the most social among the animals (Herrmann, Call, Hernàndez-Lloreda, Hare, Tomasello, 2007). An even closer look at the cognitive building blocks that provide us with these capacities reveals a long chain of processes underlying what one might mundanely call 'a successful (social) interaction with Gábor'. While a comprehensive review of these processes would be beyond the scope of this work, in this chapter I will mark out and discuss some of the important building blocks - alongside visual perspective-taking - that enable social interactions, in order to embed the phenomenon of VSP in the broader context of social cognition research.

What are the most important building blocks that enable humans to interact with each other in such a sophisticated manner as was illustrated in the earlier example? Undoubtedly, one crucial aspect in the nursing home example was that the observer was able to relate to (or trace) the specific thoughts, beliefs and desires of the resident. One could go as far as to say that the observer successfully grasped what the resident was pondering on – *in his mind*. Analogously, this chapter starts with a discussion on the extremely sophisticated capacity of humans to successfully *read the minds of others*.

However, a closer look at the variety of cognitive challenges in the above example also revealed that other phenomena significantly contributed to the success of the act of mindreading and the subsequent ability to help. For instance, it was crucial that the observer followed the wandering gaze of the resident (to understand that the resident was searching for something), that she attended to where the resident attended to (e.g. in order to restrict the search area to the resident's duvet), and finally, that she took into account that both of them operated from differing viewpoints (to understand that only she could see the remote control while the resident could not, and that from the perspective of the resident the watch appeared as if it showed a different time). One might even argue that it was the aggregate of these close observations that substantially contributed to the particular alignment of those two minds in the introductory example. Correspondingly, this theoretical chapter will discuss the phenomena of a Theory of Mind, Gaze Following, Joint Attention, Task Co-Representation and finally, Visual Perspective-Taking.

1.1 Theory of Mind

Probably among the most complex building blocks for social interactions is the ability to understand that other people, just like ourselves, hold certain mental states that motivate their behavior. This ability to attribute mental states to others - such as what they think, feel, believe or see - has been referred to as *mentalizing*, but it has also been taken as an indication for possessing the capability for a "*Theory of Mind*".[2] In their influential paper, Premack and Woodruff (1978) coined the term "Theory of Mind" when they raised the possibility that chimpanzees might also be able to ascribe mental states to others. More precisely, they listed arguments that would suggest that chimpanzees can solve tasks in which they are required to take into account another's mental state, rather than basing their behavior on non-mental physical constraints of a situation (Premack & Woodruff, 1978).

Likewise, when trying to understand the motivation behind and predict the behavior of other people, we often rely on our ability to reason about others' mental states. For example, when thinking about another person we often wonder what that person is thinking, feeling or believing at any given moment. Some have argued that these kind of mentalizing activities can only be achieved through the efficient use of a Theory of Mind (Premack & Woodruff, 1978). So how can it be experimentally measured whether somebody possesses a Theory of Mind?

Considering the large array of mental states - such as percepts, beliefs, goals, or desires - that can be attributed to others, the so called *representational* mental states seem to have drawn the attention of researchers that are interested in Theory of Mind in a particular manner. The reason behind this is that representational mental states mean to represent the

[2] Notice that throughout this dissertation, the terms 'Theory of Mind', 'mentalizing', 'mindreading' and 'mental state attribution' will be used interchangeably, and that I do not want to commit to any implications of these terms unless specified otherwise.

external world (such as 'I believe it is sunny outside') without necessarily matching the actual state of the world (e.g. when it is actually raining cats and dogs at the moment). This turns out to be a crucial feature for investigating how social agents attribute mental states to others. To briefly illustrate this point, imagine that the mental states of two agents were exactly identical. If these agents were to predict each other's behavior in the future, from a third-person point of view it would become impossible to disentangle whether their predictions would be based on their own or on the other agent's mental state (Dennett, 1978).

Representational mental states offer a unique solution to this problem. Specifically, one can develop scenarios with them in which two agents share a propositional attitude (e.g. about the current weather) but differ in assigning truth-values to it. For instance, subject A believes it is sunny outside, while subject B knows it is not the case that it is sunny outside. This way, one can test whether subject A can ascribe an epistemic state to subject B *that differs from her own* and furthermore, whether she can use this ascription to successfully predict the subsequent behavior of subject B.

Traditionally, the literature on Theory of Mind has been focusing a lot on the phylogenetic and ontogenetic time-course of understanding representational mental states. For example, in the developmental literature, the capability to understand and compute representational mental states has, for a long time, been seen as a hallmark in the development of an adult-like Theory of Mind (Wellman, Cross, & Watson, 2001). However, this assumption has been put into question, given new evidence both on the actual competencies infants show already at an early age when it comes to representing others' mental states (cf. Onishi, & Baillargeon, 2005, and see detailed discussion on p. 11–13), as well as the difficulties human adults exhibit when facing others with differing mental states (cf. Keysar, Barr, Balin, & Brauner, 2000, and see detailed discussion on p. 13–16). In consequence, these new pieces of evidence argue for a re-assessment of what has traditionally been seen as a litmus tests for Theory of Mind capabilities (the so called *false belief task,* see next section), and for a specification of what can rightfully be claimed to belong to a Theory of Mind repertoire.

In the next three subsections, I will briefly portray the most important milestones in the history of Theory of Mind research and summarize what has recently led to a paradigm shift in the debate on how to capture Theory of Mind abilities.

The Classic False-Belief Task

To test the developmental trajectory of Theory of Mind abilities, Wimmer and Perner (1983) came up with what is known as the *classic false-belief task*. Since then, different versions of this false-belief task have been investigated among several cultures (Barret et al., 2013) as well as with typical and atypical human populations (Baron-Cohen, Leslie, & Frith, 1985). However, the central idea behind the false-belief task has always remained the same: the false-belief-task examines whether subject A is able to ascribe a certain representational mental state to subject B and to use that ascription to predict B's following behavior. Crucially, A and B hold different mental states or, more specifically, they hold two differing beliefs about what the true state of their environment is.

According to the classic version of the task children are presented with the following story: M. puts her chocolate in the blue cupboard before going outside to play. While M. is away, M's mother replaces the chocolate from the blue cupboard to the green cupboard. After a while M. returns home, at which point of the story the children are asked by the experimenter where M. will look for her chocolate (cf. Wimmer & Perner, 1983).

Up until the age of three, children systematically respond that M. will look in the green cupboard for her chocolate, indicating that they do not differentiate between the actual state of the world (where the chocolate is indeed in the green cupboard), and what M. beliefs to be true (that it is in the blue cupboard). Typically, it is only from the age of four years that children will respond correctly by saying that M. will look in the blue cupboard as this was the last place where M. saw the chocolate, inferring that M. now holds a false belief about the true location of the chocolate (for a comprehensive review see Wellman, Cross, & Watson, 2001).

Across a wide range of variations of the false-belief paradigm, it was then argued that typically developing children between the age of three and five must consequently undergo a sudden improvement of their Theory of Mind abilities – that is, when they are first able to successfully ascribe a false-belief to another agent (Wellman, et al., 2001). Hence, for a long time passing the false-belief paradigm has been seen as a universal milestone in the development of a Theory of Mind.

However, at the same time these findings did not give an explanation about the nature of 3-year-olds' *failure* in the false-belief task. Specifically, they left the question open whether 3-year-olds fail the task because they do not possess the cognitive resources that are necessary to process false-belief-reasoning or because the task is simply not suitable for children younger than four years (cf. Bloom, & German, 2000). In turn, if the classic false-belief task were indeed too complex in its structure, then other (less taxing) tasks should be able to show false-belief-reasoning even in children younger than four years of age.

Early Signs of Theory of Mind Competencies

In a substantial shift of the debate about the age at which children acquire the ability to capture representational mental states, Onishi and Baillargeon (2005) demonstrated that infants as young as 15-months-old display signs of understanding others' false beliefs. In their study, infants observed an agent placing an object in one of two boxes. In the following, a barrier occluded the visual access of the agent such that the agent could not see how the object actually moved over to the other box. Thus, only the infant observed the displacement of the object, leaving the agent with the false belief that the object is still in its original place. Finally, the occluder was removed and the agent either reached to the box where she had originally placed the object (i.e. in accordance with the agent's false belief), or to the box in which the object really was (i.e. in violation of the agent's false belief).

Importantly, Onishi and Baillargeon (2005) used a violation-of-expectation manipulation in order to investigate whether the 15-month-olds differentiated between these two conditions. And indeed, the looking times of the 15-month-olds demonstrated that they were more surprised (indicated by longer looking times) when the agent reached into the box the object 'secretly' had moved into, compared to when the agent reached into the box in which she had originally placed the object (Onishi, & Baillargeon, 2005).

Onishi and Baillargeon (2005) took this as evidence that even from a very young age, infants comprehend representational mental states of others in order to better explain their behavior. Furthermore, other studies have corroborated that before children are able to process someone else's belief *explicitly,* they have - already at three years of age - acquired the ability to process someone else's belief *on an implicit level* (Clements & Perner, 1994; Garnham & Ruffman, 2001), just as in the study by Onishi and colleagues (2005).

The question was then just how young infants are when they first show these implicit signs of attributing beliefs to others. A few years later another paradigm provided one answer to this question through modifying the complex original false-belief task in order to make it better suitable for infants. Kovács and colleagues (2010) used an object-detection task to investigate the effects of somebody else's belief on 7-month-old infants. In their paradigm, participants watched a short animation sequence in which an agent observed a ball rolling behind an occluder. In the next scene, the ball continued to roll from behind the occluder to then either exit the scene or, after stopping half-way, return to its original position behind the occluder. Importantly, the agent in the animation sequence could either observe the relocation of the ball (and thus, hold the same belief as the participant), or she would exit the scene earlier and not know about the relocation of the ball (hence, holding a different belief than the participant). In the final scene, the occluder fell over to then always show that the ball was absent.

The results showed that the 7-month-old infants were more surprised (that is, they looked longer) when both the participant and the agent thought that the ball would be present, compared to when they both believed it to be absent. Crucially though, the infants also looked longer when they themselves knew that the ball would be absent, but the agent believed that it would be present. According to Kovács and colleagues (2010), these findings clearly demonstrate that already by the age of 7-months human infants show a sensitivity to other people's mental states (in this case: beliefs, Kovács, Teglas, & Endress, 2010; Surian, Caldi, & Sperber, 2007; but see also Sodian, Thoermer, & Metz, 2007 for 14-month-olds representing somebody else's visual perspective; and see Song, & Baillargeon, 2008 for 14.5 month-olds representing others' false perceptions).

However, just because humans seem to be sensitive towards others' representational mental states from a very young age on, one should not mistake mindreading for being a cognitively effortless phenomenon that humans master to perfection once they reach adulthood. In fact, there is accumulating research showing how difficult the integration of somebody else's mental state is and that even as adults humans are quite prone to making errors while applying their mentalizing abilities.

(Late) Signs of Theory of Mind Difficulties in Adults

Being able to adequately integrate someone else's diverging mental state and resisting the interference from one's own knowledge (that is, not being sensitive to an *egocentric bias)* has been shown to require cognitively effortful processes that take time to develop (see Birch & Bloom, 2004; Epley et al., 2004). Even for adults task performance has been found to be slower and more error-prone when participants needed to make judgments about another person's false (compared to true) belief (German & Hehman, 2006; Back & Apperly, 2010, but see Cohen & German, 2009). In fact, humans appear to be so strongly affected by egocentric biases, that it almost seems as if they were *cursed* to disregard other people's perspectives and instead stick with whatever knowledge they have previously acquired themselves (see Birch & Bloom, 2007).

In a series of experiments, Keysar and colleagues (2000; 2003) tried to further investigate this egocentric bias. Specifically, they examined the difficulty that adults have in incorporating other's diverging *visual* perspectives. What they found was that adults sometimes failed to use their conceptual knowledge for other people's visual perspectives when facing an online referential communication game in which their own and another person's perspective should be taken into account at the same time. The protocol was as follows.

Participants viewed a 4 x 4 grid containing various objects in different slots. By following the instructions of a 'director' (a confederate) participants then needed to move certain objects around the grid. Crucially, certain slots in the grid were selectively occluded so that the director could only see *some* of the objects while the participant always maintained visual access to *all* of the objects. Thus, there was an information mismatch between the director and the participant. In the critical experimental trials, participants then needed to utilize their ancillary information about the director's perspective in order to appropriately interpret his instructions.

For example, the grid would contain - amongst other things - three similar objects that differed in their size (say, a small, a medium and large sized ball). While all objects would be clearly visible to the participants, the smallest object would be selectively occluded from the point of view of the director. The director would then ask the participant to move "the small ball", which, from the director's perspective, actually referred to the medium sized ball.

Although clearly capable of understanding that the director would have a differing perspective to their own, the results showed that adult participants occasionally failed to use this information when interpreting the director's instructions. Specifically, eye-tracking data revealed that participants first looked at the object that the director could not see (i.e. the small ball). Furthermore, in 20% of the cases participants even reached for that object without even realizing that it was impossible for the director to refer to it (Keysar et al., 2000).

As several follow–up studies then replicated those results (see Keysar, Lin, and Barr, 2003; Epley et al., 2004), Keysar suggested that these pieces of evidence essentially prove that humans are very much prone to an egocentric bias –and that the integration of another person's perspective must therefore be understood as a cognitively demanding - rather than effortless - process (Keysar et al., 2003).

Thus, in light of this specific line of research it seems as if even as adults we frequently disregard others' perspectives in interactive contexts to instead fall back to our own point of view (Keysar et al., 2000; 2003; Epley et al., 2004). What's more, a closer look on the actual differences between children and adults regarding their automatic and controlled perspective-taking capabilities reveals that "[…] egocentrism isn't outgrown so much as it is overcome each time a person attempts to adopt another's perspective" (Epley et al., 2004:

765). These findings therefore suggest that - rather than being automatically very fine–tuned towards other people's visual perspectives - humans are also drawn to their egocentric perspective and might have to engage in effortful processing in order to disregard and / or overcome it.

Yet, while humans seem to have difficulties *on a conceptual level* integrating or making sense of another person's visual perspective (e.g., 'the other person cannot see the object and hence, cannot refer to it', cf. grid-perspective-taking task above), they are exceptionally good at picking up others' perspectives *on a perceptual level,* for example when it comes to tracking what others are gazing at.

The next section discusses how studies on gaze following can effectively demonstrate how much humans are drawn to where other people look, how gaze following develops during infancy, and what benefits it can generate in social interactions.

1.2 Gaze Following

In contrast to other species, the human eye with its black pupil on a white sclera has evolved to provide maximal information about where somebody is gazing. And indeed, humans seem to be particularly sensitive to picking up where other people are looking (Von Grünau & Anston, 1995). Following the gaze of another person does not only provide an important source of information with regards to what is potentially interesting or dangerous in a common environment, it also proves to be an incredibly powerful tool to acquire social information (Langton & Bruce, 2000).

Evidence for Gaze Following in Adults and Infants

In a seminal study, Driver and colleagues (1999) tested the automaticity of orienting towards the direction of a seen gaze. In their paradigm, participants did a peripheral letter discrimination task (i.e. target letters appeared on one or the other side of the screen and participants were asked to respond to the letters as quickly as possible) while seeing a computerized face in the center of the screen.

Crucially, throughout the experiment, Driver et al. (1999) manipulated the direction of the gaze of the computerized face such that in some trials the direction of the gaze would coincide with the location of the letter, whereas in other trials it would not. Importantly, they emphasized that participants could completely ignore the face as its gaze direction was uninformative of the location of the target letter (Driver et al., 1999).

Nonetheless, their results revealed that participants were faster responding to stimuli appearing at the gazed at location compared to stimuli appearing at the non-gazed at location. Moreover, in another experiment, Driver and colleagues (1999) showed evidence for such automatic gaze following even when participants knew that target letters were four times as likely to appear on the *opposite* side of the gazed-at direction (Driver et al., 1999).

This demonstrates that, at least in simple visual displays, observing somebody else looking into a certain direction elicits a strong and reflexive orienting towards the direction of the seen gaze – even if there is no relationship between the gazing person and the task at hand (Driver et al., 1999), and even if following another's gaze is detrimental to task performance (Driver et al., 1999; see also Ricciardelli, Bricolo, Aglioti, & Chelazzi, 2002).

But it is not only in simple visual displays that we tend to show this reflexive pattern of following other people's gaze. Investigating the influence of another person's gaze direction in real world scenes, Castelhano and colleagues (2007) showed that observers were highly sensitive to an agent's direction of gaze and effectively used it to help guide their own eye movements.

Specifically, in their paradigm, participants watched an unfolding story of a janitor cleaning a university office. With the help of an eye-tracker Castelhano et al. (2007) then measured participants' eye movements in order to assess how much participants were influenced by the gaze direction of the janitor. Their results revealed that participants were indeed very much affected by where the janitor looked during the unfolding story. For instance, objects that the janitor gazed at were fixated sooner, more often, for a longer duration and on a larger percentage of trials than any other scene region (Castelhano, Wieth, & Henderson, 2007).

These findings demonstrate that, regardless of the complexity of our visual environment (i.e. irrespective of whether we face simple visual displays or more complex real-world scenarios), what another person is looking at has a strong influence on what we will look at. But how does this ability develop and is it only triggered in the presence of other social agents?

Besides being called by our name, how are we able to infer communicative intention towards us? Eye contact provides a particular effective means to deliver and exchange socially relevant information. It communicates to the gazed-at person that he or she is the addressee of an informative intent and that the upcoming gaze of the addresser is going to be meaningful (Csibra & Gergely, 2009). From an early age on humans follow the gaze direction of their task partner in order to make sense of their environment and to develop an understanding of the social world (Csibra & Volein, 2008).

In real-life-like situations gaze following has been shown to occur in infants as young as 3-6 months (D'Entremont, Hans, & Muir, 1997). In laboratory set-ups, this behavior typically emerges from about 10 months of age when infants spontaneously shift their gaze to the same object another agent is looking at (Carpenter, Nagell, & Tomasello, 1998). While it was first argued that gaze following is independent from communicative contexts and just indicates a reflexive shift of attention to the same direction as a perceived head motion (Moore & Corkum, 1994), other studies have demonstrated that 6-month-old infants only follow an adult's gaze when it is preceded by communicative cues such as direct eye contact or infant-directed speech (Senju & Csibra, 2008; Csibra & Gergely, 2009).

Yet, while these results suggest that the presence of a communicative context can effectively trigger gaze-following, it does not imply that infants will *only* follow the gaze of another *human* agent. For example, in a study conducted by Deligianni and colleagues (2011), 8-month-old infants were presented with automated objects (lacking any human features) that either did or did not react contingently to the infant's fixation. They found that infants only followed the turning of the automated object towards a target object if the automated object had previously responded to them in a contingent fashion (Deligianni, Senju, Gergely, & Csibra, 2011). This shows that gaze following can in fact be triggered in the absence of a social agent as long as ostensive (communicative) cues are salient enough.

After having shown how much humans are prone (even from an early age on) to following others' gaze, the question arises what the function of gaze following is and exactly how it benefits social interactions.

Benefits of Gaze Following

Humans' sensitivity to gaze cues is astonishing. Not only does gaze effectively draw our attention to specific locations (Driver et al., 1999), we also recognize other people's faces faster and memorize them better when those faces are looking straight to us (i.e. depict direct gaze) compared to when faces depict averted gaze (Hood, Macrae, Cole-Davis, & Dias, 2003; Vuilleumier, George, Lister, Armony, & Driver, 2005; Senju & Hasegawa, 2005). Another striking example for the benefits of automatic gaze cueing has been demonstrated by a study conducted by Brennan and colleagues (2008).

In their study pairs of participants performed an O-in-Qs search task alone, or in different collaborative conditions (namely, shared-gaze vs. shared-voice vs. shared-gaze plus voice). Their results revealed that collaborative pairs performed better than solitary searchers (Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008). Interestingly, they also found that the most effective search strategy was actually the one in which participants coordinated their searching labor using shared-gaze alone (i.e. without talking to each other at the same time). More specifically, the results by Brennan and colleagues (2008) showed that in the shared-gaze condition, participants spontaneously adopted what turned out to be the most efficient strategy (i.e. "look where I am not looking"). In turn, this indicates that participants spontaneously learned to utilize the gaze of their task partner in the most effective way, "(…) without explicit coaching or training, typically within practice trials", (Brennan et al., 2008: p. 1474).

Thus, this finding suggests that our susceptibility to the way others look at the world seems to be especially suitable for fostering collaborative actions in an efficient manner. But how exactly do people establish perceptual common ground for joint actions or, put differently, how does following each other's gaze successively lead to performing actions together in a well-coordinated manner?

1.3 Joint Attention

By now there are many studies indicating that the phenomenon of *joint attention* serves as an important building block in establishing perceptual common ground between different actors. Joint attention has been conceptualized as the ability to share a common point of reference; i.e. to mutually attend to a certain object with an interactional partner (Tomasello, 1999). More specifically, in joint attention subject A coordinates her attention to a certain object *and* to subject B, while subject B coordinates her attention to the same object *as well as* to subject A (Tomasello, 1995). Crucially, the coordination that takes place during joint attention can only be accomplished through the comprehension that the other agent's focus of attention is directed to the same entity as the self. This implies an understanding of the other person as an agent who *intentionally* gazes at a certain object in the environment that is the same as one's own (Tomasello, 1995). The phenomenon of joint attention has consequently been argued to be one of the manifestations of humans' ability to share psychological states with each other (Tomasello & Carpenter, 2007). Thus, people that are engaging in joint attention are sharing an intentional relation to the world (Hobson, 1989).

The understanding that other people are intentional agents allocating their attention to selective parts in their environment is pivotal in order to learn about new objects – and for acquiring language. For instance, when infants begin to view others as intentional agents they begin to comprehend that the other person might selectively attend to a certain object in the environment (and ignore others), or that the other person might intend for them to selectively attend to a certain object in the environment – and ignore others (Tomasello, 1995). In turn, this enables infants to determine which part of the environment the communicated content most likely refers to, which provides the basis for acquiring knowledge about external referents and linguistic conventions (Tomasello & Todd, 1983; Tomasello, 1986). Tomasello and Carpenter (2007) have argued that joint attention therefore creates a shared space of (psychological) common ground enabling everything from collaborative activities with shared goals to the unique ways of human-style cooperative communication (Tomasello & Carpenter, 2007).

1.4 Task Co-Representation

One instance in which people seem to be particularly sensitive to differences in attentional relations between themselves and another person is when they are performing tasks together (Boeckler, Knoblich, & Sebanz, 2012). More specifically, it has been shown that knowledge about the focus of somebody else's attention modulates subsequent task performance (Boeckler, Knoblich, & Sebanz, 2012). Furthermore, it seems as if in social contexts people also take into account others' *intentional* relation to the environment (Sebanz, Knoblich, & Prinz, 2003). That is, besides being sensitive to *where* a task partner is looking, people also seem to take into account *what actions* the other person is supposed to be doing. In the following I will briefly discuss each of these phenomena.

Representing Another Person's Focus of Attention

When two people perform a task next to each other they seem to take into account certain aspects of each other's task even if this is not required for their own task performance (Boeckler et al., 2012). More precisely, a study by Boeckler and colleagues (2012) has shown that if two actors are attending to the same scene with differing attentional foci (e.g. one has a local and the other a global focus) their individual performance will be slowed down compared to when their foci of attention is the same (Boeckler et al., 2012).

In their study pairs of participants sat next to each other and performed a go-nogo Navon task together. Each participant was asked to respond to different letters and instructed either to focus on the local features (i.e. the small letters), or the global features (i.e. the large letters) of the stimulus. The Navon letters on the screen could then be either congruent (meaning that they were linked to the same response) or incongruent (meaning that they were linked to different responses). Boeckler and colleagues (2012) hypothesized that if the task partner's focus of attention was represented, then participants should be slower during the congruent compared to the incongruent condition. And indeed, the results revealed an increase of RTs when the two foci of attention differed compared to when they were the same (Boeckler, Knoblich, & Sebanz, 2012). These findings indicate that participants experience a conflict on the level of selecting and applying the appropriate focus when two foci differ from each other in a joint task set-up. Furthermore, the results suggest that in joint task set-ups people do not only represent their own task, but also take into account (aspects) of their partner's task.

Representing Another's Task during Joint Actions

In addition to being sensitive to where others' gaze and attention is directed to, people also seem to be receptive to others' intentional relations to the environment. Specifically, when performing actions next to another person, people seem to form representations of the other person's task, even if this is not required to succeed in their own task (Sebanz, Knoblich, & Prinz, 2003, Sebanz, Knoblich, & Prinz, 2005).

For example, Sebanz and colleagues (2003) conducted a study in which they compared three conditions of the Simon task (Simon, 1990; Simon et al., 1970). During the individual condition, participants were asked to respond to one stimulus attribute (say, a red ring on a left vs. right pointing finger) and not to respond to another stimulus attribute (say, a green ring on the pointing finger). In the binary choice condition, participants were instructed to respond to both stimulus attributes (i.e. red and green rings) with two different responses. Finally, in the joint condition the same task was distributed among two participants (such that each participant would respond to one stimulus attribute, respectively). Importantly, the direction in which the finger pointed was the irrelevant stimulus feature in all three conditions, whereas the color of the ring of the finger was the relevant feature.

In line with previous results, Sebanz and colleagues (2003) found a Simon effect in the binary choice condition and not in the individual go-nogo condition. However, they also found a Simon effect in the joint condition indicating that participants represented both their own and their partner's task during the joint condition (Sebanz et al., 2003). More specifically, the results suggest that participants co-represented their co-actor's task share in the joint condition, such that they knew when and how the other person needed to respond to a certain stimulus. As the representation of the co-actor's task then involved the same spatial coding of responses as for the participants, this could bring about the 'joint Simon effect' (Sebanz & Knoblich, 2009).

According to the authors, the function of such a task co-representation mechanism is to effectively predict and simulate what another person is supposed to do given a certain task set-up (Sebanz et al., 2006; Sebanz & Knoblich, 2009). In turn, this can facilitate the prediction of each other's actions, which is fundamental for performing joint actions together.

Taken together, the results of this section demonstrate what probably constitutes an important foundation for our unique capabilities to coordinate our actions with somebody else (cf. Clark, 1996); if I am aware of where (and what) you are focusing on that gives me a 'head-start' in successively engaging in future interactions with you, as it facilitates my understanding of how you look at the world.

In the next section on visual perspective-taking, I will further discuss how such an understanding of somebody else's point of view can be experimentally captured both in humans buts also in non-human animals. This evolutionary perspective will demonstrate that we are not the only species eliciting the cognitive mechanisms enabling social interactions and that in fact, there is a range of other species with whom we share some of the building blocks to understand and relate to others.

1.5 Visual Perspective-Taking

Taxonomy and Terminology

Ever since the early 1980s, a distinction has been made between so called *Level–1 and Level–2 perspective-taking abilities* that originated in the developmental literature (Flavell, Everett, Croft, & Flavell 1981). The goal of this differentiation was to capture childrens' differential understanding about *whether* another person can see a given object or not (referred to as Level-1 perspective-taking), versus *how* another person sees an object (referred to as Level-2 perspective-taking, cf. Flavell et al., 1981).

To experimentally test Level-1 perspective-taking abilities, Flavell and colleagues (1981) presented children a card with a picture of a cat on one side and a dog on the other side. This card was then held vertically between the child and the experimenter and the child was asked which animal the experimenter saw. From around three years of age children successfully reported that the experimenter saw, say the dog, while they themselves saw the cat (Flavell et al., 1981).

In contrast, in the original Level-2 perspective–taking task, subjects were presented a picture of a turtle lying on a table in front of them and were then asked whether it looks upside-down or the right-way-up to someone sitting at the opposite side of the table (Flavell et al., 1981). Here, prior to around four years of age, children typically judged incorrectly that the person opposite of them will see the turtle the same way they do (Flavell et al., 1981). These developments are generally seen as part of a pattern in developing sophistication in children's capacity to understand and properly represent the mental states of others (e.g. Perner, 1991). Yet, the precise cognitive processes underlying the different levels of perspective-taking are still debated in the literature.

Spontaneousness and the Social Nature of Visual Perspective-Taking

According to the so called 'minimal theory of mind' account (Low, Apperly, Butterfill, & Racoczy, 2016; Butterfill & Apperly, 2013) it has been proposed that humans possess two complementary mechanisms in order to capture the visual perspectives of others. On the one side, there is a mechanism that is fast and efficient in tracking *whether* another person can or cannot see a certain object (*visual perspective-taking*, also referred to as Level-1 perspective-taking). On the other side, there is a mechanism that is comparably slow as it operates in a more resource-demanding way, although it is therefore also capable to compute *how* somebody else sees a target object in her environment (*visuospatial perspective-taking*, or Level-2 perspective-taking). As a consequence, one should find evidence for spontaneous Level-1 perspective-taking but not for spontaneous Level-2 perspective-taking (Surtees, Samson, & Apperly, 2016; Butterfill & Apperly, 2013; Surtees, Butterfill, & Apperly, 2012).

This distinction between two perspective-taking mechanisms, one being fast, effortless and spontaneous, the other one being slow, effortful and not spontaneously triggered, turned out to be congruent with some findings. For example, in their influential 'dot perspective-taking task', Samson and colleagues (2010) showed evidence for Level-1 perspective-taking by demonstrating that adults would spontaneously compute whether or not an avatar would see the same amount of objects in a scene as they did.

Specifically, when being asked to make judgments about what they themselves could see, participants reliably showed altercentric intrusion, that is they spontaneously took into account what the avatar could or could not see – even if this was completely irrelevant for their task (see Samson et al., 2010). For instance, when participants saw three disks hanging on a wall while an avatar could only see one disk, participants' response times showed interference, i.e. they were slower and made more errors compared to a condition in which both saw the same amount of disks (Samson et al., 2010).

Furthermore, as the same interference effects were shown under a cognitive load manipulation of the dot perspective-taking task (Qureshi, Apperly, & Samson, 2010) and regardless of perspective-taking actually being irrelevant and costly for one's own task performance (Samson et al., 2010), this corroborated the hypothesis that Level-1 perspective-taking might indeed be grounded in a fast and efficient mechanism that is spontaneously triggered. In addition, Surtees, Butterfill and Apperly (2012) found no evidence for spontaneous perspective-taking using a Level-2 task in which they tested whether participants where sensitive to *how* an object would look like from another's perspective (see also Surtees, Samson, & Apperly, 2016). Taken together, these findings therefore indicated that one crucial difference between Level-1 and Level-2 perspective-taking might indeed be traced along the fault line of spontaneousness and that consequently, there should not be spontaneous Level-2 perspective-taking.

However, recent research seems to challenge this distinction of Level-1 perspective-taking being fast, effortless and spontaneous vs. Level-2 perspective-taking being slow, effortful, and not spontaneously triggered. For example, Elekes and colleagues (2016) found evidence for spontaneous Level-2 perspective-taking when two actors shared the same focus of attention. In their task, participants needed to make numerical judgments while sitting opposite of a task partner. Their results showed that participants were slower to indicate whether a number was smaller or larger than five when the numerical value of a digit was different for their task partner (e.g. on trials where they saw a '6' while their partner saw a '9'). This indicates that participants also computed the symbol from the other's viewpoint, that is, they spontaneously took into account *how* the stimulus looked from the point of view of their task partner (Elekes, Varga, & Király, 2016).

Importantly, they found no evidence for adults spontaneously adopting their task partner's perspective if their task partner's focus of attention differed from their own (Elekes et al., 2016). They hypothesized that for Level-1 perspective-taking to be triggered, knowledge about the other's visual access to a mutually looked at object should be sufficient. However, in order for Level-2 perspective-taking to be triggered, it might be necessary to share the same focus of attention (e.g. on the same feature of the stimulus, cf. Elekes et al., 2016).

Another study using a similar numerical judgment paradigm corroborated that humans are indeed sensitive to spontaneously computing *how an object looks to another person,* even if it conflicts with their own perspective (Surtees, Apperly, & Samson, 2016). Importantly, in this study by Surtees and colleagues (2016) evidence for Level-2 perspective-taking only occurred if the task partner had previously shown an active engagement in the task. However, in contrast to the study by Elekes et al., (2016), participants also adopted the other's perspective in situations where the task partner was not attending to the same stimulus aspect (Surtees et al., 2016). That is, when participants had to judge the magnitude of a number stimulus, they spontaneously took into account how a task partner saw the numerical (e.g. as a '6' while participants saw it as a '9'), regardless of the fact that the task partner's goal was to judge the surface pattern of the number (plain vs. spotted, Surtees et al., 2016).

Taken together, these results suggest that there seem to be certain circumstances (namely, sharing the same focus of attention and / or being mutually engaged in the same task) under which human adults actually do engage in Level-2 perspective-taking also spontaneously and hence, that the distinction between Level-1 and Level-2 perspective-taking might not run along the fault line of spontaneousness after all.

Another important issue in the debate on perspective-taking revolves around the question whether one can actually claim that perspective-taking effects are essentially social phenomena (that is, whether they are only triggered by intentional agents) or whether they can be entirely explained by lower-level mechanisms that are in fact agent-independent, such as attention re-orienting (Heyes, 2014). For example, it has been argued that the findings of Samson and colleagues' (2010) dot perspective task could be interpreted in the context of a domain-general mechanism picking up directional cues such as another person's gaze or the direction of her forehead and thereby orienting an agent's attention to what the other person is focusing on (Heyes, 2014). Crucially, such a domain-general mechanism could operate independently of other people's (higher-level) mental states such as their thoughts, beliefs or percepts (i.e. what they see). In fact, the same mechanism should be at work even if the other is not an intentional agent but a symbol which effectively captures one's attention (Heyes, 2014).

More specifically, re-interpreting the results by Samson and colleagues (2010), Heyes (2014) argued that the interference effect (which the authors claimed to be indicative of perspective-taking) was not due to the participants processing a mismatch between what they themselves saw and what the avatar could see. Instead, the specific body orientation of the avatar acted as a spatial cue highlighting the disks on the wall, which in turn led to more efficient processing on the side to where the avatar was oriented to (Heyes, 2014). In consequence, the experimental effect could have been elicited without involving any ascription of a mental state (such as 'the avatar *sees* 2 discs') at all.

In an attempt to verify this alternative explanation, Santiesteban and colleagues (2014) used a similar task set-up as in the original study by Samson et. al (2010), but replaced the avatar with an arrow of the same size as the human. And indeed, their findings showed the same pattern of results as in the original study (Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014). This led them to claim that the effects that had previously been argued to demonstrate fast and efficient reasoning about another person's *mental state*, are actually due to low-level attentional mechanisms that do not differentiate between intentional agents and non-agentive directional cues (Santiesteban et al., 2014; Heyes, 2014).

Interim Summary

The previous debate highlights, on the one hand, that future studies investigating perspective-taking should ideally be able to unequivocally address the possibility of potential lower-level explanations. On the other hand, it also indicates that there is an on-going controversy about how one can conceive the adoption of another agent's visual perspective, and what kind of cognitive architecture might enable such perspective-taking.

One way to approach these questions is to ask whether the phenomenon of perspective-taking is exclusively found among humans, or whether we actually share this capacity (at least partially) with other non-human animals. Crucially, the goal of comparing perspective-taking abilities between human and nonhuman animals should not be to search for and emphasize humans' uniqueness in the animal kingdom. Instead, such a contrast might reveal that there are both similarities and differences in the way human and non-human animals cope with other agents' visual perspectives and that consequently, some cognitive mechanisms enabling perspective-taking might be the same while others might differ. Put differently, this contrast might help to better define the cognitive architecture underlying perspective-taking. Finally, this approach will also make it easier to single out and focus on those aspects of perspective-taking (such as visuospatial perspective-taking), which have been understudied until this point.

Perspective-Taking in Non-Human Animals

The debate about *exactly what* cognitive abilities humans must possess in order to successfully engage in perspective-taking and *exactly how* we develop visual perspective–taking abilities (cf. Apperly & Butterfill, 2009) has also attracted comparative psychologists' attention. While the ability to adopt others' visual perspectives is a crucial component for humans to successfully navigate the social world, it is not only humans who live in groups and exhibit social behavior such as competition, cooperation and reciprocity. In other words, other group-living species probably benefit from the ability to take into account others' visual perspectives in much the same way as humans do. And indeed, although there is no evidence so far that any non–human animal succeeds on tasks that require complex mentalizing skills (e.g. for chimpanzees see Call & Tomasello, 1998), more and more data has now accumulated showing that an astonishing range of non–human animals seem to be able to reliably cope with what Flavell and colleagues (1981) coined as Level–1 visual perspective taking scenarios. In the following we will discuss how chimpanzees, ravens, and goats exhibit a formidable understanding of their conspecifics' visual perspectives.

Until the early 2000s, there was an interesting puzzle in the study of primate social cognition. On the one hand, evidence accumulated that many primate species (especially chimpanzees, or *Pan troglodytes*) possessed the abilities to follow the gaze direction of their conspecifics (Povinelli & Eddy, 1996a). What's more, rather than exhibiting a simple learning or search mechanism (in the sense of 'turn in the direction of where the other is looking and search randomly until you find something interesting'), it seemed as if they were actually able to geometrically track the gaze direction of others to specific targets, including locations behind barriers, behind themselves and past distractors (Tomasello, Call, & Hare, 1998). Finally, if the chimpanzees would not find anything at the target locations, they were even found to 'check back', that is, they looked back to the other individual's face and tracked its gaze direction another time (Call & Tomasello, 1998; 2005). These sophisticated gaze-following abilities are adaptive as they enable individuals to obtain useful information about group-mates, predators and the location of food. Indeed, knowing what conspecifics can or cannot see may be especially advantageous in social foraging scenarios (cf. Bräuer, Call, & Tomasello, 2007).

However, on the other hand, when being faced with tasks that actually required them not only to track the gaze of others, but to base a foraging decision on the visual behavior or experience of another individual, chimpanzees seemed to be completely insensitive to these kinds of cues (Povinelli & Eddy, 1996b). If, for example, chimpanzees had to decide on one out of two opaque containers in order to retrieve a piece of food, and the only cue that was accessible to them was the gaze of another (in this case a human experimenter), they were utterly unable to use this cue in order to successfully retrieve the food item (Call & Tomasello, 1998). Thus, even though there was solid evidence showing that chimpanzees reliably followed the gaze direction of others, it seemed as if they could not transfer or use this knowledge in simulated foraging situations.

One explanation for the negative results was that the experimental set–ups posed quite unnatural situations for chimpanzees. In fact, as it is primarily the competition for limited food that is characterizing the social life of chimpanzees, it would be highly unlikely that a conspecific would indicate the location of food in natural environments– as this implied that the conspecific would not be able to keep it for him- or herself anymore (Sterck, Watts, & Shaik, 1997). And indeed, in a seminal study Hare and colleagues (2000) advanced the understanding of visual perspective-taking in primates by devising a new task that came closer to the primate's natural foraging behavior.

In their experiment, two individuals, one dominant to the other, were placed in rooms on opposite sides of a test room, in which food was placed at strategic locations. For example, two pieces of attractive food were positioned between the subjects so that one individual had visual access to both pieces, whereas the other could only see one (as the other piece was hidden behind an occluder). Both individuals were then simultaneously released into the test room in order to retrieve the food.

Their initial results showed two 'hierarchy-depending' strategies; while subordinates preferentially retrieved the food that was hidden from the dominant, the dominant individuals preferentially retrieved the food that was visible to the subordinates (before they then continued to pick up the hidden piece as well). Interestingly, as the same individuals were tested in differing hierarchical dyads (i.e. sometimes they were subordinate to the other while on other trials they were the dominant individual), Hare and colleagues (2000) found that individuals would not stick with one of the strategies throughout all experimental trials but actually switched according to their specific hierarchical roles. That is, rather than following some blind behavioral rules or contingencies, the same individuals successively exhibited both hierarchy-depending strategies, emphasizing the sophistication of their visual perspective-taking abilities (see Hare, Call, Agnetta, & Tomasello, 2000).

However, an alternative explanation could argue that the chimpanzees were merely reacting to behavioral cues, given that they were released at the same time. Specifically, as both individuals approached the food, subordinates might have monitored the movements of the dominant individual and observed that the dominant would head to the visible food item. In turn, this could have led subordinates to adjust accordingly and go for the other food item – not because it was out of the dominant's sight, but because this followed from the dominant's behavior. Thus, in a follow–up experiment Hare et al. (2000) replicated their paradigm, the only difference being that one individual would always get a slight head–start over the other (Hare et al., 2000). As a result, the subjects had to decide which of the two pieces they would go to *before they saw their competitor approaching*.

As the results of these ‘delay trials’ then turned out to be virtually identical to the first experiment (that is, subjects did not merely change their behavior according to the arrival of the dominant individual) Hare and colleagues (2000) concluded that chimpanzees indeed take into account what conspecifics do and do not see and that, on top of that, they can also transfer this knowledge in order to perform successful foraging strategies (Hare et al., 2000; see also Hirata & Matsuzawa, 2001; and see Hare, Call, Tomasello, & Visalberghi, 2003 for capuchin monkeys failing to show such a sophisticated understanding of their conspecifics’ visual perspective).

Specifically, another study has demonstrated that chimpanzees use their perspective-taking skills to deceive a human competitor. Hare et al. (2006) placed pieces of food on two separate trays on either side of a booth. If the human saw the chimpanzee approaching either tray, he retracted it, thereby preventing the chimpanzee from retrieving the food. In the first of the two experiments, the human looked exclusively on one food tray while having his back to the other. In the subsequent experiment, one piece of food was placed behind an opaque barrier, while the other was placed behind a clear barrier. Interestingly, in both of these experiments chimpanzees preferentially approached the piece of food to which the human did not have visual access to (Hare, Call, & Tomasello, 2006; for a similar approach with rhesus macaques see also Flombaum & Santos, 2005). Taken together, these findings suggest that some nonhuman primates not only compute what others can or cannot see, they also use this knowledge in order to maximize their food intake in competitive situations.

However, this leaves the question open whether these cognitive skills are confined to some selected primates or whether similar abilities can also be found in other group-living species that exhibit social behavior such as competition, cooperation and reciprocity.

In similarity with primate species, also ravens (*Corvus corax)* have been shown to follow gaze directions around obstacles (Bugnyar, Stöwe, & Heinrich, 2004). In addition, they reliably remember and raid food caches they have seen others make, which actually makes them a suitable candidate to further investigate visual perspective-taking during social foraging. In the wild and in captivity, ravens compete both against individuals that store food but also against individuals that could potentially raid their own caches (see Heinrich & Pepper, 1998; Bugnyar & Kotrschal, 2002). In addition, as scavengers ravens show a natural tendency to observe the behavior of others (be it conspecific or other species) that are in possession of food.

This led Bugynar (2011) to test whether ravens that observed an experimenter caching food, were sensitive to other raven bystanders' visual access, and whether they efficiently used this knowledge to predict the bystanders' subsequent behavior. Bugnyar's experimental protocol was the following. Controlling the number and location of food caches being made, Bugnyar (2011) let a human experimenter store two food items in the presence of a 'focal raven' (who had uninterrupted visual access to the caching area) and raven bystanders whose view of the caching area was individually manipulated by opaque curtains. After the caches were made, focal ravens were confronted with a conspecific that had been visually present either at both, none or just one of the caching events. In consequence, the conspecific possessed either full, no or partial information about the cache locations.

Bugnyar predicted that if ravens were aware of their conspecifics' visual perspectives, they should exhibit differential behavior in accordance with the conspecific's individual knowledge. More specifically, while fully informed conspecifics would pose a high risk for pilfering either caches, partially informed conspecifics would pose a risk only to the specific cache they had seen being made. Accordingly, focal birds should hurry to retrieve both caches when confronted with a fully informed conspecific (compared to non-informed conspecifics) and they should deliberately select between the caches when confronted with partially informed conspecifics (Bugnyar, 2011).

The results seem to confirm that ravens successfully differentiated between these different knowledge states of their conspecifics. As fully informed conspecifics potentially posed a high risk of raiding either cache, focal ravens were indeed quicker to collect (or 'pilfer') both human made caches compared to situations in which they were confronted with non–informed subjects, in which case they did not show a preference for choosing a particular cache first (Bugnyar, 2011). In contrast, when confronted with partially informed conspecifics, focal birds did not show a significant difference in 'pilfering latencies' (i.e. the time between entering the room and pilfering a cache), but they selectively chose to pilfer those caches first, which the observing individuals had information about (compared to the other cache sight, about which the observing individual possessed no information about, Bugnyar, 2011).

These findings lead Bugnyar (2011) to the assumption that the ravens indeed used different behavioral tactics ('hurry up or choose') that were in accordance with the information their conspecifics gathered during the caching process (cf. Bugnyar, 2011). However, based on these results alone, one cannot distinguish whether ravens really understood what their conspecific could or could not see or whether they merely based their behavior on a learned rule (e.g. 'compete with those that could be seen at the time of caching' vs. 'compete with those that have actually seen the caching', see Bugynar, 2011).

Thus, in a follow–up experiment, Bugnyar (2011) tried to replicate his findings while on half of the trials, he now blocked only the competitors' view towards the caches while the view between the focal subject and the competitors remained unhindered. If focal subjects were merely basing their behavior on the presence of the competitor, so his reasoning, the additional occluder should not change the way they successively pilfered the caches. Yet, the results revealed that focal subjects significantly differed in their selectivity of pilfering caches; only when their conspecifics had unhindered visual access to the caches did the focal subjects show a significant preference for choosing a particular cache, whereas this preference effectively disappeared when their conspecifics' view on the cache sight was blocked (Bugnyar, 2011).

Taken together, these results strongly suggest that ravens have a rather sophisticated understanding of visual barriers, that they are capable of geometrically judging the other's view relative to their own and, finally, that they effectively incorporate this knowledge when retrieving (see Bugnyar, 2011; Bugnyar & Heinrich, 2005) or caching food (see Bugnyar & Kotrschal, 2002; Bugnyar & Heinrich, 2005; for similar results with scrub jays, see Emery & Clayton, 2001; Dally, Emery, & Clayton, 2006). Finally, these results mirror the findings in great apes and thus provide further support for a convergent cognitive evolution between corvids and primates (see Emery & Clayton, 2004).

So far, it seems as if both chimpanzees (Hare et al., 2000) and ravens (Bugynar, 2011) can successfully take the visual perspective of others into account when competing with a conspecific for food. However, such cognitive skills may not be exclusively confined to primates or corvids. Other mammalian species also possess complex social structures in which competition is an important part of everyday interactions.

Kaminsky et al. (2005) examined whether goats (*Capra hircus)* take into account what conspecifics can or cannot see. In similarity with primates and corvids, goats seem to be sensitive to the gaze direction of other individuals, as they reliably follow the gaze of conspecifics to objects that are behind and above their current visual field (Kaminsky, Riedel, Call, & Tomasello, 2005). Furthermore, like non–human primates and corvids, goats live in fission–fusion societies and form coalitions and alliances, and are known to reconcile after fights (see Kaminsky et al., 2005). Put together, these features make them another suitable candidate for investigating visual perspective-taking during food competition.

Kaminsky's experimental protocol was similar to the conspecific competition paradigm (see Hare et al., 2000). In three different test populations (i.e. in three different zoos in Germany) a dominant and a subordinate goat always competed over food across situations in which different kinds of visual barriers were present. Thus, in some situations there was a mismatch of (food–related) visual information between dominant and subordinate.

Interestingly, what Kaminsky et al. (2005) then found was that the preferences subordinate goats showed in such competitive food situations actually depended on whether or not they had received aggression from the dominant goats during the experiment. More specifically, subordinate subjects that *had* received aggression preferred to approach the hidden piece of food, whereas subjects that *had not* received aggression preferred to approach the visible piece of food first (Kaminsky et al., 2005). Similar to the dominant chimpanzees in the original study by Hare et al. (2000) this latter strategy (of approaching the visible piece of food first) endowed the subordinate goats with significantly more food, as they later on continued to retrieve the other piece (i.e. the one that was initially hidden from the dominant's view) as well. In contrast, the other group of goats that retrieved the hidden piece first, successively ‘lost’ the visible piece as their competitors ate it once they were released into the room (Kaminsky et al., 2005).

Remarkably, these diverging behavioral patterns were perfectly reflected in the differences across the three test groups. Aggression was only found in the group that was exposed to different (that is, more restrictive) feeding procedures as well as differences in their overall living space (that is, less space for each individual) compared to the other test subjects.[3] More importantly, the authors argue that although their test subjects used two distinct strategies, they effectively revealed a sensitivity for the visual perspective of their conspecifics (Kaminsky et al., 2005). On top of that, this sensitivity subsequently led them to maximize the overall food that was available to them (sometimes even monopolizing the total amount). Taken together, Kaminsky and colleagues (2005) take these results as clear evidence that also goats are able to compute - and base their successive behavior on the grounds of – what their conspecifics can or cannot see (Kaminsky et al., 2005).

Contrasting Perspective-Taking in Human vs. Non-Human Animals

If chimpanzees, ravens and goats are able to consider their conspecifics’ visual perspective, one might ask: what is the nature of this understanding in those non-human animals and (how) does it differ from perspective-taking abilities in humans?

According to Hare and colleagues (2000) the cognitively strongest hypothesis would be to argue that these non–human animals understand the visual perception and experience of other individuals in much the same way as humans. That is, they are able to understand that: 1) others can or cannot see the same objects as themselves depending on the specific environment (e.g. because of an occluder); 2) others can have different perspectives on the same

[3] According to Kaminsky et al. (2005), these strategy differences across the subordinate groups stress the general importance of contextual factors (such as space availability) one must take into account when measuring observable behavior (for similar observations in chimpanzees see Bräuer et al., 2007).

object they are perceiving depending on their spatial positioning (e.g. because of their specific spatial orientation); and 3) the visual experience of somebody else is prima facie similar to their own, allowing them to mentally simulate and adopt other perspectives and thereby experience how they themselves would perceive it (cf. Hare et al., 2000).[4]

Judging from the evidence above, the first point seems to be unequivocal; chimpanzees, ravens and goats seem to be sensitive to whether or not a conspecific can or cannot see a certain target object. Furthermore, their behavior demonstrates that they can effectively use this knowledge in order to optimize their foraging behavior. This indicates that chimpanzees, ravens and goats can take into account somebody else' visual perspective and hence, that humans share this cognitive ability with other non-human animals.

The second and the third point (that is, arguing that non-human animals also take into account how conspecifics see a certain target object) seem more difficult to defend. First of all, based on the studies that were discussed here, one cannot make an argument about non-human animals showing an understanding of how their conspecifics' saw the target object, as the experimental set–ups exclusively tested whether the target object would fall into another's line of sight. In turn, the results of these studies are illustrative only with regards to the question whether a conspecific can or cannot see a target object. However, other studies have actually failed to show a human-like understanding of others' visual perspectives in chimpanzees (cf. Call & Tomasello, 1998; Tomasello & Call, 1997).

On the other end of the spectrum, the cognitively weakest explanation would be that the behavior non–human animals exhibited in these experiments were exclusively based on learned behavioral contingencies without any understanding of the other's visual experience

[4] Notice that the first point resembles what Flavell and colleagues (1981) have coined as Level-1 perspective-taking, while the second and the third point correspond to what Flavell denoted as Level-2 perspective taking abilities (Flavell et al., 1981).

at all (Hare et al., 2000). Yet, also this objection can be largely rejected as most of the experimental evidence above actually falsified these types of explanations (cf. the adaptive behavior of chimpanzees that successively had both the dominant and the subordinate role but also the delay trials in which a modulation of behavior was measured if one of the two agents received a head-start).

Thus, in light of the data that was discussed in the previous section, both of these extreme hypotheses seem unlikely. Somewhat in between these two extremes, Hare et al. (2000) argue for a 'mixed hypothesis'. This mixed hypothesis suggests that the ability of gaze following constitutes one of the fundamental building blocks in perspective-taking. As chimpanzees, ravens and goats naturally follow the gaze of their conspecifics in order to predict their behavior, they have - through individual learning - probably made a particular important association. This association concerns the relationship between the visual access of others, its likely target, and how that relates to their behavior in a variety of contexts. Through individual learning experiences, those non-human animals could have then gained a flexible understanding that the behavior of others is determined in some specific ways by what their conspecifics do or do not have visual access to. In turn, they successfully incorporate this kind of knowledge in situations where they have to make quick decisions in order to maximize their food intake under competitive circumstances (Hare et al., 2000). A situation, which should be fairly frequent in the lives of these group–living animals.

Summary and Outlook

After having investigated how non-human animals might conceive a conspecific's visual perspective, the question remains how this conception differs from humans' understanding of another person's visual perspective. Earlier in this discussion it was mentioned that humans do not only represent *whether* another person can or cannot see a certain object but also *how* objects look like from another's point of view (also referred to as *visuospatial* perspective-taking).

While one could easily claim now that this ability must be exactly what constitutes the signature limit between human and non-human animals' perspective-taking skills, one would blatantly disregard that, in fact, many aspects about visuospatial perspective-taking in humans are still unknown. There is an on-going controversy about whether, when and how instances of visuospatial perspective-taking occur. For example, it is still debated whether humans adopt another person's visuospatial perspective by all means spontaneously, that is without being prompted to do so. Moreover, it is still a matter of discussion whether the evidence for perspective-taking so far is exclusively 'social in its nature' (in the sense that they are only triggered in the presence of an intentional agent) or whether it can be accounted for by lower-level, agent-independent mechanisms.

Thus, although there is accumulating evidence about the phenomenon of *visual perspective-taking* (Samson et al., 2010; Ramsey et al., 2013, Baker et al., 2015), comparably little is known about the phenomenon of *visuospatial perspective-taking* in humans. The present work attempts to fill this gap by investigating whether - and if so how and under which circumstances - humans spontaneously adopt another's visuospatial perspective (Chapter 2). Furthermore, I will explore whether knowledge about another's visual access systematically modulates VSP-taking (Chapter 3). Finally, this thesis probes whether VSP-taking is specific for interactions in shared physical spaces or whether alternatively, it also occurs in mental space (Chapter 4). The findings of this work will thereby contribute to the current debate about how we can conceive the cognitive architecture (differentially) enabling social agents to adopt another's visual perspective.

Chapter 2 – Evidence for Spontaneous VSP-Taking in a Shared Physical Space

Summary

The study presented in **Chapter 2** investigates the underlying factors as well as boundary conditions that characterize the spontaneous adoption of another person's visuospatial perspective (VSP) during social interactions. To this end, a novel paradigm is presented, in which a participant and a confederate performed a simple stimulus-response compatibility (SRC) task sitting at a 90° angle to each other. In this set-up, participants would show a spatial compatibility effect only if they adopted the confederate's VSP. In a series of five experiments we found that participants reliably adopted the VSP of the confederate, as long as he was perceived as an intentionally acting agent. Our results therefore show that humans *are* able to spontaneously adopt the differing VSP of another agent and that there is a tight link between perspective-taking and performing actions together. Furthermore, the results suggest that spontaneous VSP-taking can effectively facilitate and speed up spatial alignment processes accruing from dynamic interactions in multi-agent environments.[5]

[5] The full material presented in this chapter was published in JEP:HPP; 2016 Mar; 42(3): 401-412 (Freundlieb, Kovács, & Sebanz, 2016).

Introduction

Whether we steer a remote control helicopter, navigate fictional characters through a complex maze in a video game or simply guide terribly lost friends to their destination over the phone – our daily life constantly challenges us with a plethora of visual perspectives that are often different to (if not competing with) our own point of view. Moreover, in many situations we do not have the possibility to ponder over the divergence of our own and somebody else's vantage point but instead need to make quick decisions in order to successfully interact with others. Take, for instance, a soccer player who wants to pass a ball over to a moving teammate while dribbling past a swarm of opponents during a match.

Recent research suggests that we are equipped with sophisticated mechanisms that allow us to track and flexibly integrate varying perspectives on multiple levels. This is reflected, for example, in our fundamental comprehension that one and the same thing can be viewed or construed differently depending on the chosen standpoint – whether this requires an epistemic, conceptual, affective or visuospatial perspective (see Perner, Brandl, & Garnham, 2003). Functionally, the ability to flexibly adopt another person's perspective e.g. while jointly attending to an object, is pivotal in order to enable the formation of joint goals, as well as the successive coordination of actions (Bratman, 1992).

While managing different perspectives can be challenging due to egocentric biases (see Keysar, Barr, Balin, & Bauer, 2010; Keysar, Lin, & Barr, 2003; Durmontheil, Apperly, & Blakemore, 2010; Mattan, Quinn, Apperly, Sui, & Rothstein, *in press*), accumulating evidence indicates that people are remarkably sensitive towards other agents' perspectives (Samson, Apperly, Braithwaite, Andrews, & Scott, 2010; Zwickel, 2009; Tversky & Hard, 2009; Michelon & Zacks, 2006; Vogeley & Fink, 2003). When being asked to make judgments about what they themselves could see, participants reliably showed *altercentric intru-*

sion, that is they automatically took into account what somebody else could or could not see – even if this was completely irrelevant for their task (see Samson et al, 2010). For instance, when participants saw three discs while an avatar could only see one disc, their response times were slower compared to a condition where both saw the same amount of discs (Ibid.).

Although people´s sensitivity to others' perspectives is in itself noteworthy, an important point of debate in the literature is what exactly adopting another´s perspective entails, and how it affects one's own action planning. The finding by Samson and colleagues shows that we readily compute *whether* somebody else can see a target object or not. However, one could argue that perspective-taking entails more than that. For example, we sometimes need to compute the location of objects relative to others, and to infer what these objects look like from their perspective. In other words, we need to be able to take into account the differing *visuospatial perspective* (VSP) of another person.

It has been argued that judging whether another agent can or cannot see an object, and judging the location of the object with respect to that agent (that is, taking the VSP of that agent) involves two different perspective-taking processes (Michelon & Zacks, 2006). Whereas so-called *line-of-sight* or *visibility* judgments appear to be very rapid and effortless, judging the relative location from a particular perspective requires a transformation of one's egocentric reference frame and is therefore assumed to be more effortful, leading to more errors and increased response latencies (cf. Surtees, Apperly, & Samson, 2013; Kessler & Thomson, 2010). However, given that VSP-taking is argued to play an important role in social interactions (Creem-Regehr, Gagnon, Geuss, & Stefanucci, 2013; Bratman, 1992), it should not always manifest through interference and hence, be detrimental on performance. There should be situations in which one can also find evidence for the opposite, that is where VSP-taking actually has a positive effect on performance. For instance, when we need to guide someone to a specific location on the phone, we should be able to flexibly adopt and use his viewpoint in order to achieve our communicative goals.

Visuospatial Perspective-Taking in Communicative Tasks and Memory Research

Several studies have investigated how we retrieve spatial information in the presence of another person (Sjolund, Erdman & Kelly, 2014; Galati, Michael, Mello, Greenauer, & Avraamides, 2013; Shelton & McNamara, 2004; Schober, 1995). In most of these studies, pairs of participants worked together, with one participant (the "director") being instructed to describe a layout of objects to the other participant (the "matcher"), who then attempted to recreate the layout from a perspective that was shifted from the director's perspective. Crucially, as the matcher did not have visual access to the original layout (as both the director and the matcher were separated by a visual occluder) he could only rely on the instructions given by the director.

After reconstructing the layout, both the director and the matcher then individually completed a memory task, which revealed the specific reference frame they used in order to represent the object layout. While it is somewhat unsurprising that the matcher always represented the layout using an egocentric reference frame (given that this was the only perspective the matcher had experienced), there is evidence suggesting that the director also represented the matcher's perspective.

For example, Shelton and McNamara (2004) found that after having been explicitly instructed to describe the layout from the matcher's perspective, directors incorporated the matcher's perspective into their mental representations of the layout during the memory task. Furthermore, Galati et al. (2013) showed that explicit instructions about the partner's perspective were not necessary, and that the mere presence of a partner was sufficient for the partner's perspective to influence spatial memory. In contrast, Sjolund et al. (2014) found that regardless of the presence of a collaborative partner, directors exclusively remembered the spatial layout using an egocentric reference frame. Thus, the question remains under which conditions people are able to compute somebody else's VSP spontaneously, that is without being explicitly prompted to do so.

Spontaneous VSP-taking

In most prior studies participants were explicitly asked to make judgements about the relative location of an object with respect to the perspective of another agent (see Michelon & Sacks, 2006; Surtees et al., 2013; Kessler & Thomson, 2010). However, Tversky and Hard (2009) investigated how the presence of a person on a photograph exhibiting a different spatial orientation (namely, opposite of the participants, that is, in a 180° angle from their own position) affected people's verbal description of the spatial relations among an array of objects in the photographed scene (Tversky & Hard, 2009). Interestingly, the mere presence of a person in that scene indeed led a quarter of the participants to take that person's perspective and describe the locations of the objects from the other's rather than their own point of view (Ibid.). Furthermore, when the photograph showed the person reaching out for an object and the experimenter phrased the question about the spatial relations of the depicted objects in terms of action (e.g. *In relation to the bottle where does he place the book?*), only 20% of the respondents stuck to their own point of view while the majority of them effectively adopted the other person's perspective to describe the scene in terms of his 'right' and 'left' (Ibid.). On the one hand, these findings suggest that despite the complexity of having to cope with two contrasting spatial dimensions, participants nevertheless spontaneously integrated and applied the other person's VSP. On the other hand, this study was based on verbal reports which may have prompted explicit reasoning. Thus, it is still an open question in the literature whether people spontaneously adopt others' VSP outside of a communicative task. Importantly, the study by Tversky and Hard (2009) does point to the fact that perceiving another person performing an action might play a role for perspective-taking to occur.

Acting together increases Visuospatial Perspective-Taking

Certainly, in many situations we do not only share the same visual but, in addition, also the same task environment with other people. In order to successfully plan and coordinate our actions, we need to be able to flexibly integrate the perspectives of our co–actors (cf. Creem-Regehr, Gagnon, Geuss, & Stefanucci, 2013). This prompts the assumption that in joint task settings, people should be particularly sensitive towards the respective perspectives of others. If this was the case, then one should find particularly pronounced instances of perspective-taking in joint task settings.

Supporting this claim, a study by Frischen and colleagues (2009) showed that observing another person's actions effectively triggered the same selective attention processes (namely, inhibition of salient distractors) one finds when people perform an action on their own – the crucial difference being, that in the joint task scenario these processes occurred for an allocentric, rather than an egocentric frame of reference (Ibid.).

Although this study gives an illustrative example of how the observation of action can effectively trigger a change of reference frames (see also Mazzarella, Hamilton, Trojano, Mastromauro, & Conson, 2012; Furlanetto, Cavallo, Manera, Tversky, & Becchio, 2013), the exact relationship between VSP-taking and joint task performances remains unclear. For example, we do not know whether knowledge about the other person's task is sufficient to induce VSP-taking or whether it is necessary that one can directly observe (and receive feedback of) the other person's actions. Furthermore, how much does the other person need to be involved in the task in order to spontaneously adopt her perspective – is it really necessary that the other is performing a task or could it be that the mere presence of another person exhibiting a different spatial orientation is already sufficient to trigger spontaneous VSP-taking? Finally, if there really is a tight connection between VSP-taking and performing actions together, as the study by Frischen and colleagues suggests, then how much of the other person's task is actually represented while adopting her VSP and how does this impact on one's own action planning?

Taken together, there are still a number of open questions concerning the phenomenon of perspective-taking. First of all, it is still unclear whether people actually adopt others' VSP outside of a communicative setting and, whether VSP-taking can also occur spontaneously. Second, is VSP-taking always detrimental to one's task performance or can we find situations in which it is actually facilitative during joint task performances? Finally, there are already indications in the literature suggesting that spontaneous VSP-taking might be particularly pronounced in circumstances where another person performs actions (cf. Frischen, Loach, & Tipper, 2009; Tversky & Hard, 2009). However, this link between action and spontaneous VSP-taking has - to our knowledge - not yet been systematically tested.

The aim of this study was to address these issues in a systematic manner. We used a novel paradigm in a series of five experiments to investigate whether participants would spontaneously adopt a VSP that is not their own, and to test whether and how this would affect action planning. To this end, we needed a task where a) there are two different VSPs, one of which is irrelevant for the participants' task, and where b) adopting the other's VSP has a clear effect on action performance.

Throughout this study, we placed participants in a 90° angle to a co-actor and asked them to perform an orthogonal stimulus-response (SR) compatibility task (cf. Craft & Simon, 1970; Simon, 1990) on a horizontally mounted ('table-like') computer display (see Figure 1). Given the sitting position of the confederate and the participant, the stimuli could thus be perceived from two different VSPs– either along a vertical or along a horizontal axis. Measuring responses according to the spatial position of the stimuli thereby allowed us to test effects of VSP-taking on action performance.

More specifically, the participant's own perspective always coincided with the vertical axis, so that stimuli presented along this axis did not overlap with the participant's horizontally arranged responses in terms of their spatial alignment. In contrast, the confederate's perspective coincided with the horizontal axis, creating a spatial overlap between the arrangement of the stimuli and the participant's responses. If participants showed a spatial compatibility effect in this context, this would provide clear evidence that they are performing the task relying on the confederate's rather than their own VSP.

2.1 Experiment 1

The first experiment investigated whether participants spontaneously integrate the visuospatial perspective (VSP) of a confederate, while performing an orthogonal stimulus-response (SR) compatibility task. Participants sat in front of a horizontally arranged screen and in a 90° angle to a confederate and were instructed to respond with a right or a left button press to stimuli appearing at the top or the bottom of the screen, respectively (see Figure 1). From the confederate's orientation, however, the stimuli appeared on the left and on the right side of the screen. Hence, we predicted that if participants adopted the confederate´s VSP, then they should show a spatial compatibility effect.

Method

Participants

16 participants (mean age = 20. 7 years, 11 women, 13 right-handed) signed up for this study and received gift vouchers for their participation. All were naïve to the purpose of the study, reported normal or corrected to normal vision and signed informed consent prior to the experiment. All 16 participants met the inclusion criterion of having more than 90% successful trials within each experimental condition.

Stimuli and apparatus

The stimuli consisted of a rectangle (subtending 11.35° of visual angle horizontally and 6.53° vertically) in which there were three empty circles (each subtending 3.27° of visual angle) at equal distance to each other. During the trials, one out of these three circles (either the one at the top, or the one at the bottom, but never the circle in the middle) then appeared as a black disc in place of the empty circle. These two types of stimuli were shown on a horizontally arranged 27" iMac (Mid-2011). The monitor was mounted at a height of about 25 cm from the floor. Responses were given on two button boxes (ioLab Response box), which both the participant and the confederate placed on their lap. The button boxes were partially covered with a piece of carton so that out of the default array of 7 buttons, only the ones used to respond (i.e. the buttons farthest to the left and right) were visible.

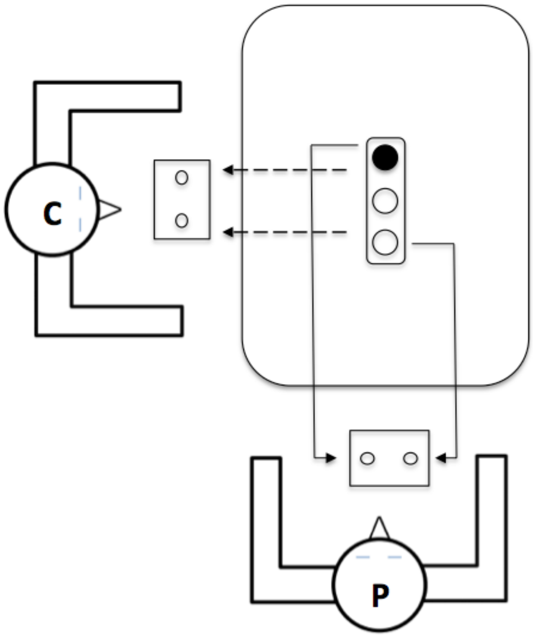

Figure 1. Experimental set-up for Experiment 1 (Ch.2).
Participant (P) sits in a 90° to the confederate (C), example of the compatible condition. Dashed lines indicate the confederate's mapping. Solid lines indicate the participant's mapping.

Design and Procedure

Both the participant as well as a confederate, who sat in a 90° angle to the participant, sat as close as possible to the screen (viewing distance was ≈ 35cm). Throughout the entire study, the same young adult male acted as the confederate. Each trial started with the presentation of a fixation cross (subtending 1.31° of visual angle, presented in the centre of the screen) for 350 ms. Subsequently, the screen turned blank for 100 ms after which, in a randomized manner, one of the two stimuli (top black disc vs. bottom black disc) was shown for 1200 ms. Participants performed two conditions (*compatible* and *incompatible*), each containing 100 trials and were asked to respond as fast and as accurately as possible.

To establish different compatibility relations, we varied the sitting position of the confederate and the SR mapping of the participants. In one half of the experiment, participants were instructed to respond to the appearance of the top black disc by pressing the right button on the button box with their right index finger and to respond to the bottom black disc by pressing the left button with their left index finger, respectively. In the other half, the mapping was reversed and they were thus instructed to respond to the appearance of the top black disc with a left and to the appearance of the bottom black disc with a right button press. In the *compatible* condition, the mapping of the participant concurred with the spatial orientation of the confederate, while in the *incompatible* condition it did not. For instance, if the confederate sat 90° to the left of the participant, participants were instructed with the 'up-left, down-right' mapping in the compatible, and with the 'up-right, down-left' mapping in the incompatible condition (see Figure 1).

Before each of the two conditions, ten practice trials familiarized the participants with the task. These were later excluded from the statistical analysis. Throughout both conditions the confederate, who sat in a 90° angle to the participant, was instructed to respond with a left button press if a black disc appeared on the left side of the screen and with a right button press if a black disc appeared on the right side of the screen. Both, the order of conditions and the position of the confederate (90° to the left vs. to the right of the participant) was counterbalanced across participants.

Data Analysis

We collected data only from participants. Errors (i.e. trials in which the wrong button or no button at all was pressed) and RTs more than two standard deviations from each participant's condition means were excluded from the analysis. Both the two condition means for correct response RTs for each participant as well as their errors were subjected to two separate two-tailed, paired-samples *t*-tests.

Results

2.7% of the trials were removed as errors and 4.8% were removed as outliers, leaving 92.5% of the raw data as correct response trials. Generally, the removal of these outliers did not result in changes of the significance patterns observed in this study. Comparing the number of errors in the compatible vs. incompatible conditions did not reveal a statistically significant result, $t(15) < 1$, $p = .94$. The RT analysis revealed that on average, participants were significantly faster in the compatible ($M = 356$, $SE = 9.8$) than in the incompatible ($M = 374$, $SE = 13$) condition; $t(15) = -3.28$, $p = .005$, (see Figure 2). In order to test whether the sitting position of the confederate (to the left vs. to the right of the participants) or the order of conditions (starting with the compatible vs. the incompatible condition) had an influence on the results, a repeated measures ANOVA was conducted with compatibility as a within subjects factor and both sitting position of the confederate and order as between subjects factors. The results yielded only a main effect of compatibility, $F(1, 12) = 9.23$, $p = .01$, $\eta_p^2 = .435$, but no effect of sitting position $F(1, 12) = 1.04$, $p = .33$, $\eta_p^2 = .08$, order $F(1, 12) < 1$, $p = .49$, $\eta_p^2 = .04$, or any interaction between them, all Fs < 1, ps $> .43$, $\eta_p^2 < .05$.

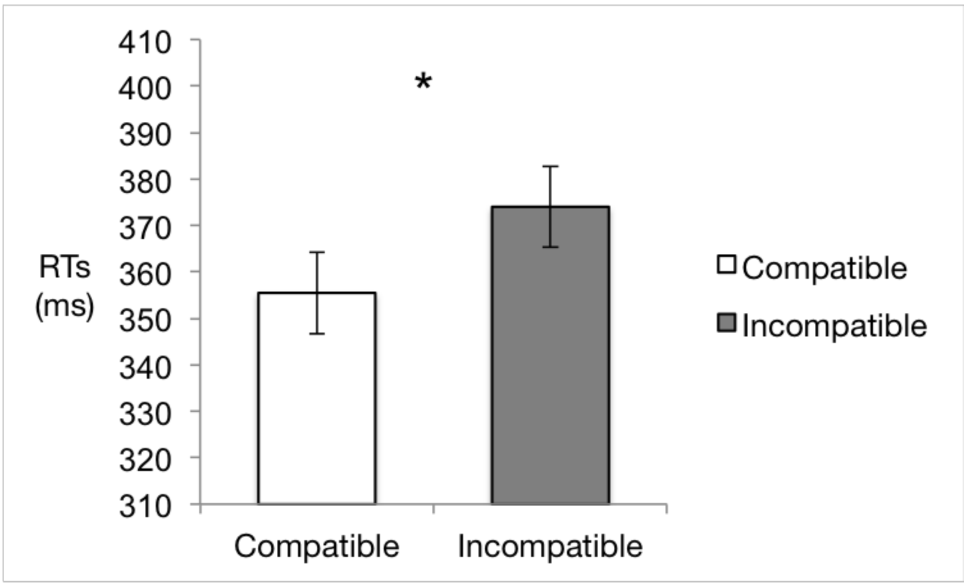

Figure 2. Mean RTs in the compatible and in the incompatible condition in Experiment 1 (Ch.2). Error bars display within-subject confidence intervals according to Loftus & Masson (1994).

Discussion

From the participants' point of view, there was no clear overlap between the stimulus dimension (which appeared on a vertical axis) and the response dimension (which was given on a horizontal axis). From the confederate's perspective though, both the stimulus and the response dimension overlapped. Hence, the assigned SR mappings were compatible or incompatible only with respect to the confederate's point of view.

The RTs of the participants showed a significant difference between the compatible and the incompatible condition. As this compatibility effect was independent of whether the confederate sat to the left or to the right of the participants, a SR-compatibility (e.g., a general and exclusive performance advantage for the 'up-right, down-left' mapping, see Cho & Proctor, 2003) cannot explain the pattern of these results. Instead, when the mapping of the participants concurred with the spatial orientation of the (left *or* right sitting) confederate, participants were significantly faster to respond compared to when the mapping did not concur with the confederate's orientation. Importantly, the task did not require the participants to compute the perspective of the confederate. All in all, these results suggest that participants spontaneously adopted the visuospatial perspective of the confederate.

However, one could argue that the overt responses given by the confederate (who performed the SR task in close proximity to the participants) might have made his specific orientation to the stimuli very salient for the participants (cf. Frischen et al., 2009). In other words, it is possible that the confederate's overtly given responses might have drawn participants' attention towards his particular spatial orientation and hence, evoked the compatibility effect. Experiment 2 addressed whether having visual and auditory access to the confederate's responses was necessary for spontaneous VSP taking to occur.

2.2 Experiment 2

Experiment 2 investigated the role of visual and auditory feedback exhibited by a co-acting confederate on spontaneous VSP taking. Previous studies have already shown that directly observing another person's actions leads people to adopt an allocentric frame of reference (Frischen et al., 2009). With regard to Experiment 1, it could therefore be argued that the overt responses of the confederate actually led the participants to pay more attention to his spatial orientation. Hence, being able to directly receive feedback from the confederate's actions might be a necessary precondition of spontaneously adopting another's VSP.

In contrast, another line of research indicates that explicit knowledge of another person's task is sufficient to form joint task representations, even if the actions are then covertly executed (see Sebanz, Knoblich, & Prinz, 2003; Tsai, Kuo, Hung, & Tzeng, 2008). If knowledge about the confederate's task in combination with knowledge about the location of his responses was sufficient to trigger a spontaneous adoption to his VSP in Experiment 1, then the previously found compatibility effect should persist regardless of whether or not feedback on the confederate's responses was available.

Method

Participants

19 new participants (mean age = 23.94 years, 9 women, all right-handed) signed up for this study and received gift vouchers for their participation. Three participants did not meet the inclusion criterion of having more than 90% successful trials within each experimental condition, leaving 16 participants (mean age = 22.25 years, 8 women) for the analysis. All participants were naïve to the purpose of the study, reported normal or corrected to normal vision and signed informed consent prior to the experiment.

Stimuli and apparatus

The stimuli were identical to Experiment 1. The crucial difference between Experiment 1 and 2 was that in half of the experimental trials (the no-feedback condition) both the participants and the confederate wore ear-plugs ('OHROPAX') as well as earmuffs ('Earline MAX200 31020') so that their responses were inaudible. Furthermore, their response boxes were placed inside of cardboard boxes so that their hands were not visible.

Procedure

Participants performed two conditions (*feedback* and *no-feedback)*. Each condition consisted of two blocks (*compatible* and *incompatible*). Each block contained 100 trials and participants were asked to respond as fast and as accurately as possible. The feedback condition was an exact replication of Experiment 1. Before the no-feedback condition, the experimenter placed the response boxes in the cardboard boxes and instructed both the participant and the confederate to put in ear-plugs and to put on the ear-muffs. Throughout both conditions, the participants and the confederate were instructed to perform the same tasks as in Experiment 1. As before, participants read through the instructions for their own and the confederate's task, which ensured that they knew about the correct response location in the no-feedback condition. Before each experimental block, ten practice trials familiarized the participants with the task. These were later excluded from the statistical analysis.

The order of feedback conditions (feedback, vs. no-feedback), the position of the confederate (90° to the left vs. to the right of the participant), as well as the order of compatibility conditions (compatible vs. incompatible) was counterbalanced across participants.

Data Analysis

Errors (i.e. trials in which the wrong button or no button at all was pressed) and RTs more than two standard deviations from each participant's condition means were excluded from the analysis. Both the two condition means for correct response RTs and errors for each participant were subjected to separate two-way, repeated measures ANOVAs with the factor Feedback (feedback vs. no-feedback) and Compatibility (compatible vs. incompatible).

Results

2.3% of the trials were removed as errors and 3.9% were removed as outliers, leaving 93.8% of the raw data as correct response trials.[6] The error analysis did not reveal any statistically significant results for Feedback $F(1, 15) < 1$, $p = .87$, $\eta_p^2 < .01$, Compatibility, $F(1, 15) < 1$, $p = .87$, $\eta_p^2 < .01$, or the interaction between the two, $F(1, 15) = 1.31$, $p = .27$, $\eta_p^2 < .08$.

The RT analysis revealed a significant main effect of Compatibility, $F(1, 15) = 14.32$, $p = .002$, $\eta_p^2 = .488$) with RTs being faster during compatible than during incompatible trials (see Figure 3). There was neither a main effect of Feedback, $F(1, 15) < 1$, $p = .52$, $\eta_p^2 = .02$, nor an interaction between Compatibility and Feedback, $F(1, 15) < 1$, $p = .99$, $\eta_p^2 < .01$.

[6] The removal of outliers did not result in any changes of the significance patterns observed in this experiment.

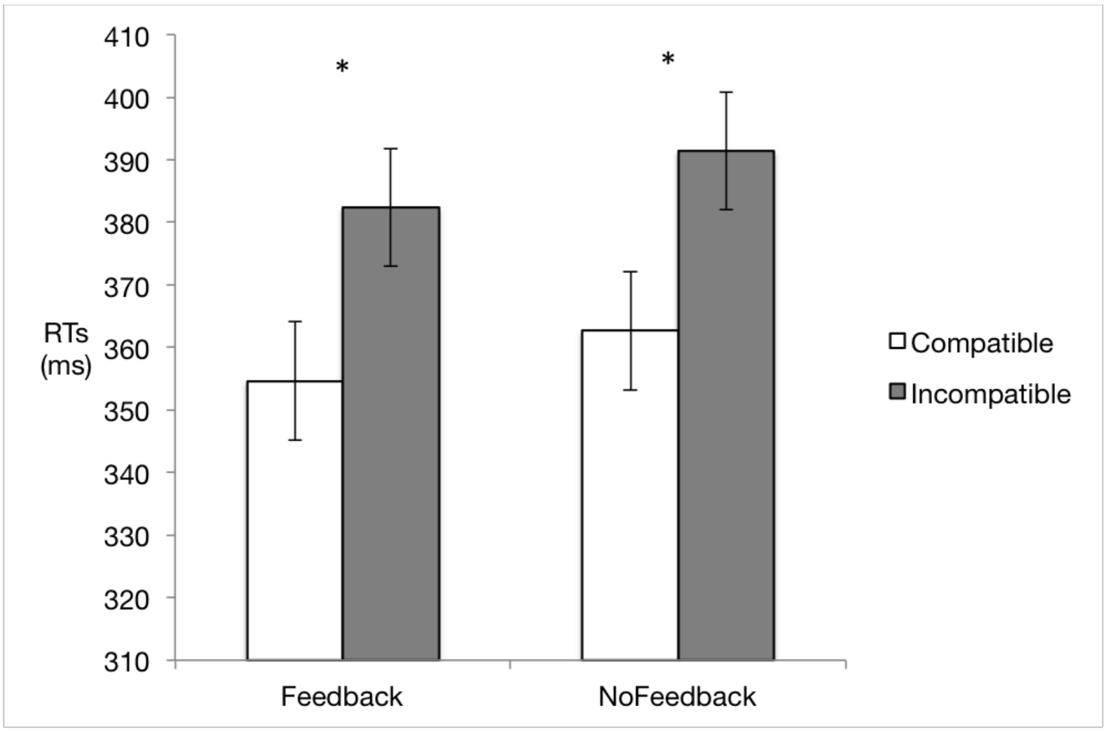

Figure 3. Mean RTs in the feedback and in the no-feedback condition in Experiment 2 (Ch.2). Error bars display within-subject confidence intervals according to Loftus & Masson (1994).

Discussion

The main effect of Compatibility indicates that participants were always faster to respond when their assigned mapping concurred with the spatial orientation of the confederate, regardless of whether they received visual or auditory feedback on the confederate's responses. Thus, this suggests that knowledge about the confederate's task together with knowledge about the location of his responses was sufficient to trigger a spontaneous adoption of the confederate's VSP. At the same time, this experiment provided evidence for VSP-taking in a context where (un-)intentional coordination of actions or entrainment (cf. Richardson, Marsh, Isenhower, Goodman & Schmidt, 2007; Schmidt & Richardson, 2008) could not have occurred because participants and confederates could not perceive each other´s actions. The results of this experiment therefore indicate that entrainment is not a necessary factor for the observed VSP-taking effect to occur.

However, other studies suggested that the mere (passive) presence of another person already suffices to change the way in which stimuli are perceived with respect to that person's frame of reference (see Tversky & Hard, 2009; Costantini, Committeri, & Sinigaglia, 2011). Looking for the minimal conditions under which spontaneous VSP taking is exhibited, this leads to the question, whether VSP taking in the present task is relying on another person's actions at all. Maybe the diverging point of view of the confederate could have already been sufficient for participants to switch to the confederate's frame of reference. This question was addressed in Experiment 3.

2.3 Experiment 3

Experiment 3 investigated, whether the confederate's differing orientation to the stimuli was already sufficient for participants to adopt his VSP. Previous studies suggest that the presence of a *passive agent* suffices to change the way in which humans perceive spatial relations among objects (Tversky & Hard, 2009) as well as their surrounding action space (Costantini et al., 2011). On the contrary, other research has shown that knowledge about the *intentional actions* of another agent is crucial in order to simulate his actions and thereby establish interpersonal links (Atmaca, Sebanz, Prinz, & Knoblich, 2008; Sebanz et al., 2003; Zwickel, 2009).

So far, the previous experiments cannot disentangle whether the task performance of the confederate was actually necessary in order for spontaneous VSP taking to occur or whether the mere presence of another agent exhibiting a different frame of reference might have already been sufficient to trigger the same effect. We hypothesized that if the diverging perspective alone sufficed to trigger spontaneous VSP taking, then it should not matter whether the confederate actually performed the SR compatibility task. If, on the other hand, the task performance of the confederate was necessary in order to evoke spontaneous VSP, then the effect should be restricted to those conditions, in which the confederate actually performs the compatibility task alongside the participants.

Method

Participants

25 new participants (mean age = 22.6 years, 13 women, 23 right-handed) signed up for this study and received gift vouchers for their participation. One participant did not meet the inclusion criterion of having more than 90% successful trials within each experimental condition, leaving 24 participants (mean age = 22.54 years, 12 women, 22 right-handed) for the analysis. All participants were naïve to the purpose of the study, reported normal or corrected to normal vision and signed informed consent prior to the experiment.

Stimuli and apparatus

These were identical to Experiment 1.

Procedure

Participants performed two conditions (*other-active* and *other-passive)* with two blocks (*compatible* and *incompatible*), respectively. Each block contained 100 trials and participants were asked to respond as fast and as accurately as possible. While the other-active condition was an exact replication of Experiment 1, in the other-passive condition the confederate was instructed not to respond but just to observe the stimuli on the screen. Before each condition, ten practice trials familiarized the participants with the task. These were later excluded from the statistical analysis.

The order of conditions (other-active, vs. other-passive), the position of the confederate (90° to the left vs. to the right of the participant), as well as the order of compatibility conditions (compatible vs. incompatible) was counterbalanced across participants.

Data Analysis

Errors (i.e. trials in which the wrong button or no button at all was pressed) and RTs more than two standard deviations from each participant's condition means were excluded from the analysis. The two condition means for correct response RTs and errors for each participant were subjected to separate two-way, repeated measures ANOVAs with the factors Role of Confederate (active vs. passive) and Compatibility (compatible vs. incompatible).

Results

2.5% of the trials were removed as errors and 4% were removed as outliers, leaving 93.5% of the raw data as correct response trials.[7] The error analysis revealed a significant main effect of Role of Confederate, $F(1, 23) = 15.9$, $p < .01$, $\eta_p^2 = .4$, showing that participants made more errors when the confederate was active (M = 3.5% errors) than when the confederate was passive (M = 1,6% errors). Neither the main effect of Compatibility, $F(1, 23) < 1$, $p = .83$, $\eta_p^2 < .01$, nor the interaction between Role of Other and Compatibility, $F(1, 23) < 1$, $p < .77$, $\eta_p^2 < .01$, was significant.

The RT analysis revealed a significant interaction between Role of Confederate and Compatibility, $F(1, 23) = 5.1$, $p = .03$, $\eta_p^2 = .18$. In post-hoc analyses, pairwise comparisons showed a significant difference in RTs between the compatible (M = 338, SD = 32) and incompatible (M = 355, SD = 38) blocks only in the 'active', $t(23) = - 4.62$, $p < .01$, two-tailed, but not in the 'passive' condition, $t(23) < 1$, $p = .85$ (see Figure 4). Furthermore, those trials in which the confederate was active but the participants' mapping was incompatible did not statistically differ from both compatible ($t(23) < 1$, $p = .4$) and incompatible trials ($t(23) < 1$, $p = .35$) in which the confederate was passive. Put differently, the data shows that - compared to all the other three conditions - participants had a particular speed advantage for the com-

[7] The removal of outliers did not result in any changes of the significance patterns observed in this experiment.

patible trials, in which the confederate was active (compatible active vs. compatible passive: $t(23) = -2.47$, $p = .021$; compatible active vs. incompatible passive: $t(23) = -3.05$, $p = .006$). In addition, there was a marginally significant main effect of Role of Confederate, $F(1, 23) = 3.88$, $p = .06$, $\eta_p^2 = .14$, with faster RTs when the confederate was active, and a tendency for Compatibility $F(1, 23) = 3.21$, $p = .09$, $\eta_p^2 = .12$.

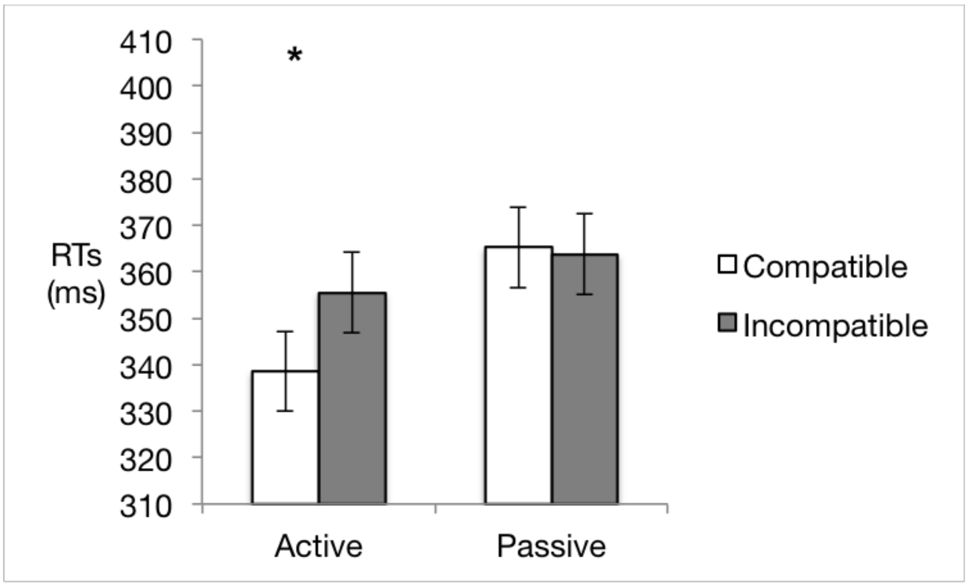

Figure 4. Mean RTs in the other-active and in the other-passive condition in Experiment 3 (Ch.2). Error bars display within-subject confidence intervals according to Loftus & Masson (1994).

Discussion

Most importantly, Experiment 3 elicited a significant interaction effect between the factors Role of Other and Compatibility. Only during the other-active condition were participants significantly faster to respond to compatible vs. incompatible trials. This pattern of results suggest that spontaneous VSP taking crucially relies on the confederate being perceived as an intentional co-actor (cf. Zwickel 2009; Sebanz et al., 2003) and that the mere presence of a passive confederate is not sufficient in order to induce a spontaneous adoption of his point of view.

Interestingly, the *post-hoc* comparisons indicated that, compared to the other conditions, participants were significantly faster only during compatible trials and thus, that VSP-taking might have actually facilitated their task performance. We will return to this point in the general discussion.

The marginally significant main effect of Role of Confederate (p = .06) in RTs, together with the significant main effect of Role of Confederate (p < .01) in terms of errors suggest that there was a speed-accuracy trade-off during the blocks where the confederate was active. However, the fact that participants were faster and made more errors in the other-active condition does not explain the compatibility effect in this condition.

The results suggest that knowledge about the confederate's task (Experiment 2) together with the actual task performance of the confederate (Experiment 3) is necessary to trigger spontaneous VSP taking. This indicates that the underlying mechanism leading to spontaneous VSP taking might not only hinge on the other's visuospatial perspective but - to some degree - also on a representation of the other person's task.

In order to find out exactly what role the confederate's task played and how much of his task was actually represented by the participants, we conducted Experiment 4. One confound that needed to be ruled out was that in all the experiments thus far, the responding hands of the participant and the confederate always changed together with the assigned mappings. That is, in compatible conditions, both of them always used the same responding hand and in incompatible conditions they always used different hands on any given trial. According to the task co-representation account from the literature on joint action, people form representations of each other's stimulus-response mappings when acting next to each other (see Sebanz, Knoblich, & Prinz, 2005; Sebanz & Knoblich, 2009; Butterfill & Sebanz, 2011). In line with this account, one could therefore argue that the participants in the above experiments might have co-represented the confederate's exact SR-mappings, which in turn could have caused the compatibility effect. In other words, the knowledge about when the confederate needed to push which button with which hand could have sped participants up during compatible trials, in which they had to execute the *same actions* (e.g. both of them needed to push the 'left' button with their left hand), and slowed them down during incompatible trials, in which they had to execute *opposite actions* (e.g. one needed to push the left, while the other needed to push the right button). Importantly, such a process could be independent of the actual VSP of the confederate. This alternative explanation was examined in Experiment 4.

2.4 Experiment 4

In the previous experiment it was shown that the confederate needed to perform a task alongside the participants in order for spontaneous VSP to occur. Experiment 4 investigated exactly what role the confederate's task played and how much of his task was actually represented by the participants.

The rationale behind this experiment was as follows: If participants actually represented the confederate's task in terms of SR rules (Sebanz et al., 2005), seeing a stimulus that requires a particular response from the confederate would activate a representation of this response. Regardless of the visuospatial orientation of the confederate, they should thus be faster when their responding hands coincided with the confederate's, and slower when their responding hands differed.

In contrast, one could also contend that the presence of an active confederate was already sufficient to modulate participants' frame of reference with respect to the stimuli (cf. Mazzarrella et al., 2013). Hence, if exact knowledge about the other's task in terms of SR rules was not necessary and instead, the presence of a co-acting confederate already sufficed to trigger spontaneous VSP taking, then one would expect the previously found effect to be independent of whether the same or different hands were used to respond.

Method

Participants

19 new participants (mean age = 24.32 years, 11 women, 18 right-handed) signed up for this study and received gift vouchers for their participation. Three participants did not meet the inclusion criterion of having more than 90% successful trials within each experimental condition, leaving 16 participants (mean age = 23.1 years, 10 women, 15 right-handed) for the analysis. All participants were naïve to the purpose of the study, reported normal or corrected to normal vision and signed informed consent prior to the experiment.

Stimuli and apparatus

These were identical to Experiment 1.

Procedure

Participants performed two conditions (*same hands* and *different hands)* with two blocks (*compatible* and *incompatible*), respectively. Each block contained 100 trials and participants were asked to respond as fast and as accurately as possible. Their task was identical to Experiment 1. However, while in the previous experiments the task of the confederate never changed (that is, he was always assigned to a congruent mapping throughout the entire experiment) it now switched; in half the trials, the confederate performed a congruent mapping (i.e. when a stimulus appeared on his right-hand side, he needed to respond with a right button press and vice versa), in the other half, he was now assigned to an incongruent mapping (i.e. when a stimulus appeared on his right-hand side, he needed to respond with a left button press and vice versa). The instructions for both the confederate and the participant resulted in half the trials being performed with their same hands (hence, the *same hands* condition) and the other half being performed with different hands (i.e. the *different hands condition*, see Figure 5). Before each condition, ten practice trials familiarized the participants with the task. These were later excluded from the statistical analysis.

The order of conditions (same hands vs. different hands), the position of the confederate (90° to the left vs. to the right of the participant), as well as the order of compatibility conditions (compatible vs. incompatible) was counterbalanced across participants.

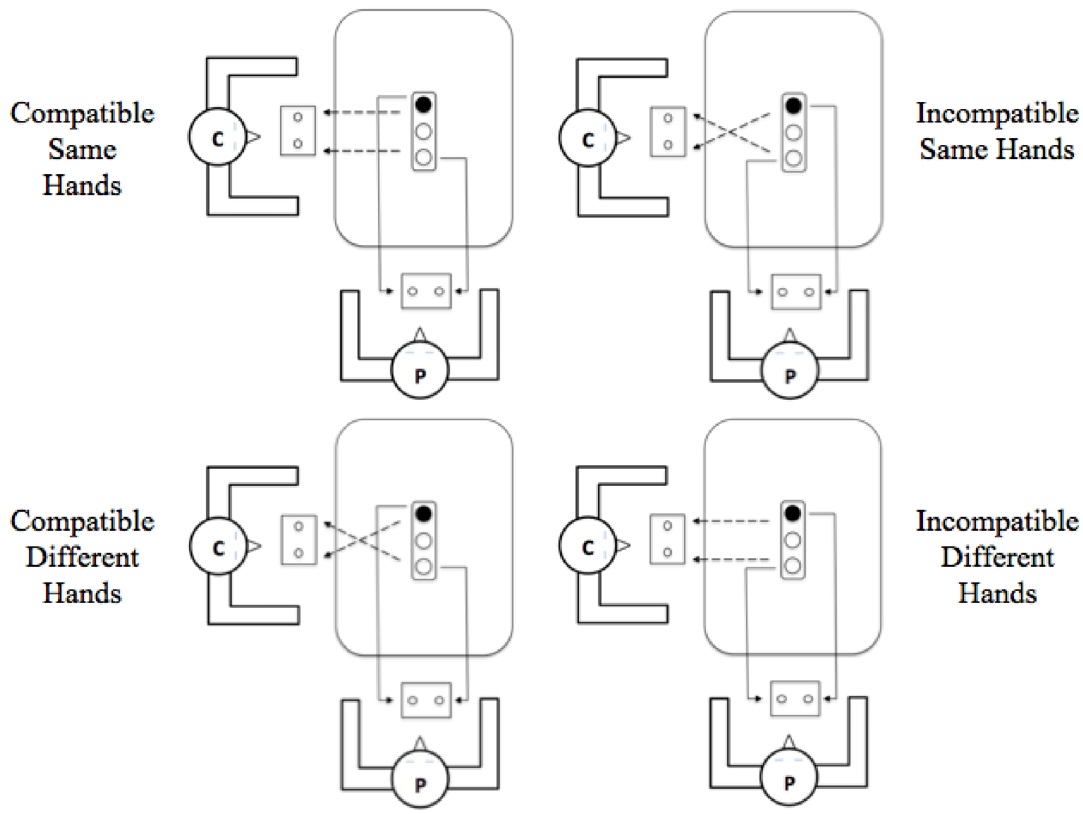

Figure 5. Overview of experimental set-up in Experiment 4 (Ch. 2).

Data Analysis

Errors (i.e. trials in which the wrong button or no button at all was pressed) and RTs more than two standard deviations from each participant's condition means were excluded from the analysis. Both the two condition means for correct response RTs and errors for each participant were subjected to separate two-way, repeated measures ANOVAs with the factors Hands (same vs. different) and Compatibility (compatible vs. incompatible).

Results

2.5% of the trials were removed as errors and 3.4% were removed as outliers, leaving 94.1% of the raw data as correct response trials.[8] The error analysis did not reveal any statistically significant results for Hands, $F(1, 15) < 1$, $p = .99$, $\eta_p^2 < .01$, Compatibility $F(1, 15) < 1$, $p = .56$, $\eta_p^2 = .02$, or the interaction between the two $F(1, 15) < 1$, $p = .89$, $\eta_p^2 < .01$.

The RT analysis revealed a significant main effect of Compatibility, $F(1, 15) = 30.22$, $p < .001$, $\eta_p^2 = .67$, with RTs being generally faster during compatible than during incompatible trials (see Figure 6). Furthermore, there was a significant interaction between Compatibility and Hands, $F(1, 15) = 5.65$, $p = .03$, $\eta_p^2 = .27$. The difference score between incompatible and compatible trials was significantly higher in the same ($M = 32.97$, $SD = 21.94$), compared to the different hands condition ($M = 17.54$, $SD = 23.05$), $t(15) = 2.37$, $p = .03$. Post-hoc two-tailed *t*-tests then revealed a significant compatibility effect between compatible and incompatible trials both in the same hands, $t(15) = - 6.01$, $p < .01$, as well as in the different hands condition, $t(15) = - 3.04$, $p < .01$. In addition, participants were slower on incompatible trials, in which they used the same hand as the confederate ($M = 368$, $SD = 41$), compared to incompatible trials, in which the two used different hands ($M = 357$, $SD = 38$). However, this comparison was only marginally significant; $t(15) = 2.04$, $p = .06$. Finally, there was no main effect of Hands $F(1, 15) < 1$, $p = .34$, $\eta_p^2 = .06$, suggesting that participants were not overall faster or slower when responding with the same or different hand as the confederate.

[8] The removal of outliers did not result in any changes of the significance patterns observed in this experiment.

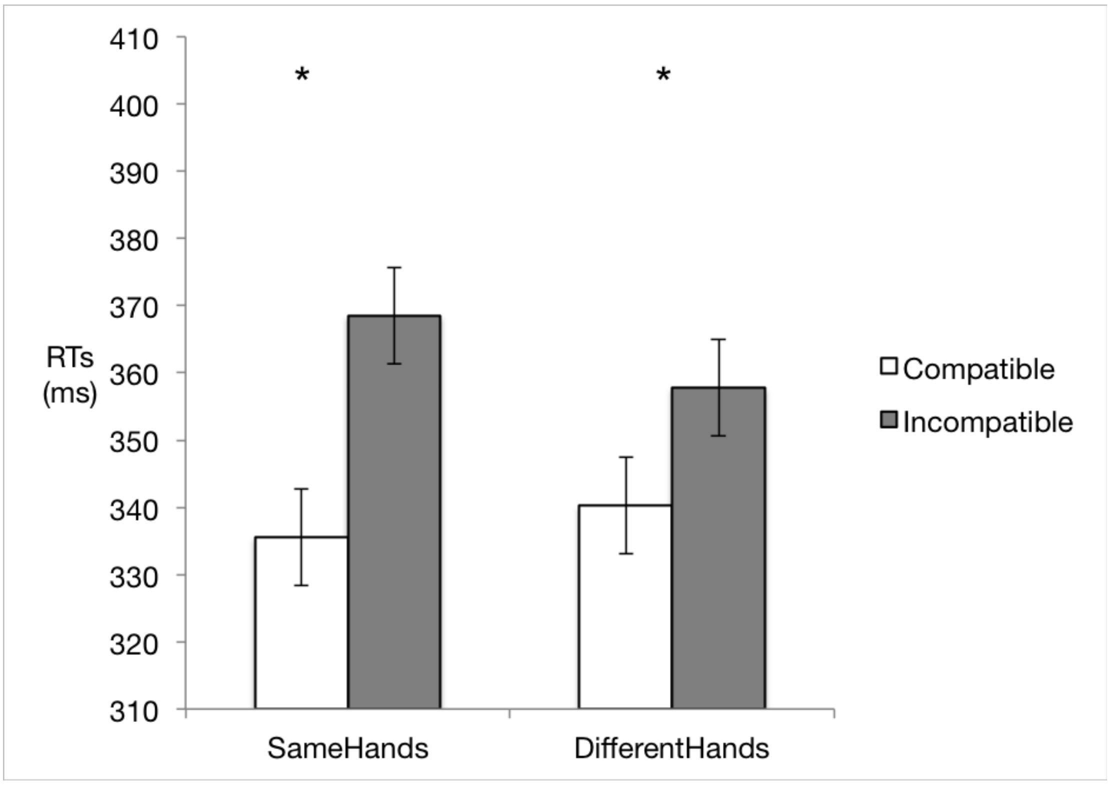

Figure 6. Mean RTs in the same-hands and in the different-hands condition in Experiment 4 (Ch.2). Error bars display within-subject confidence intervals according to Loftus & Masson (1994).

Discussion

Experiment 4 investigated whether responding with the same or with a different hand as the confederate affected the participants' compatibility effect. The results showed that the compatibility effect did not depend on but was influenced by the relation between the responding hands. The significant main effect of Compatibility together with the absence of a significant main effect of Hands suggests that regardless of whether participants had to respond with the same or different hands as the confederate, they were always significantly faster to respond during compatible compared to incompatible trials.

Indicated by the interaction effect, we also found evidence suggesting that participants' responses were not completely independent from the responses given by the confederate. More specifically, it seems as if using the same hands during incompatible trials particularly hampered participants' responses. One could hypothesize that during this condition, the mismatch between the VSP of the confederate and the required SR mapping was particularly salient because the confederate simultaneously also responded in an incongruent manner. However, as the group comparison was only marginally significant, this remains a tentative conjecture.

Taken together with the results from the previous experiment our findings suggest that participants' spontaneous adoption of the confederate's VSP did rely on the confederate performing a task next to them (Experiment 3) while it did not rely on an exhaustive representation of the confederate's task in terms of SR mappings (Experiment 4). This raises the question whether *any task* performed by the confederate can trigger the mechanism underlying spontaneous VSP taking or whether it has to be a *spatially matching one*, that is, a task in which the stimulus and the response dimension overlap (cf. Kornblum, Hasbroucq, & Osman, 1990).

2.5 Experiment 5

In the previous experiments it was shown that participants reliably adopted a confederate's VSP, as long as the confederate was performing a SR task (cf. Experiment 3) but regardless of whether this SR task was congruent (cf. Experiment 1-3) or incongruent for the confederate (cf. Experiment 4) in terms of the hands used for responding.

Experiment 5 investigated whether a spatially neutral S-R arrangement (that is, neither congruent nor incongruent, e.g., vertically presented stimuli in combination with laterally arranged responses, cf. Kornblum et al., 1990) performed by the confederate is already sufficient to trigger spontaneous VSP taking. This allowed us to find out more about the underlying mechanism of the spontaneous adoption of another's VSP.

Theoretically, there can be two competing explanations for the compatibility effect found in the previous experiments. On the one hand, it could have been the case that participants adopted the confederate's point of view (seeing the stimuli as 'left and right' rather than 'up and down') as well as the spatial dimension of his responses - which also entailed a 'left' and a 'right' dimension. As a result this would allow for an overlap between SR dimensions and hence, lead to the compatibility effect. On the other hand, one could claim that participants adopted the point of view of the confederate but disregarded his response dimension and instead 'superimposed' their own response dimension. As the participants' own responses were also given laterally (i.e. as left and right), this could have also lead to a conflict between the stimulus and the response dimension and hence, to a compatibility effect.

To disentangle these two alternatives, we rotated the confederate's response dimension so that also from his point of view, it no longer overlapped with the stimulus dimension (see Figure 6). The rationale behind this manipulation was the following: if participants adopted both the point of view *and* the spatial response dimension of the confederate, then the compatibility effect should disappear once the confederate responded orthogonally (that is with an 'up' and a 'down' button press) to the stimuli. Alternatively, if participants took the point of view of the confederate but retained their own response dimension (that is, perceive the responses as 'left' and 'right'), then the compatibility effect should persist regardless of the confederate's response dimension.

Method

Participants

19 new participants (mean age = 21 years, 8 women, all right-handed) signed up for this study and received gift vouchers for their participation. Three participants did not meet the inclusion criterion of having more than 90% successful trials within each experimental

condition, leaving 16 participants (mean age = 21.88 years, 8 women, all right-handed) for the analysis. All participants were naïve to the purpose of the study, reported normal or corrected to normal vision and signed informed consent prior to the experiment.

Stimuli, apparatus and design

The stimuli were identical to Experiment 1. The only difference in the apparatus was that the confederate's button box was rotated 90° on his lap so that instead of being located left and right, the two response buttons were now oriented up and down with respect to the confederate. Hence, there was no longer an overlap between the spatial dimension of the stimuli (appearing to the left and right of the confederate) and the spatial dimension of the confederate's responses (now requiring an 'up' and 'down' response). Thus, for both, the participants and the confederate, the stimulus and response dimensions were now orthogonal to each other (see Figure 7).

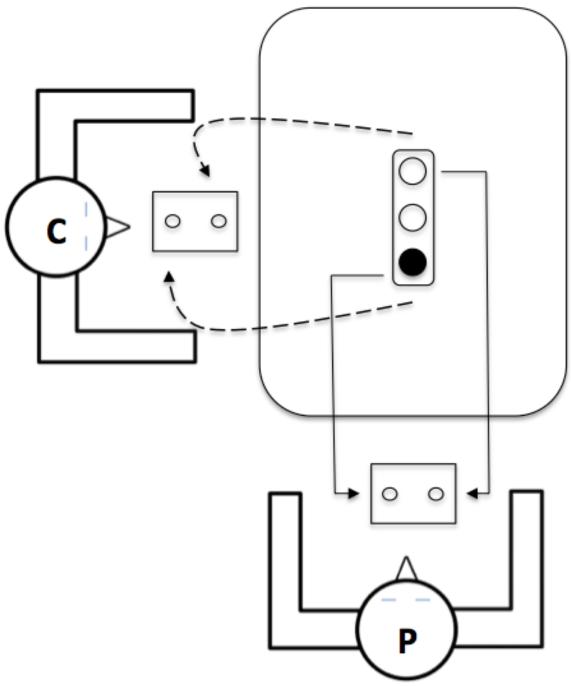

Figure 7. Experimental set-up in Experiment 5 (Ch.2).
Participant (P) sits in a 90° to the confederate (C), example of the incompatible and different hands condition.

Procedure

Participants performed two conditions (*same hands* and *different hands)* with two blocks (*compatible* and *incompatible*), respectively. Each block contained 100 trials. The participants' task was identical to Experiment 1 and they were asked to respond as fast and as accurately as possible. The confederate was instructed to respond with an 'up' button press whenever a stimulus appeared to his right and a 'down' button press whenever a stimulus appeared to his left side, respectively.

In order to control for same and different hand responses between the participants and the confederate, the confederate performed half the trials with his right hand on top and the other half with his left hand on top. The instructions for both the confederate and the participant then lead to half the trials being performed with their same hands (hence, the *same hands* condition) and the other half with different hands (i.e. the *different hands condition*).[9] Before each condition, ten practice trials familiarized the participants with the task. These were later excluded from the statistical analysis.

The order of conditions (same hands vs. different hands), the sitting position of the confederate (90° to the left vs. to the right of the participant), as well as the order of mappings (compatible vs. incompatible) was counterbalanced across participants.

Data Analysis

Errors (i.e. trials in which the wrong button or no button at all was pressed) and RTs more than two standard deviations from each participant's condition means were excluded from the analysis. Both the two condition means for correct response RTs and errors for each

[9] If, for instance, the confederate sat to the left of the participant and was instructed to have his left hand on top, while the participant was instructed to respond according to the 'up-right, down-left' mapping, they would then use the same hands in this block (cf. Figure 6).

participant were subjected to separate two-way, repeated measures ANOVAs with the factors Hands (same vs. different) and Compatibility (compatible vs. incompatible).

Results

2.17 % of the trials were removed as errors and 4.01 % were removed as outliers, leaving 93.82% of the raw data as correct response trials.[10] The error analysis did not reveal any statistically significant results for Hands, $F(1, 15) < 1$, $p = .94$, $\eta_p^2 < .01$, Compatibility $F(1, 15) < 1$, $p = .67$, $\eta_p^2 = .01$, or the interaction between the two, $F(1, 15) = 1.38$, $p = .26$, $\eta_p^2 = .08$.

The RT analysis revealed a significant main effect of Compatibility, $F(1, 15) = 16.51$, $p = .001$, $\eta_p^2 = .524$, with RTs being faster in compatible, than in incompatible trials (see Figure 8). There was neither a significant main effect of Hands, $F(1, 15) < 1$, $p = .84$, $\eta_p^2 < .01$, nor a significant interaction effect, $F(1, 15) < 1$, $p = .71$, $\eta_p^2 = .01$.

[10] The removal of outliers did not result in any changes of the significance patterns observed in this experiment.

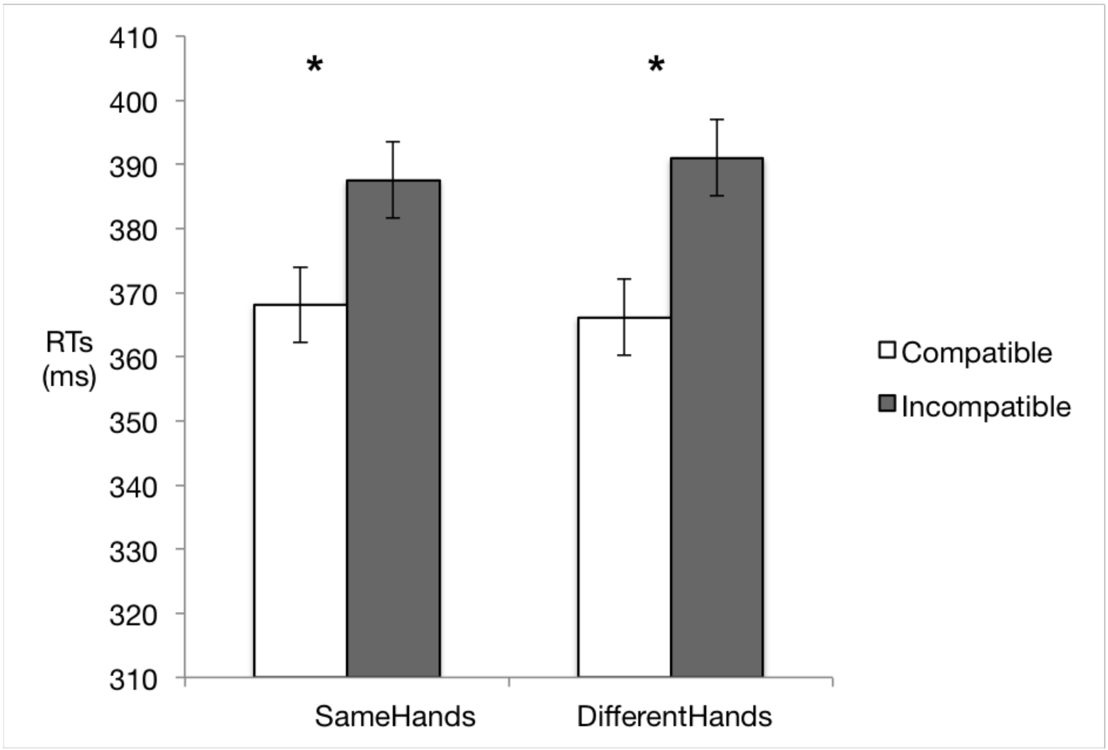

Figure 8. Mean RTs in the same-hands and in the different-hands condition in Experiment 5 (Ch.2). Error bars display within-subject confidence intervals according to Loftus & Masson (1994).

Discussion

The significant main effect of compatibility indicates that participants were generally faster to respond during compatible compared to incompatible trials, regardless of whether the confederate was assigned to a spatially neutral (that is, neither congruent nor incongruent) task and regardless of whether the two were instructed to use the same or different hands to respond.

The results suggest that while it is necessary that the confederate is involved in a task in order for spontaneous VSP taking to occur (cf. Experiment 3), it is sufficient if this task is a neutral SR task (cf. Experiment 5) rather than a spatially matching one (cf. Experiment 1, 2 and 4). Taken together, these results also shed light on the underlying mechanism of spontaneous VSP taking. More specifically, they bolster the claim that participants adopted the confederate's point of view while upholding their own response dimension. In other words, it seems as if participants perceived the stimuli in a similar manner as the confederate did (namely, as ‘left and right’ rather than from their own point of view: ‘up and down’), while they disregarded the way in which the confederate’s responses were orientated. Instead, they seem to have superimposed their own response dimension (i.e. ‘left and right’) onto the confederate’s (‘up and down’). Coding both the stimuli and the necessary responses as ‘left and right’ created a dimensional overlap which could have resulted in the observed compatibility effect.

2.6 General Discussion

The aim of this study was to investigate the underlying factors as well as boundary conditions that could lead to the spontaneous adoption of another person’s VSP during social interactions. In a nutshell, we found that, throughout the course of five experiments, participants reliably adopted the VSP of a co-acting confederate who sat in a 90° angle to the participants, as long as he was perceived as an intentionally acting agent.

More specifically, while performing an orthogonally arranged SR task (i.e., stimuli appeared vertically, while responses were given laterally), participants reliably showed a compatibility effect that corresponded to the confederate's visuospatial perspective. For example, if the confederate sat to the right of the participants, they were significantly faster to respond to an ‘up-right, down-left’ mapping, compared to ‘up-left, down-right’ and vice versa if the confederate sat to the left of the participants (Experiment 1). As the confederate performed a congruent SR task (that is, responding to a left stimulus with a left button press and vice versa) in close proximity to the participant in Experiment 1, we then investigated whether perceiving the confederate acting constituted a boundary condition for the effect to occur.

Experiment 2 demonstrated that regardless of receiving auditory and visual feedback from the confederate, participants again showed the compatibility effect. This result suggests that knowledge about the confederate's task was sufficient for the participants to adopt his VSP. This led to the question whether the confederate needed to perform a task at all or whether his passive presence together with his diverging orientation in relation to the stimuli was enough to trigger spontaneous VSP-taking. Crucially, in Experiment 3 we found that it was necessary that the confederate performed the task in order for the participants to adopt his VSP and hence, that the mere presence of a passive confederate was not sufficient.

While the active vs. passive role of the confederate was crucial, Experiment 4 revealed that spontaneous VSP-taking was largely independent of the overlap of the specific SR mappings between the two actors. Put differently, while encoding the stimulus relative to the VSP of the confederate, participants were not overall faster or slower to respond when using the same or different hands as the confederate.

Finally, Experiment 5 showed that participants adopted their confederate's VSP even if the confederate performed an orthogonal SR task. Thus, it seems as if during the task participants adopted the point of view of the confederate (i.e. perceiving the stimuli as ‘left and right’ rather than from their own point of view, that is, as ‘up and down’), while retaining their own (that is, ‘left and right’) response dimension.

Taken together, the findings of the present study therefore show that participants spontaneously adopted a differing VSP while performing a SR task when there was an intentionally acting agent alongside of them. Importantly, the activity of the confederate constituted a boundary condition for spontaneous VSP-taking to occur. Hence, rather than reflecting an automatic process which was activated whenever there was another agent having a differing perspective on the stimuli, our data suggests that participants instead required sufficient information to believe that the other person was actively involved in the task in order to adopt his VSP.

Spontaneous VSP-taking in our study was tightly connected to the partner´s actions (cf. Creem-Regehr et al., 2013). While our findings are therefore in line with other studies highlighting the link between action and PT (Tversky & Hard, 2009; Mazzarella et al., 2013; Furlanetto et al., 2013, and see Costantini et al., 2011) they also show, for the first time, that action-related VSP-taking can take place even outside of a communicative setting. Furthermore, our results demonstrate an effect of VSP on one's own action planning, extending previous studies that have reported effects of VSP-taking in tasks where participants made judgments about the location of objects or had to indicate what could be seen from a particular perspective.

Even though perspective-taking in our study occurred spontaneously, it could be argued that diverging from one's own VSP in order to adopt somebody else's must nevertheless require extensive processing (Keysar et al., 2000; Keysar et al., 2003; Durmontheil et al., 2010; Mattan et al, *in press*). This would mean that the found compatibility effect was most likely driven by an interference effect during the adoption of the confederate's perspective. If the adoption to another person's perspective is already effortful, then having to deal with an incompatible SR arrangement on top of that must be reflected in particularly increased response latencies on incompatible trials. However, a closer look at the results of Experiment 3 points to a different direction. The post-hoc comparisons of Experiment 3 revealed that the interaction effect between Role of Other and Compatibility was driven by participants being significantly faster to respond during *compatible* trials in which the confederate was *active* compared to each of the other three conditions. Rather than causing interference and increased response latencies, it therefore seems as if adopting the VSP of an active confederate could effectively *facilitate* the processing of the task. While it would be illustrative to explore how the confederate's presence changes the performance compared to an individual baseline, one could argue that the passive condition is a more convincing baseline, as it generally rules out possible social facilitation effects. Thus, one might conjecture that instances of spontaneous VSP-taking in our study may have been driven by facilitative processes. Such mechanisms might be particularly useful in situations where multiple agents have constantly changing perspectives but nevertheless need to coordinate their actions under time pressure (e.g., passing a ball during a soccer match).

Finally, when there could not be any intentional or unintentional interpersonal coordination of actions (as in Experiment 2), we did not find evidence that the lack of these coordination processes diminished the compatibility effect. However, it may be interesting to investigate in future studies whether instructing participants to act in synchrony with the confederate could further boost the effects shown in our study.

We believe that the mechanism underlying the observed effects involves a switch from an egocentric to an allocentric reference frame (Mazzarella et al., 2012). Switching reference frames might have been prompted by the left-right-dimension of the participants' response locations. Experiment 5 indicated that the effect is driven by an overlap between the spatial dimension of the participants' responses (left-right) and the spatial dimension of the stimuli from an allocentric perspective (also left-right) because the effect persisted regardless of the particular spatial arrangement of the confederate's responses. In future studies it would therefore be interesting to determine whether the spatial dimension of the participants' responses is a necessary factor in order to trigger VSP-taking.

In conclusion, the findings of the present study show that participants spontaneously adopted a differing VSP while performing a SR task, given there was an intentionally acting agent alongside of them. In consequence, the current study extends our prior understanding on perspective-taking in two ways. To our knowledge, these are the first results showing that humans adopt another person's VSP by all respects *spontaneously*; that is, in the absence of a communicative context and without being prompted to do so. Secondly, our data suggests that, given the right circumstances, spontaneous VSP-taking might effectively facilitate and speed up spatial alignment processes accruing from dynamic interactions in multi-agent environments.

Chapter 3 –
Another's Visual Access Modulates Spontaneous VSP-Taking

Summary

After having shown first evidence for spontaneous VSP-taking in humans in Chapter 2, Chapter 3 examines whether knowledge about another's visual access systematically modulates spontaneous VSP-taking. To this end, we conducted another study in which participants performed the same spatial compatibility task that was presented in Chapter 2. Importantly though, we manipulated the visual access of the confederate during the task by means of glasses with adjustable shutters that allowed or prevented the confederate from seeing the visual stimuli. The results of two experiments then showed that people only adopted their task partner's VSP if that person had unhindered visual access to the stimuli. Furthermore, provided that the confederate had visual access to the participant's stimuli, VSP-taking occurred regardless of whether the confederate performed the same visual task as the participant (Experiment 1) or a different, auditory task (Experiment 2). The results therefore suggest that knowledge about another's visual access is in fact pivotal for triggering spontaneous VSP-taking while having the same task is not. Finally, we will discuss the possibility that spontaneous VSP-taking can effectively facilitate spatial alignment processes in social interactions.[11]

[11] The full material presented in this chapter was published in JEP:HPP; 2017 Jun;43(6): 1065-1072 (Freundlieb, Sebanz, & Kovács, 2017).

Introduction

Being able to relate to multiple individuals' viewpoints is a key component of social interactions. Our own visual perspective is virtually never perfectly aligned with the perspectives of the people we are interacting with. Instead, we might sometimes even find our perspectives to be opposite to each other. For instance, imagine that you want to draw your friend's attention to the fact that there is an eyelash on her cheek. In telling her you need to take into consideration that her right is your left hand side and vice versa (Kessler & Thomson, 2010). Moreover, often we have practically no time to ponder on the other's perspective but instead need to quickly react in order to successfully interact with each other. Take a basketball player who - just at the right moment - needs to pass the ball at a particular angle to her teammate's appropriate hand so that he can go for an easy layup. In order to successfully interact with others, we need to be able to *spontaneously* understand and integrate information about differing perspectives.

Recent studies suggest that humans have a remarkable ability to take somebody else's perspective. However, as our review of the literature will show, little is known about the mechanisms underlying spontaneous visuospatial perspective-taking, which involves computing how an object is perceived from somebody else's perspective. In particular, does knowledge about another's ability to see the object on which we have a different perspective play a role? Or are there more low-level mechanisms at work that are independent of knowledge about another's visual access? Addressing this question was the aim of the present study.

In the following, we first summarize research showing that people spontaneously process whether another agent can see a target object or not (visual perspective-taking, also referred to as Level-1 perspective-taking in the literature, cf. Flavell, Everett, Croft, & Flavell, 1981). Afterwards, we discuss whether humans also spontaneously compute the location of objects relative to another person, infer what these objects look like from his/her point of view and how that impacts action planning (visuospatial perspective-taking, also referred to as Level-2 perspective-taking, cf. Flavell et al., 1981). Finally, we move on to discuss how knowledge about another person's visual access might modulate the spontaneous adoption of her visuospatial perspective.

Spontaneous Visual Perspective-Taking

Recent research has investigated how and under which circumstances we spontaneously adopt somebody else's perspective (Surtees, Apperly, & Samson, 2016; 2013; Freundlieb, Kovács, & Sebanz, 2015; Furlanetto, Becchio, Samson, & Apperly, 2015; Mazzarella, Hamilton, Trojano, Mastromauro, & Conson, 2012; Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010; Tversky & Hard, 2009). For example, Samson and colleagues (2010) have shown that participants automatically process the content of an avatar's perspective, regardless of whether or not it is important for their task. Specifically, when making judgments about the total amount of objects being visible in a given scene, participants spontaneously computed the number of objects that the avatar could see. The results showed that participants were significantly faster on trials in which their own perspective was consistent with the avatar's perspective, compared to when it was inconsistent (Samson et al., 2010).

It has been argued that in such a set-up participants might not actually process the visual perspective of the human-like avatar, as the same results were obtained when the avatar was exchanged with a mere direction indicating symbol, like an arrow (cf. Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014). If an arrow produced the same results as the avatar, then - according to the alternative explanation - spontaneous visual perspective-taking effects are probably based on general attention-shifting mechanisms, rather than a process that is specifically sensitive to the agentive features and the perspective of somebody else (cf. Heyes, 2014). Some evidence against this alternative hypothesis has recently been put forward by Furlanetto and colleagues (2015) who replicated the original study by Samson et al. (2010) with a crucial modification. Specifically, the avatar wore transparent or opaque goggles. Interestingly, they only found evidence for automatic visual perspective-taking in the presence of an avatar wearing goggles when participants believed the goggles to be transparent but not when they believed the goggles to be opaque (Furlanetto et al., 2015). This suggests that people are sensitive towards the visual access of somebody else when automatically computing the content of what another can see.

Spontaneous Visuospatial Perspective-Taking and Action Planning

Recent studies indicate that in interactive settings people spontaneously compute not only whether another can see an object, but also how an object or scene appears to a task partner (Elekes, Varga, & Kiraly, 2016; Surtees et al., 2016). This has been addressed in studies on *visuospatial perspective* (VSP) taking where the perception of the participant and a second person differs not in terms of the visibility of objects, but in terms of how they appear from two different perspectives. In a study by Surtees, Apperly, and Samson (2016), participants were instructed to judge the magnitude of a single number either sitting alone or opposite to a partner. The results showed that participants' task performance was systematically modulated when sitting opposite to a partner such that responses were significantly faster on trials in which their perspective was consistent with that of their task partner (e.g., on trials in which an '8' or a '5' would be shown), compared to trials in which their point of view was inconsistent with their task partner's (e.g. on trials in which a '6' or '9' would be shown).

Freundlieb et al. (2015) proposed that spontaneous VSP-taking affects not only perceptual judgments but also action planning processes. They found that when participants responded to stimuli arranged vertically from their perspective with left and right responses, they showed a spatial compatibility effect when a task partner was sitting at a 90 degree angle, so that the stimuli were arranged horizontally from the point of view of the task partner. More specifically, participants' reaction time patterns showed that they were faster to respond on trials that were compatible compared to incompatible with regard to the task partner's perspective (Freundlieb et al., 2015). Interestingly, this study also indicated that when acting together, adopting another's VSP can have facilitatory effects on participants' performance.

Visual Access and Sponteanous Visuospatial Perspective-Taking

While evidence for spontaneous VSP-taking has been accumulating, many aspects concerning the underlying mechanisms remain unknown. In particular, it is still unclear to what extent spontaneous VSP-taking is modulated by knowledge about the visual access of another person and how that impacts one's own action planning. Importantly, the study by Furlanetto and colleagues (2015) investigated visual perspective-taking and showed that humans indeed spontaneously encode whether or not an object can be seen by another person. In contrast, VSP-taking (that is, *how* an object is seen by another person) seems to emerge much later in development (Apperly & Butterfill, 2009; Flavell, Everett, Croft, & Flavell, 1981) and has yet to be shown in non-human animals (Call & Tomasello, 2008), indicating that VSP-taking might be cognitively more effortful than visual perspective-taking. Thus, the question remains whether VSP-taking effects can be explained with encoding how an object is perceived by another person, or by lower level processes not entailing such computations. One way to investigate this is to ask whether visual access modulates VSP-taking. Is it crucial for spontaneous VSP-taking that people attribute particular perceptual or knowledge states to the other person?

On the one hand, it could be argued that knowledge about the other's visual access plays a crucial role in triggering VSP-taking. We are sensitive to others' epistemic access from early on in life, as indicated by research in infants showing that their gaze following (Melztoff & Brooks, 2008) and their eye movement behavior in a false belief task (Senju, Southgate, Snape, Leonard, & Csibra, 2011) depends on whether an observed actor is wearing an opaque or transparent blindfold. Rather than merely following directional cues, evidence suggests that adults engage in fairly elaborate computations of what others can see depending on their line of sight (Baker, Levin, & Saylor, 2015). Furthermore, joint attention modulates how people process images of hands (Böckler, Sebanz, & Knoblich, 2011) and faces (Böckler & Zwickel, 2013), leading to a switch from an egocentric to an altercentric reference frame specifically if a task partner is attending to the stimuli. Knowing that another individual can in principle observe the same object from a different VSP might thus constitute a necessary factor for triggering the spontaneous adoption of the other's VSP.

On the other hand, it has been proposed that VSP-taking could be based on an embodied cognitive process during which the self-perspective is physically aligned with the target perspective – regardless of whether or not the target perspective entails seeing the world from a social agent's viewpoint or, say, from a predefined point in space such as an empty chair (see Kessler & Thomson, 2010). For instance, Kessler and Thomson (2010) investigated whether participants used different strategies if they had to adopt the perspective of a differently oriented chair, compared to adopting the perspective of a differently oriented human agent. Their results indicated that participants used the same kind of motoric embodiment (i.e., a computation of the sensory consequences of a mental rotation of the self perspective) in order to change their VSP (Kessler & Thomson, 2010). This suggests that in order to understand where something is located relative to someone or something else, we need not necessarily attribute mental content, or, as Surtees and colleagues (2013) put it: "(…) for me to know that something is to your left is in no way dependent on you representing it as such" (p. 427, Surtees, Apperly & Samson, 2013; cf. also Surtees, Noordzij & Apperly, 2012).

This means that the attribution of a particular perceptual or knowledge state might not be a necessary prerequisite for spontaneous VSP-taking to occur. In particular, in the task used by Freundlieb et al. (2015) the mere presence of a task partner with a different spatial orientation might, by itself, have posed a sufficient cue in order to trigger the adoption of his perspective. While the mere presence of a passive individual did not trigger VSP-taking in this study, it could still be the case that participants mentally rotated themselves into the other's position when he was actively performing a task, without attributing perceptual content to the other.

To further specify the mechanism underlying spontaneous perspective-taking, we investigated whether the visual access of another person affects spontaneous VSP-taking. Specifically, we aimed to disentangle whether spontaneous VSP-taking is merely based on physical alignment processes (the rotation of the self into the target perspective) or whether specific forms of spontaneous VSP-taking depend on knowledge about the visual access of the other in a task where the other's perspective was never mentioned and thus not highlighted in any way.

To this end, we used a task in which participants were seated in a 90° angle to a co-actor. They were instructed to perform an orthogonal stimulus-response compatibility task (SRC, cf. Craft & Simon, 1970; Simon, 1990) on a horizontally mounted computer display (see Figure 9). Given the sitting position of the confederate and the participant, the stimuli could be seen from two different VSPs – they either appeared vertically (from the participant's perspective) or horizontally (from the confederate's perspective). As the participant saw the stimuli along a vertical dimension (up/down) and responded to them on a horizontal dimension (that is, with left/right button presses) there was no spatial overlap. From the confederate's point of view though, stimuli appeared horizontally (left/right) and could therefore coincide with the participant's responses (left/right). Thus, measuring responses according to the spatial position of the stimuli allowed us to test effects of VSP-taking on participants' performance (Freundlieb et al., 2015).

We modulated the visual access of the confederate by using goggles that would either allow or prevent the confederate from seeing the stimuli on the screen. If participants showed a spatial compatibility effect depending on whether or not the confederate had visual access to the stimuli, this would provide evidence that they are relying on the confederate's visual access during spontaneous VSP-taking.

3.1 Experiment 1

To investigate how knowledge about another's visual access modulates spontaneous VSP-taking, we compared two conditions in which the confederate's actions remained the same but his visual access to the stimuli was manipulated. While the participants performed a visual SRC task in both conditions, the confederate performed either a visual task (seeing condition) or was blindfolded and performed an auditory task (blindfolded condition). In both conditions, the confederate was instructed to give the same (right and left button press) responses.

If there was a spatial compatibility effect only in the condition in which the confederate could *see* the stimuli, then this would support the claim that the other's visual access is a necessary factor for triggering spontaneous VSP-taking. In contrast, if participants' responses were unaffected by the confederate having (or *not* having) visual access to the stimuli, then this would suggest that the other's body orientation and the fact that he performed responses along a right-left dimension triggered a mental rotation into his position – independent of his visual access.

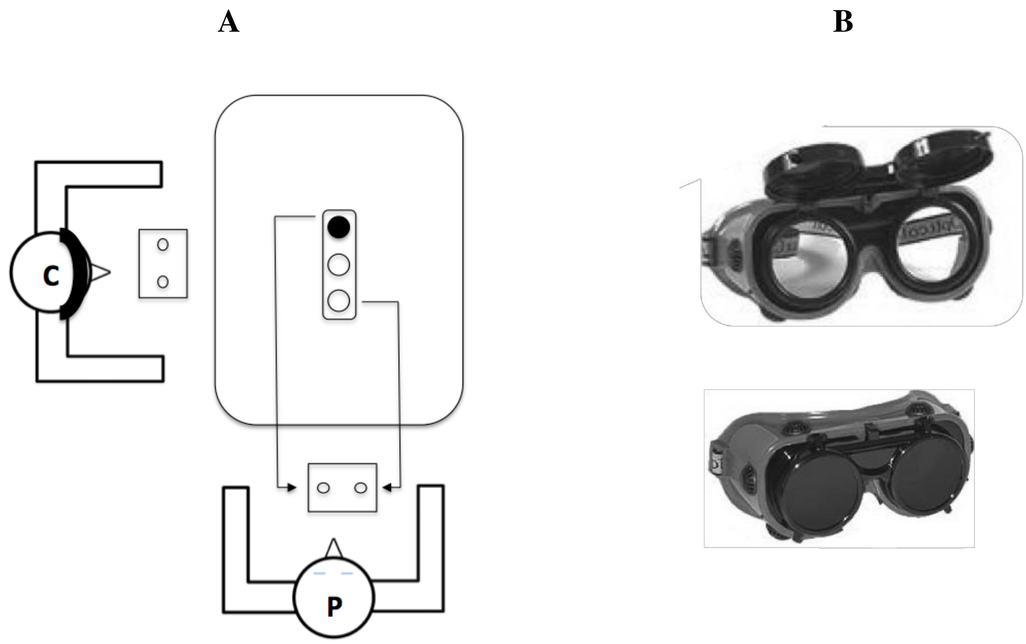

Figure 9. (A) Sketch of the experimental set-up for Experiment 1 (Ch.3). (B) Photograph of the shutter glasses in the transparent and opaque state.
Participant 'P' sat at a 90 degree angle to the confederate 'C', example of the compatible block during the blindfolded condition. The arrows indicate the participant's mapping.

3.2 Method

Participants

18 participants (mean age = 20.8 years, 13 women, 3 left-handed) signed up for this study and received gift vouchers for their participation. Two participants did not meet the inclusion criterion of having more than 90% successful trials within each experimental condition, leaving 16 participants (mean age = 20.7 years, 11 women, 3 left-handed) for the analysis. All participants were naïve to the purpose of the study, reported normal or corrected to normal vision and signed informed consent prior to the experiment.

Stimuli and apparatus

The stimuli consisted of a rectangle (subtending 5.73° of visual angle vertically and 3.27° horizontally) containing three empty circles (each subtending 1.64° of visual angle) at equal distance to each other. During the trials, one out of these three circles (either the one at the top, or the one at the bottom, but never the circle in the middle) then appeared as a black disc in place of the empty circle (see Figure 9A). These two types of stimuli were shown on a horizontally arranged 27" iMac (Mid-2011). The monitor was mounted at a height of about 25 cm from the floor. Responses were given on two button boxes (ioLab Response box), which both the participant and the confederate placed on their lap. The button boxes were partially covered with a piece of carton so that only the two buttons used to respond (i.e. the buttons farthest to the left and right) were visible. Throughout the experiment, the confederate wore a pair of lift-front goggles ('Lux Optical', see Figure 9B). These goggles had small shutters that could either be lifted up (in which case one had unhindered vision through transparent plexiglass), or flapped down (in which case black tape on the shutters blocked vision).

Design and Procedure

For the participant as well as for the confederate, who was oriented in a 90° angle to the participant, viewing distance was ≈ 70cm. Throughout the entire study, the same young adult male acted as the confederate. Each trial started with the presentation of a fixation cross (subtending 0.66° of visual angle, presented in the centre of the screen) for 350 ms. Subsequently, the screen turned blank for 100 ms after which, in a randomized manner, one of the two stimuli (top black disc vs. bottom black disc) was shown for 1200 ms. Participants performed two conditions (*blindfolded* and *seeing)* with two blocks (*compatible* and *incompatible*) each. Each block contained 100 trials and participants were asked to respond as fast and as accurately as possible. To establish different compatibility relations, we varied the sitting position of the confederate and the stimulus-response-mapping of the participants.

In one half of the experiment, participants were instructed to respond to the appearance of the top black disc by pressing the right button on the button box with their right index finger and to respond to the bottom black disc by pressing the left button with their left index finger, respectively. In the other half, the mapping was reversed and they were thus instructed to respond to the appearance of the top black disc with a left and to the appearance of the bottom black disc with a right button press. In the *compatible* condition, the mapping of the participant concurred with the spatial orientation of the confederate, while in the *incompatible* condition it did not. For instance, if the confederate sat 90° to the left of the participant, participants were instructed with the 'up-left, down-right' mapping in the compatible, and with the 'up-right, down-left' mapping in the incompatible block (see Figure 9A).

Crucially, the task of the confederate changed throughout the experiment. During the *seeing* condition the confederate was asked to flap the shutters of his goggles up and respond to the visual stimuli on the screen. Specifically, the confederate was instructed to respond with a left button press if a black disc appeared - from his point of view - on the left side of the screen and with a right button press if a black disc appeared on the right side of the screen. In the other half of the trials (during the *blindfolded* condition) the confederate was given a pair of headphones and was asked to flap the shutters of his goggles down to respond to auditory stimuli. He was instructed to respond to a high tone with a 'right' and to a low tone with a 'left' button press on the button box. The high and low tones appeared at the same time as the visual stimuli but could only be heard through the headphones. To ensure that the confederate performed the same actions in the two conditions, the high tone always appeared together with the up stimulus, whereas the low tone appeared together with the low stimulus.

Before the blindfolded condition started, a short practice block was conducted, during which the participant and the confederate switched tasks. This practice block was conducted in order to familiarize participants with the task involving the goggles. Hence, for the duration of the practice block, the experimenter asked the participants to sit where the confederate would sit later on (and vice versa). The instructions that were given (both for the auditory and for the visual task) during this practice block were identical to the instructions given during the experimental condition. After ten practice trials, the practice block was over and both the participant and the confederate were instructed to swap places, so that for the experimental trials, the participant always performed the visual SRC task while the confederate consecutively performed the auditory task as well as the visual SRC task. Importantly, participants could not hear the tones to which the confederate responded when he was performing the auditory task. Before each block, ten practice trials familiarized the participants with the task. These were later excluded from the statistical analysis.

The order of conditions (blindfolded vs. seeing), the position of the confederate (90° to the left vs. to the right of the participant), as well as the order of mappings (compatible vs. incompatible) was counterbalanced across participants.

Data Analysis

Errors (i.e. trials in which the wrong button or no button at all was pressed) and RTs more than two standard deviations away from each participant's condition means were excluded from the RT analysis. Both the two condition means for correct response RTs and errors for each participant were subjected to separate two-way, repeated measures ANOVAs with the factors Vision Other (blindfolded vs. seeing) and Compatibility (compatible vs. incompatible).

Results

For the RT analysis 2.03 % of the trials were removed as errors and 3.77 % of the trials were removed for being more than two standard deviations away from each participant's condition means, leaving 94.2% of the raw data as correct response trials. The error analysis revealed a tendency towards a significant main effect of Compatibility, $F(1, 15) = 15$, $p = .06$, $\eta_p^2 = .23$) showing that participants made more errors during compatible ($M = 2.25\%$ errors) than during incompatible ($M = 1.06\%$ errors) trials.[12] Neither the main effect of Vision Other, $F(1, 15) < 1$, $p = .99$, $\eta_p^2 < .01$., nor the interaction between Vision Other and Compatibility, $F(1, 15) = 3.08$, $p = .10$, $\eta_p^2 = .17$, was significant.

The RT analysis revealed a significant interaction between Vision Other and Compatibility, $F(1, 15) = 5.35$, $p = .03$, $\eta_p^2 = .26$. In post-hoc analyses, pairwise comparisons showed a significant difference in RTs between the compatible ($M = 349$, $SD = 26$) and incompatible ($M = 363$, $SD = 29$) blocks only in the seeing condition, $t(15) = - 2.63$, $p = .02$, two-tailed, but not in the blindfolded condition ($M = 359$, $SD = 38$ and $M = 362$, $SD = 42$, for compatible and incompatible trials, respectively), $t(15) = - .53$, $p = .60$, two-tailed (see Figure 10). None of the other pairwise comparisons were significant (all ps > .15). There was neither a significant main effect of Vision Other $F(1, 15) < 1$, $p = .53$, $\eta_p^2 = .03$, nor of Compatibility, $F(1, 15) = 2.9$, $p = .11$, $\eta_p^2 = .16$.

[12] The fact that there was a tendency towards a significant main effect for Compatibility in the error analysis could indicate that there was a speed-accuracy trade-off. However, as participants were not generally faster in compatible vs. incompatible trials (there was a difference between compatible and incompatible trials only in the seeing condition), such a speed-accuracy trade-off is not supported by the data.

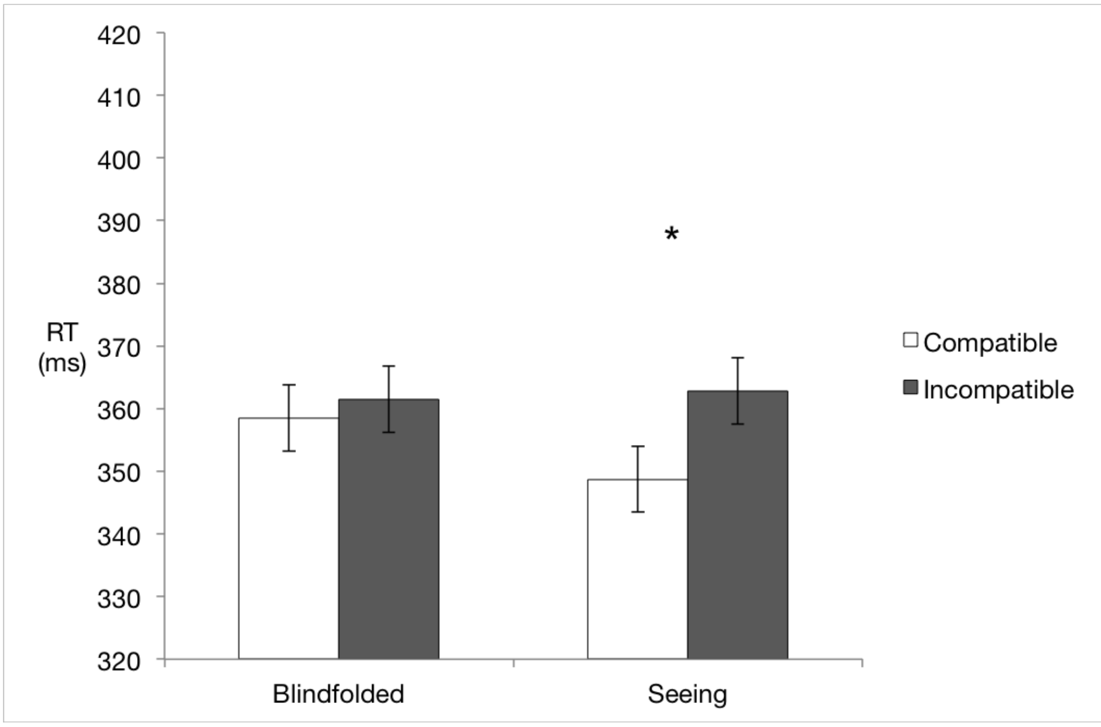

Figure 10. Meant RTs in the blindfold and in the seeing condition in Experiment 1 (Ch.3).
Error bars display within-subject confidence intervals according to Loftus & Masson (1994).

Discussion

Experiment 1 showed a spatial compatibility effect selectively when the confederate was able to see and responded to the visual stimuli. This suggests that participants spontaneously adopted the other’s VSP depending on the other’s visual access. Because the confederate's ability to see the stimuli was necessary in order to evoke the compatibility effect, one could therefore contend that participants indeed computed how the stimuli were *seen* by the confederate. Because the confederate's position/posture did not change between conditions, it can be ruled out that it was merely the directionality of his front features (that is, the direction of body, forehead, nose etc.) that had triggered spontaneous VSP-taking.

However, in Experiment 1 the confederate performed a visual task in the seeing condition and an auditory task in the blindfolded condition, while the participant performed a visual task in both conditions. It could be argued that the data of Experiment 1 can be explained in virtue of the (changing) tasks that both the participant and the confederate needed to perform. In particular, it might be possible that spontaneous VSP-taking only occurs if both people are performing the same task. If this was the case, then the absence of the effect during the blindfolded condition in Experiment 1 could be merely due to the participant and the confederate performing two different tasks – the former a visual, the latter an auditory task.

Second, one could argue that if participants monitored the confederate's task performance, this led to differences in terms of general task complexity between the two conditions. While in the seeing condition participants were performing the very same visual SRC task as the confederate, in the blindfolded condition monitoring the other's task implied inferring the auditory stimulus based on the confederate's key presses. In order to better understand the role of task similarity and to rule out potential confounds we conducted Experiment 2.

3.2 Experiment 2

Experiment 2 investigated the alternative explanation that spontaneous VSP-taking was hinging on differences between the tasks that the participant and the confederate needed to perform in the seeing and in the blindfolded conditions of Experiment 1, rather than on knowledge about the other's visual access. To rule out potential confounds, the confederate now performed the same auditory task throughout the two conditions. The predictions of Experiment 2 were as follows: if spontaneous VSP-taking depends on performing the same task as the other person, then the differences between conditions should disappear when the participant and the confederate perform two different tasks in both conditions. Importantly, this should be independent of whether or not the confederate has visual access to the stimuli. In contrast, if spontaneous VSP-taking depends on whether or not the other has unblocked visual access to the stimuli, then the compatibility effect should only occur when the confederate has visual access to the stimuli – regardless of the fact that the confederate's task is different from the participant's.

Method

Participants

18 new participants (mean age = 21.83 years, 9 women, 1 left-handed) signed up for this study and received gift vouchers for their participation. Two participants did not meet the inclusion criterion of having more than 90% successful trials within each experimental condition, leaving 16 participants (mean age = 21.75 years, 7 women, 1 left-handed) for the analysis. All participants were naïve to the purpose of the study, reported normal or corrected to normal vision and signed informed consent prior to the experiment.

Stimuli and apparatus

The stimuli and the apparatus were identical to Experiment 1.

Procedure

Participants performed two conditions (blindfolded and seeing*)* with two blocks (compatible and incompatible), respectively. Each block contained 100 trials. The participants' task was identical to Experiment 1 and they were asked to respond as fast and accurately as possible. In contrast to Experiment 1, the confederate now performed an auditory task throughout the entire experiment. The occurrence of the type of tone (high vs. low) was independent of the location of the visual stimulus in Experiment 2. That is, the appearance of a stimulus at the upper side of the screen could now co-occur with a high or with a low tone (with the same holding true for the stimuli at the lower end of the screen). The tones were played through headphones so that they were only audible for the confederate but not for the participant. Other than that the procedure was identical to Experiment 1. In half of the trials the confederate's shutters were flapped down (blindfolded), while in the other half the shutters were flapped up (seeing) and the confederate had unblocked visual access to the stimuli on the screen. Before each condition, ten practice trials familiarized the participants with the task. These were later excluded from the statistical analysis.

The order of conditions (blindfolded vs. seeing), the position of the confederate (90° to the left vs. to the right of the participant), as well as the order of mappings (compatible vs. incompatible) was counterbalanced across participants.

Data Analysis

Errors (i.e. trials in which the wrong button or no button at all was pressed) and RTs more than two standard deviations from each participant's condition means were excluded from the RT analysis. Both the two condition means for correct response RTs and errors for each participant were subjected to separate two-way, repeated measures ANOVAs with the factors Vision Other (blindfolded vs. seeing) and Compatibility (compatible vs. incompatible).

Results

For the RT analysis 1.11 % of the trials were removed as errors and 4.03 % were removed for being more than two standard deviations away from each participant's condition means, leaving 94.86% of the raw data as correct response trials. The error analysis did not show any effect of Compatibility, $F(1, 15) < 1$, $p = .99$, $\eta_p^2 < .01$), Vision Other, $F(1, 15) < 1$, $p = .79$, $\eta_p^2 < .01$., or the interaction between the two, $F(1, 15) < 1$, $p = .82$, $\eta_p^2 < .01$.

The RT analysis revealed a significant interaction between Vision Other and Compatibility, $F(1, 15) = 18.93$, $p < .01$, $\eta_p^2 = .56$. In post-hoc analyses, pairwise comparisons showed a significant difference in RTs between the compatible ($M = 366$, $SD = 27$) and incompatible (M = 401, SD = 42) blocks only in the seeing condition, $t(15) = -3.69$, $p = .002$, two-tailed, but not in the blindfolded condition (M = 397, SD = 61 and M = 382, SD = 25 for compatible and incompatible trials, respectively), $t(15) = 1.3$, $p = .20$ (see Figure 11). Furthermore, there was a significant difference in RTs between the seeing compatible (M = 366, SD = 27) and blindfolded incompatible (M = 382, SD = 25) trials, $t(15) = 2.58$, $p = .02$, two-tailed, and a tendency towards a difference between seeing compatible (M = 366, SD = 27) and blindfolded compatible (M = 397, SD = 61) trials, $t(15) = 2.05$, $p = .058$, two-tailed. The two remaining pairwise comparisons were not significant (both ps > .15). Finally, there was neither a significant main effect of Vision Other $F(1, 15) < 1$, $p = .61$, $\eta_p^2 = .02$, nor of Compatibility, $F(1, 15) = 1.3$, $p = .27$, $\eta_p^2 = .08$.

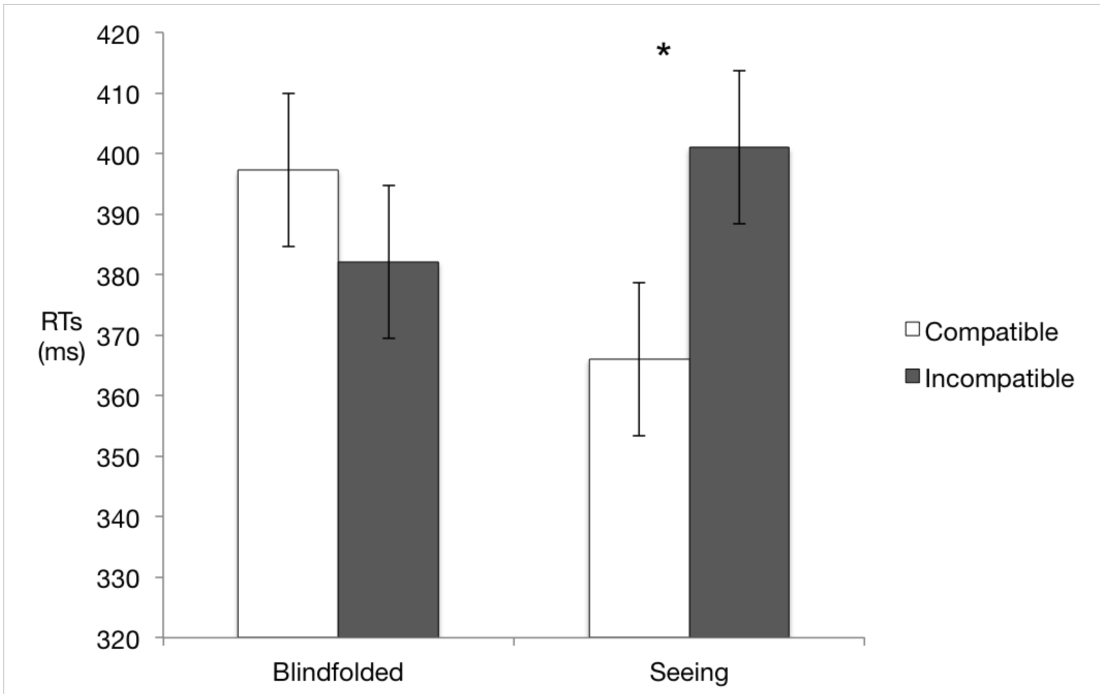

Figure 11. Mean RTs in the blindfolded and in the seeing condition in Experiment 2 (Ch.3). Error bars display within-subject confidence intervals according to Loftus & Masson (1994).

Discussion

The results suggest that participants adopted the confederate's VSP if he had unblocked visual access to the stimuli. This confirms the results obtained in Experiment 1, indicating that VSP-taking depends on others' visual access. Participants showed a spatial compatibility effect when the confederate could see the visual stimuli they responded to, even though the confederate performed an auditory task that was independent of the participants' visual task. These results are in line with findings by Surtees and colleagues (2016), showing that spontaneous VSP-taking is not restricted to situations where people perform the same tasks. While Surtees et al. found evidence for spontaneous VSP-taking when two participants responded to two different aspects of the same visual stimuli (e.g., one participant responded to number magnitude while the other responded to a surface feature, cf. Surtees et al., 2016), our results demonstrate that spontaneous VSP-taking can occur even when the tasks are performed in different sensory modalities.

Finally, the data of Experiment 2 suggests that compared to all other conditions, participants were particularly fast to respond during the compatible trials in the seeing condition. This is in line with earlier findings (Freundlieb et al., 2015) supporting the claim that, under certain circumstances, spontaneous VSP-taking might have facilitatory effects.

3.3 General Discussion

The aim of this study was to investigate the mechanisms underlying spontaneous VSP-taking. Specifically, we examined whether knowledge about another person's visual access systematically modulates perspective-taking. To this end we used a task that has previously been shown to elicit spontaneous VSP-taking (Freundlieb et al., 2015). We manipulated the visual access of the other person (a confederate) during the task by means of glasses with adjustable shutters that allowed or prevented the confederate from seeing the visual stimuli.

The results show that participants only adopted the other's VSP if he had unhindered visual access to the stimuli but regardless of whether or not he performed the same visual task or a different auditory task. Our study therefore suggests that spontaneous VSP-taking is indeed modulated by knowledge about another person's visual access.

Our findings contribute to current debates about the mechanisms underlying perspective-taking. It has been suggested that in addition to a comparatively slow but elaborate mentalizing system, humans possess another 'simple perspective-taking system' (cf. Samson et al., 2010), which enables them to quickly and efficiently process what another agent can see (cf. Qureshi, Apperly, & Samson, 2010; Ramsey, Hansen, Apperly, & Samson, 2013), especially in contexts where the other is performing actions (cf. Surtees et al., 2016; Frischen, Loach and Tipper, 2009; Tversky & Hard, 2009; Zwickel, 2009). In contrast, it has also been proposed that much of the evidence that has been connected to the concept of perspective-taking can be captured more parsimoniously through domain-general processes such as attention reorienting or spatial referencing (Heyes, 2014; Santiesteban et al., 2014; Santiesteban, Shah, White, Bird, & Heyes, 2015). On this account, if we observe somebody else confronted with a different amount of target objects than ourselves (like in the study conducted by Samson et al., 2010) we might not actually have to process *that* he or she can actually see and consequently represents the seen objects. Instead, it could be that domain-general cognitive mechanisms pick up on salient features (such as the other's body orientation), resulting in attentional reorienting and producing the same kind of responses that are typically ascribed to perspective-taking or implicit mentalizing (Heyes, 2014).

Some evidence against these low-level explanations has been provided by Furlanetto and colleagues (2015) who showed that information about the visual access of the other person is in fact pivotal in order to engage in automatic visual perspective-taking and hence, that the mere exhibition of front features (i.e., the direction of the body, forehead, nose etc.) is not sufficient to trigger visual perspective-taking (Furlanetto et al., 2015). The findings of the present study further extend this claim to the domain of VSP-taking. Using a visuospatial paradigm where the other's perspective was not prompted in any way (participants never had to consider the other's perspective, in contrast to Furlanetto et al., 2015), we showed that, beyond automatically processing the content of what another agent can see, humans are also able to spontaneously process *how* something is seen from another person's point of view. Furthermore, while other studies have reported effects of perspective-taking in tasks that required participants to make judgments about the location of objects (Tversky & Hard, 2009; Kessler & Thomson, 2010) or had to indicate what could be seen from a particular perspective (Furlanetto et al., 2015; Samson et al., 2010), our results suggest that participants adopted another's VSP by all means spontaneously, that is, without being prompted to do so. Importantly, spontaneous VSP-taking seems to hinge not only on the other person being actively engaged in a task (cf. Surtees et al., 2016; Freundlieb et al., 2015; Frischen et al., 2009) but also on the other person having visual access to the stimuli. Only if participants knew that the other had unhindered visual access to the stimuli did they spontaneously adopt his perspective and processed the stimuli as if they were seen from the other person's perspective. Thus, we believe this is the first study that shows how visual access triggers the spontaneous integration of somebody else's VSP into one's own action planning.

We believe that the mechanism underlying the observed effects entails that participants shifted from processing the scene according to their own point of view (or *egocentrically*) to processing the scene from the other person's point of view (or *altercentrically*, cf. Samson et al., 2010; Ramsey et al., 2013). Previous studies have already shown that task performance can be affected in the presence of another person whose viewpoint differs from our own (Furlanetto et al., 2015; Ramsey et al., 2013; Conson, Mazzarella, Donnarumma and

Trojano, 2012; Böckler et al., 2011; Samson et al., 2010), especially when the other person is perceived as potentially interacting with the object in the common focus of attention (Surtees et al., 2016; Freundlieb et al., 2015; Furlanetto, Cavallo, Manera, Tversky and Becchio, 2013; Mazzarella et al., 2012; Tversky & Hard, 2009; Frischen et al., 2009). The switching of reference frames in our study might have been prompted by the fact that the other person was performing a task while having visual access to the stimuli. Processing the stimuli in an altercentric manner then led to a spatial overlap between the left-right dimension of the stimuli and the left-right dimension of participants' responses. Finally, this overlap is reflected in the spatial compatibility effect that we observed in both experiments.

Functionally, such a mechanism could be helpful during interpersonal coordination as it could facilitate the integration of diverging spatial perspectives into one common format. Specifically, the spontaneous integration of somebody else's VSP into one's own action planning might serve the function of aligning actions that are performed in close vicinity – but from different VSPs (cf. Creem-Regehr, Gagnon, Geuss, & Stefanucci, 2013). The fact that we found evidence for such an integration even when both actors performed different tasks suggests that this mechanism is quite general, i.e., it does not depend on performing the same task together — as long as the other person has visual access to the same stimuli and is an intentionally acting agent (see Freundlieb et al., 2015; Surtees et al., 2016). Interestingly, a closer look at Experiment 2 of this study reveals that adopting the other's VSP actually sped up participants' performance during the task. Numerically, this pattern also seems to be present in Experiment 1, however, the statistical comparison failed to reach significance. Thus, further experiments are required in order to make a more compelling argument for the hypothesis that, given the right circumstances, spontaneous VSP-taking can effectively facilitate spatial alignment processes, which are required in many social interactions (cf. Freundlieb et al., 2015).

In conclusion, we found that knowledge about another person's visual access systematically modulated the spontaneous integration of another person's VSP into one's own action planning. Our findings show that participants only adopted the other person's VSP if he had unhindered visual access to the stimuli but regardless of whether or not he performed the same task or a different task. Furthermore, our data suggests that, when people perform a task together, adopting the other's VSP might be facilitating and possibly lead to improved task performance. In turn, this might assist with interpersonal coordination in situations where we need to quickly integrate the diverging perspectives of multiple agents.

Chapter 4
Evidence for Spontaneous VSP-Taking in Mental Space

Summary

Most of the evidence for visuospatial perspective-taking (including the evidence from Chapter 2 and 3) comes from tasks where an observed agent or task partner can physically act upon objects. **Chapter 4** investigates whether spontaneous VSP-taking extends to mental space where another's perspective matters for mental activities rather than physical actions. Specifically, a study will be presented in which participants sat a 90° angle to a confederate and performed a semantic categorization task on written words. From the participants' point of view, words were always displayed vertically, while for the confederate, these words either appeared the right way up or upside down, depending on the confederate's sitting position. The results then showed that participants took longer to categorize words that were upside down for the confederate, suggesting they adopted the confederate's VSP without being prompted to do so. Importantly, the effect disappeared if the other's visual access was impeded by opaque goggles. This demonstrates that humans show a spontaneous sensitivity to others people's VSP in the context of mental activities such as joint reading.[13]

[13] The full material presented in this chapter is currently under revision in *Psych. Science*.

Introduction

From passing the basketball to a fellow team-mate, to handing over a knife at the dinner table, being able to adopt other people's visuospatial perspectives (VSP) is pivotal for successfully engaging in a large variety of social interactions. Recent research has provided evidence that people adopt others' VSP spontaneously, computing the relative location of objects from another's orientation without being prompted to do so. We seem to be equipped with mechanisms allowing us to spontaneously take into account not only *whether* somebody else can or cannot see a certain object (visual perspective-taking, or Level-1 perspective-taking, see Flavell, Everett, Croft, & Flavell, 1981) but also *how* objects look like from another's point of view (visuospatial perspective-taking, or Level-2 perspective-taking, cf. Flavell, et al. 1981).

For example, when being asked to give verbal descriptions of the spatial relation among an array of objects, observers spontaneously adopted the VSP of another person facing them (Lozano, Hard, & Tversky, 2007; Tversky & Hard, 2009; cf. also Cavallo, Ansuini, Capozzi, Tversky, & Becchio, 2016). Furthermore, when participants were asked to indicate the spatial locations of stimuli arranged vertically in front of them with left and right responses, and a task partner was sitting at a 90 degree angle next to them, they spontaneously adopted the other person's spatial reference frame, processing the stimuli in terms of the other's left and right (Freundlieb, Kovács, & Sebanz, 2016; Freundlieb, Sebanz, & Kovács, 2017).

Nearly all of the evidence for spontaneous VSP-taking comes from tasks where an observed agent or task partner could physically act upon objects. In these tasks the physical location of an object (say, an apple, Cavallo et al., 2016) varied along a spatial dimension (e.g. appeared to the right vs. to the left of somebody else) and participants' left vs. right responses reflected how they, or the other person, would physically interact with the object.

However, it is unknown whether spontaneous VSP-taking extends to mental space[14] in which spatial relations matter for cognitive processes rather than for physical actions. For example, when the newspaper is oriented at a right angle from you at the breakfast table, will it be easier for you to read its headlines if the paper happens to be aligned with your partner's perspective? Reading is a prototypical case of a mental activity where objects (words) are manipulated by the mind rather than by our hands.

Some first evidence for VSP-taking in mental space comes from a task involving numerical cognition (Surtees, Apperly, & Samson, 2016; see also Elekes, Varga, & Király, 2016). In a joint numerical judgement task, participants were slower to indicate whether a number was smaller or larger than five when the numerical value of a digit was different for a task partner sitting opposite (e.g. on trials where they saw a '6' while their partner saw a '9'). This indicates that participants also computed the symbol from the other's viewpoint. However, importantly, as participants were asked to respond to smaller numbers with a left and to larger numbers with a right response, one could still argue that these results are based on the spatial-numerical associations of response codes (SNARC, cf. Dehaene, Bossini, & Giraux, 1993) and thus, still strongly relate to the action space.

To investigate whether spontaneous VSP-taking occurs not only in physical but also in mental space, we developed a novel task where participants were required to read words in order to perform a semantic categorization task. Across three experiments we asked whether participants are faster in processing words when they are oriented such that they could be easily read by another individual, compared to being in an orientation that is the same from the participants' point of view, but difficult to read from another's perspective.

14 Note that we use the term mental space to highlight the contrast to the physical (inter-) action space. It therefore differs from the notion of mental space as defined by Fauconnier (1994).

4.1 Experiment 1

Participants sat at a 90° angle to a confederate and performed a semantic categorization task with words being displayed on a horizontally mounted computer screen (see Figure 12). The stimuli always belonged to one of two categories and participants were instructed to only respond to stimuli from their own category, and not to respond to stimuli from the other category. All word stimuli were displayed in the same orientation (90° angle clockwise) from the participants' perspective. The confederate sat to the right or to the left side of the participant. From the confederate's perspective, words thus either appeared the right way up (*congruent condition*) when he sat to the participant's left, or upside down (*incongruent condition*) when he sat to the participant's right. If participants spontaneously adopted the confederate's VSP, then it should be easier for them to read words that are the right way up for the confederate and harder to read words that are upside down for the confederate, resulting in a congruency effect.

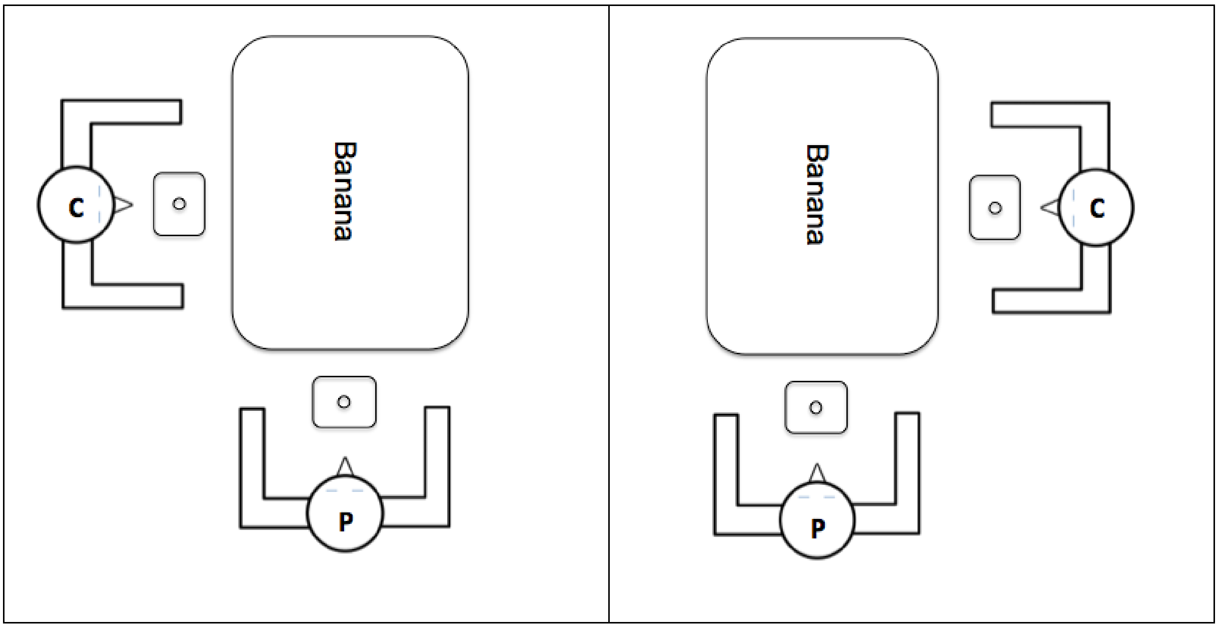

Figure 12. Experimental set-up for Experiment 1 (Ch.4).
Participant (P) sits at a 90° to the confederate (C). Left panel: example of a trial in the congruent condition. Right panel: example of a trial in the incongruent condition.

Method

Participants

We based our sample size on a previously published study testing visuospatial perspective-taking in a paired samples design (Freundlieb et al., 2016). Prior to data collection, we decided to test 16 participants and set the significance level to $\alpha = .05$. 16 participants (mean age = 20.06 years, 12 women, 15 right-handed) signed up for this study and received gift vouchers for their participation. All were naïve to the purpose of the study, reported normal or corrected to normal vision and signed informed consent prior to the experiment. All 16 participants met the inclusion criterion of having more than 90% successful trials in each experimental condition.

Stimuli and apparatus

The stimuli consisted of single nouns in Hungarian (subtending between 3.3° and 7.3° of visual angle, depending on the length of the word, see Appendix). Each word item belonged to one of two categories: animals or fruit/vegetables. Each of the two categories contained 32 items that had been used in a prior study on social memory and were controlled for frequency (see Elekes et al., 2016, and Table A1 in the Appendix presents a list of all word items). In order to rule out carry-over effects between the two experimental conditions, we randomly split the 32 items from each of the two categories in halves, resulting in 4 sub-lists (animals-1, animals-2, fruit/vegetable-1, fruit/vegetable-2), each containing 16 items. This way, participants responded to a unique list of word items in each of the two conditions. During the trials, single word items were always presented in the same orientation (90° clockwise from the participants' perspective) and at the same central position on a horizontally arranged 27" iMac (Mid-2011) (see Figure 12). The monitor was mounted at a height of about 25 cm from the floor. Responses were given on two button boxes (ioLab Response box), which both the participant and the confederate placed on their lap. The button boxes were partially covered with a piece of carton so that only the button used to respond (i.e. the most central button) was visible.

Design and Procedure

Viewing distance to the screen was ≈ 70cm, both for the participant and for the confederate, who was oriented in a 90° angle to the participant. A young adult male acted as the confederate. During the instruction phase the experimenter assigned both the participant and the confederate to one of the two categories (animals vs. fruit and vegetables) and asked them to respond with a button-press only to word items from their own category and not to respond to word items from the other's category (*go/no-go-task)*. Each trial started with the presentation of a fixation cross (subtending 1.31° of visual angle, presented in the center of the screen) for 350 ms. Subsequently, the screen turned blank for 100 ms after which a word item was shown for 1200 ms. The word items were randomly chosen from two sub-lists (e.g. animals-1 and fruit/vegetable-1 in the first condition and then animals-2 and fruit/vegetable-2 in the second condition) and each sub-list was - consecutively - repeated four times per condition, with items presented in a random order. Participants performed two conditions (congruent and incongruent) each containing 128 trials (2 [categories] x 16 [items per sub-list] x 4 [repetitions of each sub-list]). They were asked to respond as fast and as accurately as possible and not to tilt their heads during the experiment.

To establish different congruency relations, we varied the sitting position of the confederate. While the participant always sat at the narrow end of the rectangular screen, the confederate switched between the two long ends during the experiment. In the *congruent* condition the confederate sat to the participant's left, so that words were oriented towards him, or the right way up. In contrast, in the *incongruent* condition the confederate sat to the participant's right, so that words appeared upside-down (see Figure 12). Before each condition, eight practice trials familiarized the participants with the task. These were later excluded from the statistical analysis.

The order of conditions (congruent vs. incongruent), the assigned category (animals vs. fruit and vegetables) as well as the starting sub-list (animals-1 vs. animals-2 vs. and fruit and vegetables-1 vs. fruit and vegetables-2, respectively) was counterbalanced across participants.

Data Analysis

Data was only collected for participants. Errors (i.e. missed button presses during participants' own trials or button presses during the confederate's trials) and RTs more than two standard deviations from each participant's condition means were excluded from the RT analysis. Both the two condition means for correct response RTs and errors for each participant were subjected to separate two-tailed, paired-samples *t*-tests.

Results

For the RT analysis 0.34% of the trials were removed as errors and 4.49% were removed for being more than two standard deviations away from each participant's condition means, leaving 95.17% of the raw data as correct response trials. Generally, the removal of these outliers did not result in changes of the significance patterns observed in this study. Comparing the number of errors in the congruent vs. incongruent conditions showed that participants made significantly more errors in the congruent (M = 1.56, SD = 1.15) compared to the incongruent condition (M = 0.56, SD = 0.81), $t(15) = 4.14$, $p = .001$.[15]

[15] The fact that there was a significant effect of Congruency in the error analysis could indicate that there was a speed-accuracy trade-off. However, this pattern was not observed in subsequent experiments. As participants

The RT analysis revealed that participants were significantly faster in the congruent (M = 578, SE = 15.29) than in the incongruent (M = 618.4, SE = 19.47) condition; $t(15) = 4.49$, $p < .001$, two-tailed (see Figure 13). A post-hoc power analysis (using G* Power, cf. Faul, Erdfelder, Lang, & Buchner, 2007) revealed that, given the mean of difference M_z = 40.33, the SD of difference SD_z = 35 and the effect size of d = 1.13, we achieved a power of 1 - β = .98. In order to test whether either the specific category assigned to the participants (animals vs. fruit and vegetables), the order of conditions (starting with the congruent vs. the incongruent condition), or the order of the sublists (animals-1, animals-2, fruit/vegetable-1, fruit/vegetable-2) influenced the results, a repeated measures ANOVA was conducted with congruency as a within-subjects factor and category, order of condition and order of sub-lists as between-subjects factors. The results yielded only a main effect of congruency, $F(1, 11) = 21.11$, $p = .001$, $\eta_p^2 = .66$, but no effect of category, $F(1, 11) < 1$, $p > .250$, $\eta_p^2 < .01$, order of conditions, $F(1, 11) < 1$, $p > .250$, $\eta_p^2 = .06$, order of sub-lists, $F(1, 11) < 1$, $p > .250$, $\eta_p^2 = .09$; or any two-way interactions between congruency and the between-subjects factors, all Fs < 1.8, ps $> .211$, $\eta_p^2 < .24$.

These results suggest that participants spontaneously adopted the other's VSP when performing a semantic categorization task together, which facilitated the processing of words oriented such that the confederate could easily read them and/or impaired processing of words that were in the same orientation from the participant's point of view, but were oriented upside down from the confederate's perspective. An open question is whether the active engagement of the confederate was necessary for triggering spontaneous VSP-taking in mental space.

generally committed very few errors (1.2% and 0.4% in the congruent and incongruent condition, respectively) the speed-accuracy trade-off in this experiment is unlikely to account for the robust effect in RTs.

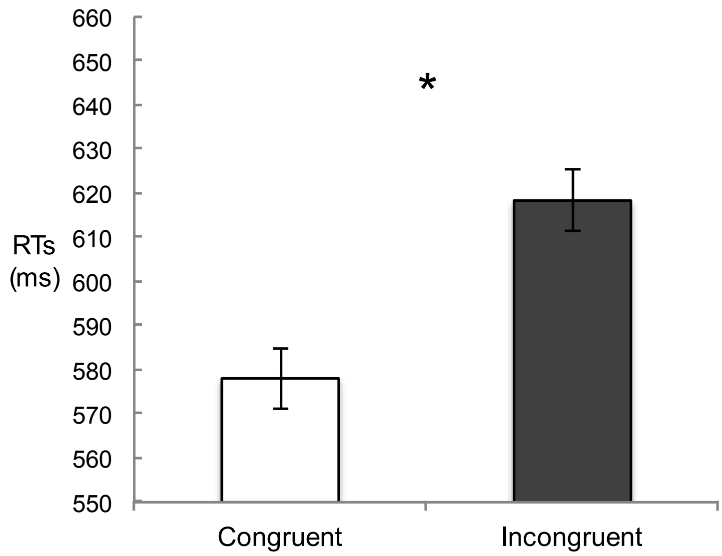

Figure 13. Mean RTs in the congruent and in the incongruent condition in Experiment 1 (Ch.4).
Error bars display within-subject confidence intervals according to Loftus & Masson (1994).

4.2 Experiment 2

Experiment 2 asked whether the mere presence of another person with a diverging VSP is sufficient for participants to spontaneously adopt the other's VSP. In one block the confederate performed the same task as in Experiment 1, while in the other block he was instructed to just watch the stimuli on the screen. As reading is a mental activity that does not necessarily manifest in physical actions, a passive individual can still engage in it. Therefore, we predicted that the presence of a passive confederate with a divergent VSP would be sufficient for participants to adopt his VSP, leading to a congruency effect.

Method

Participants

Prior to data collection, we decided to obtain data from 32 participants and set the significance level to $\alpha = .05$. Changing the experimental paradigm to a 2x2 factorial design lead us to double our initial sample size. This sample size is identical to a previously published study on VSP-taking that used a similar factorial design (Surtees et al., 2016). 33 participants (mean age = 21.68 years, 20 women, 29 right-handed) signed up for this study and received gift vouchers for their participation. One participant with severely reduced vision forgot to bring his glasses and was therefore excluded from the analysis. Each of the 32 participants (mean age = 21.66 years, 20 women, 28 right-handed) was naïve to the purpose of the study, reported normal or corrected to normal vision and signed informed consent prior to the experiment.

Stimuli and apparatus

The stimuli and the apparatus were identical to Experiment 1.

Procedure

Participants performed the two conditions (congruent vs. incongruent) in two different blocks (other-active vs. other-passive). Each condition contained 128 trials resulting in a total of 512 trials in Experiment 2. The participants' task was identical to Experiment 1. While the other-active block was an exact replication of Experiment 1, in the other-passive block the confederate was instructed not to respond to his category but instead just to watch the stimuli on the screen. Before each condition eight practice trials familiarized the participants with the task. These were later excluded from the analysis.

The order of congruency (congruent vs. incongruent), the order of the blocks (other-active vs. other-passive first), the order of the starting sub-list (animals-1 vs. animals-2 vs. fruit/vegetables-1 vs. fruit/vegetables-2) as well as the assigned category (animals vs. fruit and vegetables) was counterbalanced across participants.

Data Analysis

Errors (i.e. missed button presses during participants' own trials or button presses during the confederate's trials) and RTs more than two standard deviations from each participant's condition means were excluded from the RT analysis. Both the two condition means for correct response RTs for each participant as well as their errors were subjected to separate two-way, repeated-measures ANOVAs with the factors Congruency (congruent vs. incongruent) and Activity Other (other-active vs. other-passive).

Results

0.31% of the trials were removed as errors and 4.34% were removed as outliers, leaving 95.35% of the raw data as correct response trials. The removal of these outliers did not result in changes of the significance patterns observed in this experiment. The error analysis did not reveal any statistically significant results for Congruency, $F(1, 31) < 1$, $p > .250$, $\eta_p^2 < .01$, Activity Other, $F(1, 31) < 1$, $p > .250$, $\eta_p^2 = .01$, or an interaction, $F(1, 31) < 1$, $p > .250$, $\eta_p^2 < .01$.

The RT analysis revealed a significant main effect of Congruency, $F(1, 31) = 40.68$, $p < .001$, $\eta_p^2 = .57$, with RTs being generally faster during the congruent than during the incongruent condition (see Figure 14). This effect was moderated by a significant interaction between Congruency and Activity Other, $F(1, 31) = 5.56$, $p = .025$, $\eta_p^2 = .15$. The difference score between congruent and incongruent trials was significantly higher in the other-active ($M = 42.42$, $SE = 7.43$), compared to the other-passive condition ($M = 20.85$, $SE = 5.99$), $t(31) = 2.36$, $p = .025$, two-tailed). In addition, post-hoc *t*-tests revealed a significant congruency effect between congruent and incongruent trials both in the other-active, $t(31) = 5.71$, $p < .001$, as well as the other-passive condition, $t(31) = 3.48$, $p = .002$, two-tailed. Finally, there was no main effect of Activity Other, $F(1, 31) < 1$, $p > .250$, $\eta_p^2 = .01$.

These results replicate Experiment 1 and indicate that participants were sensitive to the fact that in the absence of any overt responses, the confederate could still read the stimuli. While this demonstrates that VSP-taking in mental space can occur in the absence of direct evidence of another's engagement, we also found that the effect of VSP-taking was larger when the confederate actively performed a task that required reading the words. A related question raised by the results of Experiment 2 is whether the congruency effect observed in the passive condition might simply be due to the bodily orientation of the observer rather than to him reading the words.

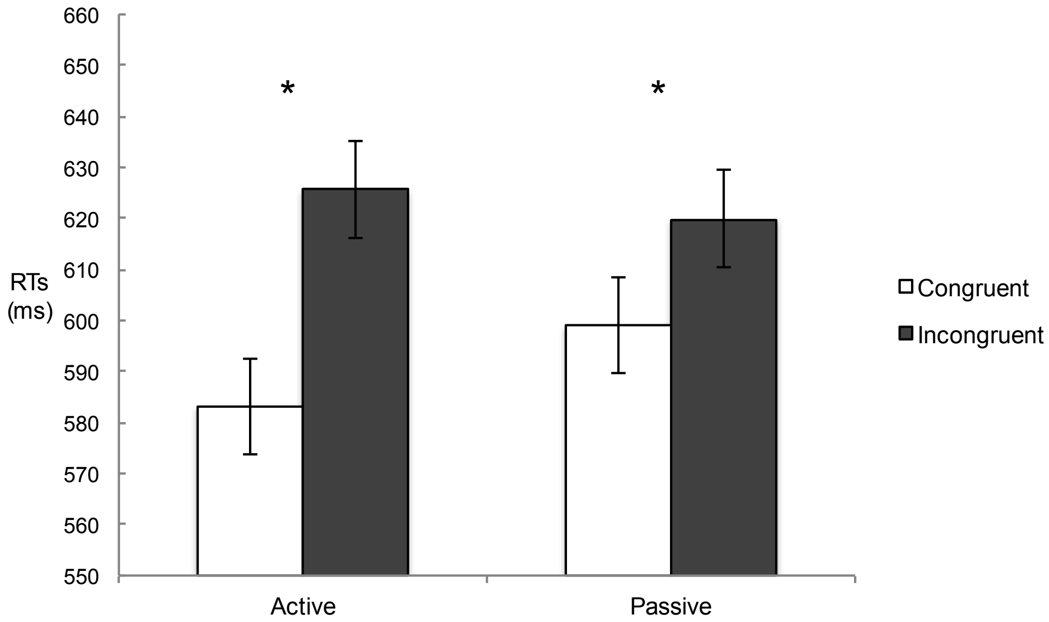

Figure 14. Mean RTs in the other-active and in the other-passive block in Experiment 2 (Ch.4). Error bars display within-subject confidence intervals according to Loftus & Masson (1994).

4.3 Experiment 3

Another agent's front features (such as his forehead, eyes, nose etc.) towards objects may automatically trigger a shift of attention regardless of whether that agent has visual access to the stimuli or not (cf. Heyes, 2014). Alternatively, the other's visual access may be crucial for spontaneous VSP-taking to occur. To address this question, we manipulated the confederate's ability to see the stimuli in Experiment 3. If participants' responses systematically changed in accordance with the confederate seeing or not seeing the stimuli, this would support the claim that ascribing visual access to another agent is a pre-condition for adopting their VSP and processing the stimuli as if seen from their point of view.

Method

Participants

32 participants (mean age = 21.94 years, 20 women, 29 right-handed) signed up for this experiment and received gift vouchers for their participation. All were naïve to the purpose of the study, reported normal or corrected to normal vision and signed informed consent prior to the experiment. All 32 participants met the inclusion criterion of having more than 90% successful trials in each experimental condition.

Stimuli and apparatus

The stimuli were identical to Experiment 1 and 2. The only difference in the apparatus was that the confederate wore a pair of lift-front goggles (Lux Optical, Worldwide Euro Protection, Luxembourg) throughout the experiment. These goggles had small shutters that could either be lifted up (in which case one had unhindered vision through transparent plexiglass), or flapped down (in which case black tape on the shutters blocked vision).

Procedure

Participants performed the two conditions (congruent vs. incongruent) in two different blocks (blindfolded vs. seeing). Each condition contained 128 trials, resulting in a total of 512 trials, as in Experiment 2. The participants' task was identical to Experiment 1 and 2. While the other-seeing block replicated the other-passive block in Experiment 2 (with the only exception that the confederate wore the transparent goggles), in the other-blindfolded block the confederate was instructed to flap the shutters of his goggles down and wait until the end of the block. Before each condition eight practice trials familiarized the participants with the task. These were later excluded from the analysis.

The order of congruency (congruent vs. incongruent), the order of the blocks (blindfolded vs. seeing first), the order of the starting sub-list (animals-1 vs. animals-2 vs. fruit and vegetables-1 vs. fruit and vegetables-2) as well as the assigned category (animals vs. fruit and vegetables) was counterbalanced across participants.

Data Analysis

Errors (i.e. missed button presses during participants' own trials or button presses during the confederate's trials) and RTs more than two standard deviations from each participant's condition means were excluded from the RT analysis. Both the two condition means for correct response RTs for each participant as well as their errors were subjected to two-way, repeated-measures ANOVAs with the factors Congruency (congruent vs. incongruent) and Vision Other (other-seeing vs. other-blindfolded).

Results

0.56% of the trials were removed as errors and 4.65% were removed as outliers, leaving 94.79% of the raw data as correct response trials. The removal of these outliers did not result in changes of the significance patterns observed in this experiment. The error analysis revealed a significant main effect of Vision Other, $F(1, 31) = 5.18$, $p = .030$, $\eta_p^2 = .14$, showing that participants made more errors when the confederate was blindfolded ($M = 1.25$, $SE = .21$) than when he had visual access to the stimuli ($M = .83$, $SE = .16$). Neither the main effect of Congruency, $F(1, 31) < 1$, $p > .250$, $\eta_p^2 = .03$, nor the interaction between the two factors, $F(1, 31) = 1.56$, $p = .221$, $\eta_p^2 = .05$, was significant.

The RT analysis revealed a significant interaction between Congruency and Vision Other, $F(1, 31) = 5.88$, $p = .021$, $\eta_p^2 = .16$. In post-hoc analyses, pairwise comparisons showed a significant difference in RTs between the congruent (M = 622.68, *SD* = 68.81) and the incongruent (M = 640.18, *SD* = 73.02) blocks only in the other-seeing condition, $t(31) = 2.22$, $p = .034$, two-tailed, but not in the other-blindfolded condition, $t(31) < 1$, $p > .250$, two-tailed (see Figure 15). The other pairwise comparisons did not reach significance (all *p*s > .218). Furthermore, there was neither a significant main effect of Congruency, $F(1, 31) = 1.72$, $p = .200$, $\eta_p^2 = .05$, nor of Vision Other $F(1, 31) < 1$, $p > .250$, $\eta_p^2 < .01$.

This indicates that the confederate's ability to see the stimuli was necessary for participants to adopt his VSP while cues about the orientation of the confederate relative to the stimuli (i.e., the direction of his body, forehead, nose etc.) were not sufficient for triggering spontaneous VSP-taking.

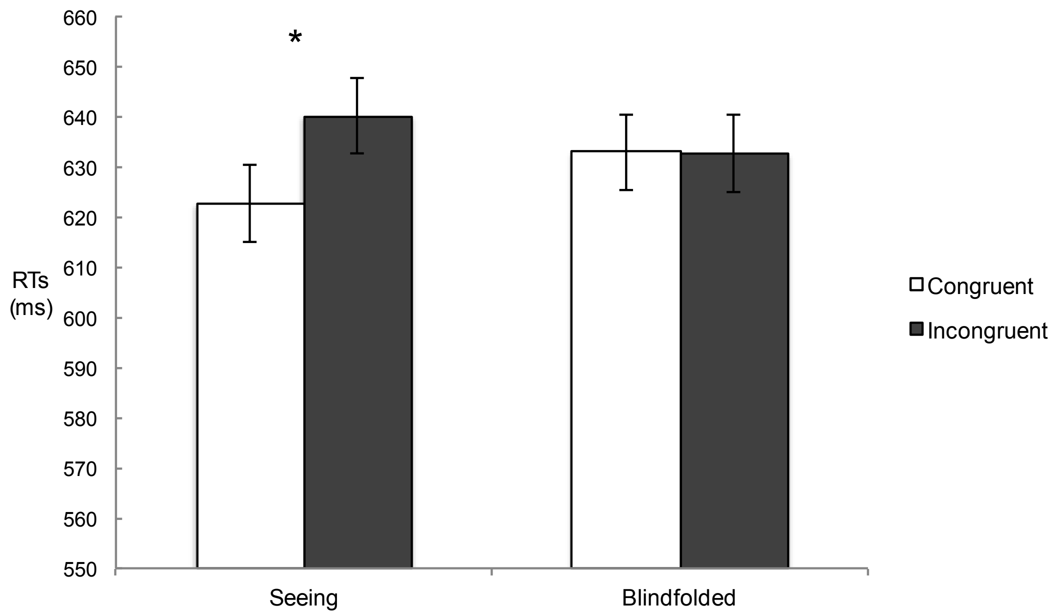

Figure 15. Mean RTs in the other-seeing and in the other blindfolded block in Experiment 3 (Ch.4). Error bars display within-subject confidence intervals according to Loftus & Masson (1994).

4.4 General Discussion

The aim of this study was to investigate whether spontaneous VSP-taking occurs in mental space, where spatial relations matter for mental rather than physical actions. Across three experiments, we found that participants reliably adopted the VSP of a confederate in the context of a semantic categorization task that involved reading words. Specifically, we found that participants were faster to categorize words that were orientated in an upright way from the point of view of a confederate, compared to words orientated upside down from the confederate's point of view. This indicates that VSP-taking is not restricted to situations involving physical interactions with objects or action planning (cf. Creem-Regehr, Gagnon, Geuss, & Stefanucci, 2013; Freundlieb et al., 2016; 2017) as it also extends to the mental space.

Furthermore, our results provide evidence for *spontaneous* VSP-taking in mental space. The orientation of the words and the sitting position of the confederate were completely irrelevant for the participants' task, and participants were never prompted to adopt the other's perspective. This extends earlier studies on VSP-taking in physical space where participants were asked to provide responses about spatial arrangements of objects from a particular perspective (Furlanetto, Becchio, Samson, & Apperly, 2016; Cavallo et al., 2016).

A further important finding of our study is that participants spontaneously adopted somebody else's VSP even if that other person was not explicitly assigned a task and was just passively observing the stimuli (Experiment 2). Previous studies have shown evidence for spontaneous VSP-taking only if the task was performed within a cooperative context (Surtees et al., 2016) or if the responses given by the other person indicated her constant engagement in the shared task (Freundlieb et al., 2016). We believe that this discrepancy between earlier results of VSP-taking in physical space and the present results can be explained by differences in the tasks involved. For example, as the rules of the task used by Freundlieb and colleagues (2016) were completely arbitrary, it was only through the confederate's overt responses that participants could verify the confederate's participation in the task. In contrast, participants in the current task could still assume that the passive confederate processed the stimuli in a meaningful way – as written words automatically trigger reading and recognition processes (especially given that the confederate was instructed to watch the stimuli, cf. Strijkers, Bertrand, & Graininger, 2015; Stroop, 1935). This 'passive' participation (likely involving reading) might have been sufficient for participants to perceive the task as being interactive (or as a "team context", see Surtees et al., 2016) and hence, to spontaneously adopt the confederate's VSP. Finally, the results of Experiment 2 suggest that although the activity of the other was not a necessary factor for the congruency effect to occur, it seemed to have further increased VSP-taking effects.

But what exactly led to the observed effects in the current study in the first place? It could be claimed that the congruency effects found in Experiment 1 and 2 were simply based on a domain-general mechanism picking up directional cues (such as somebody else's body orientation) and thereby redirecting participants' attention (cf. Heyes, 2014). Importantly, such a mechanism would elicit the same effect regardless of whether the agent exhibiting such directional cues had visual access to the stimuli or not. However, in Experiment 3 we replicated the results obtained in Experiment 2 and showed that spontaneous VSP-taking disappeared if the confederate's visual access to the stimuli was blocked. Importantly, this corroborates that participants were taking into account how the confederate *saw* the stimuli and rules out the possibility that directional cues about the confederate's front features (such as the orientation of his body, forehead, nose etc.) are sufficient for triggering spontaneous VSP-taking effects.

We think that the mechanism underlying the effects reported here involves a modulation of the processes involved in word reading that was prompted by the spontaneous adoption of the other person's VSP. Words are not processed and transformed as integral units over the entire range of orientations (Koriat & Norman, 1985). Instead, Koriat and Norman (1985) proposed that when stimuli are (close to) the upright canonical orientation (± 60°), word recognition relies on whole-word units, whereas at more extreme orientations (beyond 120° deviations) it appears to be based on sequential letter identification. They suggest that for intermediate orientations (60° - 120°, which coincides with the orientation used in our study) "word recognition may rely on units larger than single letters" (p. 507, Koriat and Norman, 1985). In our study, adopting the other person's VSP might have led participants to process the words more holistically in the congruent condition and in a letter-by-letter fashion in the incongruent condition, creating the observed differences in reaction times. Future experiments will have to determine and disentangle the effects of VSP-taking on semantic, orthographic, and lower visual levels of word processing.

4.5 Conclusions

Overall, the current study enriches our prior knowledge on spontaneous VSP-taking in three important ways. To our knowledge, these are the first results showing that people spontaneously adopt another's VSP in mental space. Thus, our findings suggest that spontaneous VSP-taking extends to situations in which spatial relations matter for mental rather than for physical actions. Secondly, we found evidence for spontaneous VSP-taking in mental space even in the absence of any overt physical response by the other. This shows that participants were sensitive to the fact that the other person automatically processed the words regardless of whether he physically responded to them. Finally, the results suggest that ascribing visual access to other agents is a pre-condition for adopting their VSP.

Chapter 5.
General Discussion
Cognitive Mechanisms underlying VSP-Taking

While all three studies empirically support the claim that humans spontaneously adopt another's visuospatial perspective (VSP) both in physical (Chapter 2 and 3) and in mental space (Chapter 4), some questions regarding the mechanisms underlying spontaneous VSP-taking still remain. Generally speaking, how can we describe perspective-taking at a cognitive level?

For instance, did VSP-taking generally *facilitate or interfere* with peoples' task performance? Furthermore, what happens exactly when we adopt somebody else's VSP – how do we navigate within the duality of our own (egocentric) and somebody else's (i.e. the altercentric) perspective? Put differently, which spatial *reference frame* do we use during instances of spontaneous VSP-taking? Finally, what are the *boundary conditions* leading to spontaneous VSP-taking and do we have evidence to claim that the perspective-taking effects we found were exclusively *triggered by intentional agents?*

This chapter will summarize and elaborate on the empirical evidence we presented in the previous chapters so that the cognitive mechanisms underlying spontaneous VSP-taking become clearer.

5.2 Facilitation and Interference

Did spontaneous VSP-taking manifest in *facilitation* or in *interference* effects? One can argue that an individual baseline would have been informative with regards to the question how VSP-taking affected task performance and processing speed throughout our studies. However, the findings of Chapter 2, Experiment 3 (where the confederate was active vs. passive) as well as from Chapter 3, Experiment 2 (where the confederate was blindfolded vs. seeing) can also shed light on this issue. Importantly, both of these results showed a significant interaction effect between Compatibility and Role of the Confederate (Chapter 2) and Vision Other (Chapter 3), respectively. This means that only when the confederate was actively involved in the spatial compatibility task (Chapter 2) and only if the confederate had visual access to the stimuli (Chapter 3) did participants take into account his VSP. In contrast, when the confederate was passive or blindfolded, there was no indication that participants adopted the confederate's VSP (i.e. no significant difference between compatible & incompatible trials). Crucially, this turned the 'Other Passive' (Chapter 2) and the 'Blindfolded' (Chapter 3) condition into baseline conditions from which we could compare the RTs in the three other conditions.

We then found partial evidence for facilitation in both of these experiments. More specifically, post-hoc *t*-tests revealed that the compatible-active (Chapter 2) and the compatible-seeing (Chapter 3) conditions were significantly faster than all three other conditions, respectively. Furthermore, in two other experiments (Chapter 3, Experiment 1 and Chapter 4, Experiment 3) numerically it also seemed as if the compatible-seeing (Chapter 3) and the congruent-seeing (Chapter 4) condition was faster than all three other conditions. However, these two patterns failed to reach significance.

Thus, while we did find some evidence for facilitation effects during VSP-taking in Chapter 2–4, further experiments would be required to make a more compelling argument about facilitation actually driving the effect. Importantly, the data points to the possibility that the adoption of somebody else's perspective can lead not only to interference effects (as found in other studies, see Samson et al., 2010), but that under specific circumstances (such as in the present experiments) it can actually facilitate task performance. In turn, this might assist with interpersonal coordination in situations where we need to quickly integrate the diverging perspectives of multiple agents, e.g. in ball games. Finally, as participants always performed the task with a partner throughout our studies, we could generally rule out social facilitation effects.

5.3 Reference Frames

Exactly whose perspective was adopted during instances of spontaneous VSP-taking in our studies and could it be that neither one's own, nor the other person's perspective was used as a reference frame, but actually an agent-independent point of view?

This section discusses what perspective-taking entails in terms of the plurality of reference frames that could have been active throughout the different perspective-taking tasks. In this respect, three hypotheses seem plausible:

1.) During spontaneous VSP-taking neither participants' own, nor the other person's VSP serves as the spatial reference frame. Instead a general, agent-independent perspective (a 'God's-eye-view', if you will) is computed that is anchored somewhere in the environment (the *God's Eye Hypothesis*).

2.) During spontaneous VSP-taking participants abandon their own VSP and completely adopt the other's VSP (the *Dissolution Hypothesis*).

3.) During spontaneous VSP-taking participants adopt the other person's VSP while retaining their own VSP (the *Multifocal Lens Hypothesis*).

In the following, I will elaborate on each of these three possibilities and discuss them in light of the data we acquired throughout Chapter 2 – 4.

Disembodied Perspective-Taking – The God's Eye Hypothesis

One way to sketch the processes underlying spontaneous VSP-taking is to assume that participants adopt an abstract point of view (also called *God's-eye view* or *allocentric* perspective) that it is not embodied through the agents that are present in the scene. Instead, such an allocentric perspective can be thought of as being linked to an agent-independent reference frame in the external environment, for example, in a prominent landmark (cf. Klatzky 1998; Soechting & Flanders, 1992; Volcic & Kappers 2008).

With regards to the three studies in Chapter 2 – 4, adopting an allocentric perspective would entail that participants encoded the stimuli neither from their own / egocentric, nor from the confederate's / *altercentric* perspective but instead from above / a third person point of view. But where exactly would such an allocentric point of view be anchored in or, in other words, which landmark would become the spatial reference point?

Importantly, the data we found in our studies unequivocally rejects the possibility of an allocentric reference frame having been used throughout the different tasks. Specifically, both in Chapter 2 and 3, as well as in Chapter 4 an allocentric perspective cannot explain the obtained pattern of results, as the stimulus-response-compatibility effect (Chapter 2 and 3) and the congruency effect (Chapter 4) exclusively correspond to the VSP of the confederate.

Thus, while the concept of a disembodied allocentric perspective seems interesting in theory, it is unclear how an allocentric viewpoint could operate in the context of environments that are not 'affording' the use of an allocentric strategy. More specifically, the selection of an agent-independent viewpoint might work in environments where the prominence of one specific external landmark is so salient that it clearly stands out for all agents. But it is not clear which external landmark could be used in order to spatially anchor the stimuli independently of the two agents' viewpoints and *in the absence of a prominent landmark*.

Hence, with regards to our studies, we might infer that the adopted perspective must have somehow been anchored in the reference frame of the other agent. The question then is: to what degree did perspective-taking rely on the other person's reference frame: completely or only partially?

Perspective-taking as becoming the Other – the Dissolution Hypothesis

Probably the most radical among the three hypotheses is the following: while adopting another person's VSP, participants completely abandon their own (egocentric) perspective and only encode the stimuli as seen from the visuospatial perspective of the other agent (i.e. altercentrically). While this hypothesis sounds rather drastic at first, it does not necessarily manifest in a dramatic manner if placed in the context of our studies.

Specifically, applied to the experimental paradigms in Chapter 2 and 3 (the two Spatial Compatibility paradigms), the Dissolution Hypothesis would assert that while performing the spatial compatibility paradigm, participants encoded the stimuli not as 'up and down' (according to their own VSP) but as 'left and right' (as seen from the confederate's VSP).

Hence, what 'dissolves' according to the Dissolution Hypothesis is not the bodies or entire minds of the actors but instead "just" how objects are visuospatially encoded (i.e. altercentrically instead of egocentrically) while two agents simultaneously attend to them. But regardless of whether the hypothesis seems far-reaching or not - can we find empirical evidence in our studies to support or reject this notion?

One indication that might actually speak against the Dissolution hypothesis is that we did not find clear evidence neither for strong facilitation nor for strong interference effects during participants' task performance. Yet, if participants completely abandoned their egocentric perspective and adopted the altercentric perspective throughout the task, their response patterns should show (in line with the response patterns one would expect from the VSP of the confederate) strong facilitation effects during the compatible and strong interference during the incompatible condition.

Remember that the idea behind our experimental set-up in Chapter 2 and 3 was that technically, there was only an overlap between the stimulus- and the response-dimension from the confederate's but not from the participant's point of view. More precisely, from the confederate's point of view, stimuli appeared as 'left and right' to which he responded with 'left and right' button presses whereas from the participant's point of view, stimuli appeared as 'up and down' to which she responded with 'left and right' button presses.

One could therefore classify the participant's task as entailing a *Type 1 Ensemble,* which is characterized by the absence of a dimensional overlap in either the relevant or the irrelevant dimension (cf. Kornblum et al. 1990). According to Kornblum (1990) such an ensemble "(…) may be useful as neutral or control conditions" (Kornblum et al., 1990; p. 264). In contrast, the confederate's task could be classified as incorporating a *Type 2 Ensemble,* where there is a dimensional overlap between the stimulus and the response dimension, therefore satisfying "(…) the requirements for obtaining mapping effects" (Kornblum et al., 1990; p. 264).

Thus, with regards to the response patterns participants *should* elicit - assuming they exclusively relied on the confederate's VSP - one would expect such mapping effects to occur, that is significantly faster RTs during the compatible and/or significantly slower RTs during incompatible trials. Put differently, one would expect to find the same response patterns the confederate would typically elicit.[16]

[16] Note that for technical reasons we could unfortunately not record the confederate's RTs. Yet, given that stimulus-response-compatibility effects have been reliably replicated for over 60 years now, it seems reasonable to

However, looking at the participants' response patterns we only find partial evidence for facilitation effects during compatible trials (see Chapter 5.1. for a more detailed analysis) and further experiments would be required to make a more compelling argument for facilitation actually driving VSP-taking effects. In consequence, as the response patterns of the three studies in Chapter 2, 3 and 4 do not show clear facilitation and / or interference effects, one must therefore conclude that the averaged RTs cannot provide strong support for the Dissolution Hypothesis. Yet, a closer look on the *time course* of the perspective-taking effects (see Chapter 5.2.4) will later on suggest that we cannot refute the Dissolution Hypothesis either.

Lastly, although radical in its implication (e.g. abandoning one's own perspective means that we disembark from our default viewpoint), the Dissolution Hypothesis would not be the only instance in the social cognition literature where humans temporarily dissolve into another person's perspective. For example, in the realm of affective perspective-taking, or empathy, there is the well-known phenomenon of *emotional contagion* that is illustrative with regards to the Dissolution Hypothesis.

Hatfield, Cacioppo, and Rapson (1993) define emotional contagion as "the tendency to automatically mimic and synchronize facial expressions, vocalizations, postures, and movements with those of another person and, consequently, to converge emotionally" (Hatfield, Cacioppo, & Rapson, 1993 p. 96). While often described as an important building block to emotionally relate to others (Decety & Ickes, 2009), emotional contagion on its own does not suffice to invoke what is known as the phenomenon of empathy. Instead, the concept of empathy has been defined as a construct to account for a sense of similarity in feelings experienced by the self and the other *without confusion between the two individuals* (Decety & Lamm, 2009; Decety & Jackson, 2004; Decety & Lamm, 2006). Hence, one also needs the

assume that the confederate's responses were much faster and less prone to errors during the compatible compared to the incompatible condition, cf. Fitts & Deininger, 1954; Fitts & Seeger, 1953; Simon et al., 1970; Simon, 1990.

knowledge that that self and other are similar but separate (i.e. *self-other awareness*). Thus, it is only in the interplay of emotional contagion on one side, and self-other awareness on the other side, that genuine empathic responses can be invoked (cf. Decety and Lamm, 2009).

Thus, with regards to the discussion about the Dissolution Hypothesis, it might be that just like for proper empathic responses one needs to be able to combine (emotional) contagion and self-other distinction, the successful adoption of somebody else's VSP involves a hybrid strategy in which agents utilize both their own (egocentric) and the other person's (altercentric) perspective. In the next section we will discuss what such a hybrid strategy could look like and what questions it would raise.

Switching between two Perspectives – the Multifocal Lens Hypothesis

The third hypothesis about how people adopt others' VSP is to think of the underlying process as wearing glasses with multifocal lenses. These special type of glasses are characterized by a gradient increase of lens power that allows the wearer to focus on both close-by objects (say, by looking through the lower section of the glasses) as well as on far away objects (say, by looking through the upper part of the glasses). Applying this metaphor onto perspective-taking, the Multifocal Lens Hypothesis claims that participants were sensitive to and took into account both their own (egocentric) perspective and the other person's (altercentric) perspective. In the following I will discuss this hypothesis further and see whether it can hold in light of the data we gathered in Chapter 2, 3 and 4.

If we assumed that humans are indeed able to retain two VSPs at the same time (or consecutively), an interesting question arises: how would these two distinct VSPs be organized in our cognitive architecture? In other words, if we were able to process two VSPs over a given period, does the underlying mechanism operate by shifting back and forth between the ego- and the altercentric VSP? Alternatively, could we be able to retain the two perspectives *at the same time*, that is process two discrete VSPs in parallel? Finally, regardless of whether such a 'perspective duality' works in parallel or in succession – what determines which of the two perspectives becomes the dominant one at any given point in time? That is, if two active VSPs are successively accessible, when and how do we shift between them? Similarly, if we processed two VSPs in parallel, are the two inputs equally weighted or does one dominate over the other?

Applied to Chapter 2, the Multifocal Lens Hypothesis would assert that participants encoded the stimuli both as 'up and down' (according to their own VSP) *and* as 'left and right' (according to the other person's VSP). What evidence can we find to support the hypothesis that participants took into account multiple spatial reference frames during VSP-taking?

Perhaps the only indication we can gather to claim that participants must have been processing the stimuli within more than one single reference frame during perspective-taking comes from Chapter 2, Experiment 5 (the Spatial Compatibility Paradigm). As part of the spatial compatibility paradigm, in Experiment 5 we wanted to investigate how to further capture and differentiate the plurality of perspective dimensions that were involved in the task. To reiterate, we differentiated between the *stimulus dimension* (actually capturing the two differing VSPs of the participant and the confederate) and the *response dimension* that captured the spatial arrangement of the two agents' responses (that is, how their respective response buttons were spatially aligned).

Up until Experiment 5 in Chapter 2 the response dimension was always horizontal, i.e. for both the participant and the confederate response buttons were always oriented on a 'left-right' dimension (see Figure 1, Chapter 2). However, this way we could not disentangle whether adopting the confederate's perspective meant that participants switched to the confederate's reference frame in terms of both the stimulus *and* the response dimension or whether they only adopted the confederate's VSP and disregarded the spatial dimension of the confederate's responses. And indeed, by varying the confederate's response dimension in Experiment 5 (such that he would now respond 'up and down' rather than 'left and right') we could show that participants adopted the confederate's stimulus dimension (or VSP) but retained the spatial dimension of their own responses ('left and right'). This interplay between the confederate's VSP and the participant's response dimension then lead to a dimensional overlap and created the stimulus response compatibility effect which we found (see Chapter 2, Experiment 5 for further details).

While this experiment shows that participants were sensitive to *at least* two spatial perspectives originating from two different actors (i.e. the stimulus dimension of the confederate and the response dimension of the participant), it does not reveal whether participants were, in addition to that, also sensitive to their own VSP. Assuming they were sensitive to their own VSP while adopting the confederate's VSP, Chapter 2 also leaves the question open how those two VSPs of the participants and the confederate could be hierarchically organized – whether they would be processed successively or in parallel and what would determine their relative weights.

Taken together, it seems as if based on the averaged RTs in our data sets alone, one cannot make strong claims for either the Dissolution Hypothesis or for the Multifocal Lens Hypothesis. In the following, I want to discuss how tentative evidence (supporting the Dissolution Hypothesis) can be constructed though by looking at the *time course of perspective-taking effects*. Specifically, I want to investigate whether specific patterns in our data sets allow us to make inferences regarding the time span during which participants adopted the other's VSP in our experiments.

The Time Course of Perspective-Taking Effects

In this section two pieces of evidence from our studies are discussed regarding the question of the time course of perspective-taking. More precisely, I will analyze temporal aspects of the perspective-taking effects we found throughout the three studies to answer the following question: *for how long might participants have adopted the confederate's VSP during the different tasks?*

Answering this question is illustrative for the discussion about the two competing hypotheses (i.e. the Dissolution vs. the Multifocal Lens Hypothesis) as it will help to clarify whether participants adopted the confederate's VSP *uninterruptedly* (which, in turn, would support the Dissolution Hypothesis), or whether participants adopted the confederate's VSP *intermittently* (which would speak against the Dissolution and for the Multifocal Lens Hypothesis).

The first piece of evidence comes from Chapter 4 and is based on an analysis of within-subjects data (and more precisely: block-by-block analyses of perspective-taking effects). The second piece of evidence comes from Chapter 2 and considers between-subjects analyses (more concretely: order-effects between counterbalancing groups). I will start with the evidence from Chapter 4.

In order to explain the congruency effect we found in Chapter 4, we argued that the mechanism underlying the effect involved a modulation of the processes involved in word reading that was prompted by the spontaneous adoption of the other's VSP (see Chapter 4). More specifically, we hypothesized that participants processed the words more holistically in the congruent and in a letter-by-letter fashion in the incongruent condition. In turn, this could have led to the observed differences in reaction times. There are at least two possibilities with regards to the length of the periods in which participants adopted the other's perspective, or *the volatility of perspective-taking* if you will.

The first possibility is that participants *uninterruptedly adopted* the confederate's VSP during both conditions, meaning that they adopted the confederate's perspective in the beginning of the experiment and only switched back to their own perspective once the experiment was over. Together with the assumption that adopting the other person's perspective means that participants abandoned their own perspective, this would mark a particular strong version of, and support the Dissolution Hypothesis.

The second possibility is that participants *intermittently adopted* the confederate's VSP – e.g. only during the congruent condition (when the holistic word processing actually gave them an advantage over encoding the words in a letter-by-letter fashion). This would mean that participants engaged in perspective-taking during half of the trials (the congruent trials) and retained their own VSP during the other half of the Experiment (during the incongruent condition). Similarly, in a more extreme version, 'intermittently' could also mean that participants continuously switched back and forth between the ego- and the altercentric perspective – as long as, on average, they remained in the altercentric perspective more often than in the egocentric perspective. The main point I am trying to raise here is that, rather than occurring continuously, perspective-taking might as well have taken place in a more volatile and erratic manner (which would speak against the Dissolution Hypothesis – and for the Multifocal Lens Hypothesis).

Importantly, the general outcome of the perspective-taking effects (in terms of averaged RTs) would be identical in both of the cases that I have just sketched: regardless of whether participants adopted the other's VSP uninterruptedly or intermittently, one would expect participants to be faster processing the stimuli in the congruent vs. in the incongruent condition – which is exactly what the pattern of results of Chapter 4 shows. Hence, based on the averaged RT patterns alone one could not make a good argument for perspective-taking occurring uninterruptedly vs. intermittently – and therefore neither for or against any of the two competing hypotheses.

However, if one argued that participants adopted the confederate's VSP only intermittently, then one would expect the congruency effect to be generally more volatile and erratic compared to if participants adopted the confederate' VSP uninterruptedly. For instance, it could be that initially, participants did not adopt the VSP of the other person but realized later on that it might be useful. In this case, one should find no evidence for spontaneous VSP-taking in the beginning of the experiment. Alternatively, it could also be that participants adopted the other's VSP mostly in the first part of the experiment and switched back to their own VSP towards the end. Or that they adopted the other person's VSP in one block but not in the other.

Thus, looking at the time course of the congruency effect in Chapter 4 might be illustrative with regards to the question of the *volatility of perspective-taking* – and thereby also with regards to the two competing reference frame hypotheses (the Dissolution Hypothesis vs. the Multifocal Lens Hypothesis).

Exploring the time course of how the effect developed during Experiment 1 of Chapter 4 *using within-subjects data*, we plotted the RTs after each of the four blocks (remember that the stimuli consisted of sublists [e.g. animals 1 & fruit & vegetables 1] that were repeated four times in the first condition [say, the congruent condition] before the complementary lists [in this case, animals 2 & fruit & vegetables 2] would be shown in a random order in the other condition for four times [e.g. the incongruent condition]). This way, we could get a grasp of how the effect developed across the experiment and, more importantly for this discussion, we might see whether the data patterns indicate that perspective-taking seemed volatile (erratic), or stable throughout the different blocks of the experiment.

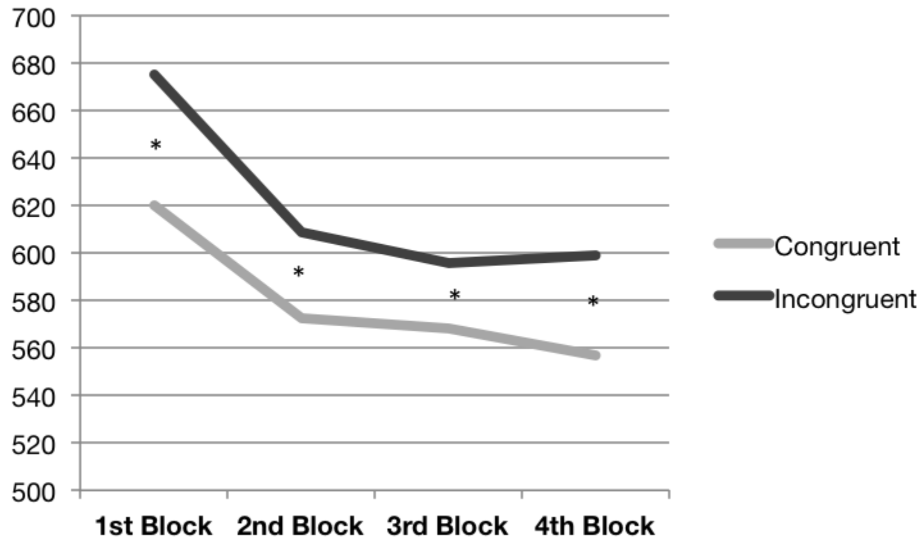

Figure 16. Time course of the congruency effect in Experiment 1 (Ch.4).
The ordinate shows RTs in ms. Asterisks depict significant differences in RTs according to paired-samples *t*-tests (two-tailed).[17]

Figure 16 clearly shows that the congruency effect was in fact quite stable across the four blocks. Although RTs became faster after the first block (due to the participants recognizing the word items that were repeated from the second block on), the differences between the congruent and the incongruent condition remained significant until the last block. This suggests that the adoption of the other person's VSP was likely not due to an intermittent process that varied across blocks[18] rather than a steady one. In turn, this fits better with the Dissolution Hypothesis asserting that participants were only using one single reference frame (namely the confederate's), rather than two at the same time.

Looking at the second piece of evidence a similar picture arises. That is, a closer look at the time course information we have available from Chapter 2 also suggests that partici-

[17] 1st Block: $t(15) = 4.9$, $p < .001$, two-tailed; 2nd Block: $t(15) = 3.3$, $p = .004$, two-tailed; 3rd Block: $t(15) = 2.3$, $p = .03$, two-tailed; 4th Block: $t(15) = 4.2$, $p = .001$, two-tailed;
[18] I am acknowledging that I cannot rule out that VSP-taking still varied on a trial by trial basis.

pants adopted the confederate's VSP uninterruptedly, and therefore provides converging evidence in favor of the Dissolution Hypothesis. Importantly, while the previous time course analysis revolved around within-participant data, we will now look at between-subjects data from Chapter 2 and, more precisely, at order effects.

The presence or absence of order effects are illustrative with respect to the discussion about the two competing hypotheses because they are indicative of the underlying perspective-taking process. Specifically, if we found order effects then this would show that participants adopted the confederate's VSP to a varying degree – *depending on the order of conditions* (i.e. their specific counterbalancing group). For instance, order effects might reveal that one participant group shows more pronounced evidence for perspective-taking while another group shows less.

Importantly, the presence of such order effects could therefore demonstrate whether - at a group level - perspective-taking likely took place intermittently (i.e. some perspective switches happened), which would support the Multifocal Lens Hypothesis. In contrast, the absence of order effects would indicate that perspective-taking occurred uniformly or uninterruptedly, which would support the Dissolution Hypothesis.

One might claim that the particular set-up we used throughout the Spatial Compatibility paradigm actually afforded order-effects to take place. More precisely, it could be argued that subjects who conducted the compatible condition first and the incompatible condition second should elicit larger compatibility effects than those who fell into the other counterbalancing group.

For the former group adopting the other's VSP meant a performance advantage in the first condition (the compatible condition) and subjects might have therefore unintentionally engaged in VSP-taking also in the following condition (i.e. the incompatible condition). This would have led to comparatively larger compatibility effects in the compatible-first group. Vice versa, for the incompatible-first group, adopting the other's VSP posed a disadvantage at first, which means that subjects could have potentially adopted an unintentional strategy of ignoring the VSP of the confederate in the successive condition. This could have led to comparatively smaller compatibility effects in the incompatible-first group.

Importantly, if it was indeed the case that (unintentionally) differing strategies were employed depending on which condition the participant started with, then we should have found clear indications for order effects in the general response patterns. In turn, these order effects would indicate that VSP-taking can be (strategically) employed to a varying degree, which would mean that the underlying processes are volatile (or subject to manipulation) rather than steady (and non-manipulable). However, we did not find evidence for order effects in any of the five experiments in Chapter 2.

The lack of order effects in Chapter 2 suggests that spontaneous VSP-taking occurred rather stably and uniformly between the counterbalancing groups. This means that the participants of Chapter 2 probably adopted the other person's VSP in a rather continuous manner throughout the experiment and that they did not engage in perspective-taking to a higher or lesser degree depending on the specific order of conditions they were in. Along with the within-subjects time course analysis of Chapter 4, this provides converging evidence to claim that perspective-taking did likely not occur *intermittently*.

In sum, one could thus say that while these rudimentary time course analyses cannot proof the Dissolution Hypothesis to be true (as other data patterns actually speak against it, see discussion on the Dissolution Hypothesis), they do not provide compelling evidence against it either. Together with the preceding discussion on the two competing reference frame hypotheses, it therefore seems as if we can only speculate about the question whether participants in our studies abandoned their egocentric perspective completely or only partially.

Summarizing, I think that clear(er) patterns for facilitation and / or interference during VSP-taking would have served as the most convincing argument with regards to the question whether participants adopted the confederate's VSP *uninterruptedly,* or only *intermittently* (and thereby also regarding the two competing reference frame hypotheses).

For instance, if participants in Chapter 2–4 would have been faster responding to the congruent stimuli, but also slower responding to the incongruent stimuli, this would have clearly demonstrated that perspective-taking was driven both by facilitation as well as by interference effects. Importantly though, it would also indicated that participants actually abandoned their egocentric perspective and instead adopted the confederate's perspective throughout the entire task. Hence, this would have posed as strong evidence for the Dissolution Hypothesis.

In contrast, if we would have found only facilitation effects but no signs of interference, then this would have supported the claim that at least during the incongruent condition participants retained their own egocentric perspective and only gathered a speed-advantage in the congruent condition. In turn, this would have supported the notion that during perspective-taking, participants switch back and forth (at least once) between the ego-, and the altercentric perspective, just as the Multifocal Lens Hypothesis claims.

5.4 Boundary Conditions

In order to better capture the cognitive mechanisms underlying spontaneous visuospatial perspective-taking, the previous sections discussed whose reference frame participants might have used while engaging in perspective-taking, and whether this led to increased task performance. Another important aspect for a better understanding of the cognitive underpinnings of VSP-taking revolves around the question how spontaneous VSP-taking is initiated, or *triggered* in the first place.

This section therefore summarizes the specific boundary conditions that led to spontaneous VSP-taking in Chapter 2 – 4. As the boundary conditions differed between the experimental paradigms, I will start discussing the boundary conditions for VSP-taking in physical space (i.e. Chapter 2 and 3) before I will then turn to the evidence we gathered in mental space (Chapter 4).

Summarizing the specific conditions that led participants to adopt the other person's VSP in the Spatial Compatibility paradigm, it seems that knowledge about the other person's task was sufficient, while receiving visual and auditory feedback of the other's response was not necessary to trigger VSP-taking (see Chapter 2, Experiment 2). Furthermore, VSP-taking in physical space crucially relied on the other person being perceived as an intentional co-actor, as the mere presence of a passive confederate proved not to be sufficient for triggering VSP-taking (Chapter 2, Experiment 3). Finally, adding to the findings from Chapter 2, the results from Chapter 3 show that knowledge about the other person's visual access systematically modulated spontaneous VSP-taking. More specifically, Chapter 3 shows that participants only adopted the other's VSP if the other had unhindered visual access to the stimuli but regardless of whether the other person performed the same or a different task (see Chapter 3, Experiment 1 and 2).

Thus, judging from the findings of Chapter 2 and 3, both the modulation of the other person's activity (Chapter 2, Experiment 2), as well as modulating the other person's visual access (Chapter 3) created effected boundary conditions determining whether VSP-taking was triggered or not. First of all, these findings emphasize the importance of interacting in a space where attention can be actively shared, i.e. where the other person can be perceived as an intentional actor selectively attending to the same object (cf. Chapter 1.3 on Joint Attention). Second, these results also neatly reflect humans' increased sensitivity during instances of joint task performance (cf. Chapter 1.4 on Task Co-Representation).

Chapter 4 (VSP-taking in Mental Space) shows that participants spontaneously adopted another's VSP in the context of a semantic categorization task. Specifically, we found evidence for spontaneous VSP-taking in mental space even in the absence of any overt physical response by the other (Chapter 4, Experiment 2), but only if the other had unhindered visual access to the stimuli (Chapter 4, Experiment 3). Hence, similar to our findings in Chapter 3, it seems that knowledge about the other person's visual access is crucial for triggering spontaneous VSP-taking in mental space. This finding directly addresses the debate on alternative (low-level) explanations for perspective-taking effects (cf. Chapter 1.5 on the Social Nature of Visual Perspective-Taking) and clearly rules out that it was the directionality of the other person's front features that had triggered VSP-taking.

Though in contrast to Chapter 2 (VSP-taking in shared Physical Space), we found that participants in Chapter 4 still adopted the confederate's VSP even if the confederate was just passively observing the screen. On a first look, this seems to indicate that different perspective-taking mechanisms could be at work in physical and in mental space, respectively. However, I would argue that the underlying perspective-taking mechanism was the same and that the differences across the two paradigms were task-specific and simply demonstrated that participants differentially mentalized during the two tasks.

More specifically, we argue in Chapter 4 that the different outcomes of the passive-other modulations in Chapter 2 and Chapter 4 can actually be explained by differences of the tasks that were involved, rather than being based on two different mechanisms. More precisely, as the rules of the stimulus-response-compatibility task in Chapter 2 were completely arbitrary (e.g. 'push the left button when you see a stimulus appearing at the upper end of the screen and push the right button when you see a stimulus at the lower end of the screen'), it was only through the confederate's overt responses that participants could verify the confederate's participation in the task. In contrast, participants in Chapter 4 could still assume that the passive confederate processed the stimuli in a meaningful way – as written words automatically trigger reading and recognition processes. The fact that we found evidence for spontaneous VSP-taking in Chapter 4 therefore showed that participants took into account that the other person automatically processed the words regardless of whether he physically responded to them or not.

Thus, instead of showing that VSP-taking in mental vs. in physical space is based on two different mechanisms, the differing outcomes between the two passive-other manipulations (i.e. Chapter 2, Experiment 2 and Chapter 4, Experiment 2) might actually indicate that different levels of mentalizing were involved in the two tasks (see next section for a more detailed discussion on mentalizing).

5.5 The Social Nature of VSP-Taking

An ongoing debate in the literature on perspective-taking revolves around the question whether effects that are usually connected to the phenomenon of perspective-taking are actually "social" in their nature – meaning that they are exclusively triggered in the presence of another intentional agent – or whether they originate in low-level mechanisms that are not specific to social encounters or interactions. To re-iterate this issue, some authors have claimed that the effects that had previously been argued to demonstrate fast and efficient reasoning about another person's *mental state*, are actually due to low-level attentional mechanisms that do not differentiate between intentional agents and non-agentive directional cues (Santiesteban et al., 2014; Heyes, 2014). Particularly, another agent's front features (such as his forehead, eyes, nose etc.) towards objects in a scene may automatically trigger a shift of attention highlighting the agent's perspective regardless of whether that agent has visual access to the stimuli or not (cf. Heyes, 2014). Furthermore, the same mechanism should be at work even if the other is not an intentional agent but a symbol which effectively captures one's attention, like an arrow (cf. Santiesteban et al., 2014). Taking into account the data we acquired from Chapter 3 and 4 we can make a contribution with regards to this debate.

In each of the two paradigms (the Spatial Compatibility and the Reading paradigm) we dedicated one experimental manipulation exactly to test whether the effects we find can actually be explained by lower level mechanism such as cueing effects. More specifically, we tested whether participants took into account the visual access of the confederate by comparing a blindfold condition with a condition where the confederate had unhindered access to the stimuli (cf. Chapter 3 and Chapter 4, Experiment 3). If the experimental effects were indeed driven by the directional cues from the confederate's body orientation (Heyes, 2014), then it should not matter whether or not the confederate had visual access to the stimuli – as the directional cues exhibited by the other remained constant throughout the blindfold manipulation.

Hence, we should have then found the compatibility / congruency effect in both the blindfolded and the seeing condition. Crucially, in Chapter 3 we only found the compatibility effect if the confederate had unhindered visual access to the stimuli. Likewise, also in Chapter 4 the congruency effect (and thereby: evidence for VSP-taking) only occurred if the confederate had visual access to the word stimuli, but not if the confederate's visual access was blocked by means of the shutter glasses. Together, these findings provide converging evidence that the visual access of the other agent creates an effective boundary condition for spontaneous VSP-taking to occur (cf. Chapter 5.4), which, in turn excludes lower level explanations such as attention re-orienting (Heyes, 2014).

Thus, judging from the data we acquired, VSP-taking seems to be a phenomenon that is 'social in its nature' – as throughout our studies, it was only triggered in the presence of another intentional agent having unhindered visual access to the stimuli, which he could interpret in a meaningful way. Future experiments will have to determine the minimal 'social requirements' triggering spontaneous VSP-taking (see also next section). For example, it would be interesting to investigate both the Spatial Compatibility paradigm as well as the Reading paradigm in the context of a robot co-actor– and manipulate the robot's ability to master spatial judgments (such as 'left' and 'right') and reading, respectively. This way, one could investigate whether spontaneous VSP-taking is triggered in the presence of an agent that has perceptional states (that is, reacting to the environment), but no mental states.

Mentalizing and Spontaneous VSP-Taking

Another important aspect about the social nature of VSP-taking concerns the question exactly what mental representations were spontaneously ascribed to the other person during perspective-taking and hence, how much participants were mentalizing. Summarizing how much evidence for mentalizing was involved in our studies *at the very least* (i.e. at a minimal level), it seems as if participants in the Spatial Compatibility paradigm took into account whether the confederate participated in the task or not (Chapter 2, Experiment 3), whether he responded with the same or with different hands (Chapter 2, Experiment 4, as indicated by the interaction effect) and whether he had visual access to the stimuli or not (Chapter 3).

In Chapter 4 it seems as if participants were sensitive to the fact that the other person could read the word stimuli even in the absence of any overt responses (Chapter 4, Experiment 2), and that the confederate could not read the stimuli while being blindfolded (Chapter 4, Experiment 3).

Thus, in a minimalistic sense, we found evidence that participants must have represented the confederate's mental state *beyond the level of percepts,* as they did not only pick up on the confederate's visual access, but also took into account specifics of the confederate's motor planning (Chapter 2, Experiment 4), as well as him being able to semantically process certain stimuli (Chapter 4). As we mention at the end of Chapter 4, future experiments will have to determine which level of word processing (e.g. semantic or orthographic) was actually employed during spontaneous VSP-taking in this specific study.

Outlook: Individual Differences and Social Modulations

Besides having shown how humans spontaneously adopt another's visuospatial perspective both in physical and in mental space, a further topic of interest would be to see how spontaneous VSP-taking is modulated both by inter-individual differences (cf. Bukowski & Samson, 2017) as well as in intergroup contexts (Simpson & Todd, 2017).

Inter-Individual Differences in Perspective-Taking

A recent study has demonstrated that attentional focus and conflict handling pose two important sources in order to explain individual differences in Level-1 perspective-taking (Bukowski & Samson, 2017). More specifically, Bukowski and Samson (2017) conducted a cluster analysis on 346 participants who completed the dot-perspective-taking task (Samson et al., 2010, and see Chapter 1 for a description of the task) and evaluated individual performances along two dimensions: 1) the ability to handle conflict between the egocentric and the altercentric perspective and 2) the relative attentional focus on the egocentric perspective versus the altercentric perspective during task performance.

Their results showed high heterogeneity along both dimensions giving a richer account on the source of inter-individual variability of the perspective-taking performance. More specifically, on one end Bukowski and Samson (2017) characterized a group of individuals that handled conflicts in an efficient manner or focused on the other person's perspective rather than on their own (the 'good perspective-takers'). On the other end, another group was characterized by individuals who either had more difficulty in handling conflicting perspectives or strongly focused on their egocentric perspective (the 'poor perspective-takers', cf. Bukowski & Samson, 2017). The authors argue that this partitioning thus allows for a better understanding of individual differences in perspective-taking tasks and, on a more broader level, provide a better understanding of the origin of successful versus unsuccessful perspective-taking (Bukowski & Samson, 2017). In this respect it would be interesting to investigate whether individual differences can be found along the same cognitive dimensions (i.e. attentional focus and conflict handling) when it comes to spontaneous VSP-taking.

Perspective-Taking in the Context of Group Affiliation

Furthermore, while we found VSP-taking to be a spontaneous process in our studies, another interesting endeavor would be to investigate whether it can nevertheless be modulated by group affiliation. Prior work on group identity has already shown that visual perspective-taking can be modulated by perspective-taking targets being part of an in-group vs. an out-group (Todd et al., 2017).

A recent study by Simpson and Todd (2017) investigated how group membership modulates Level-1 visual perspective-taking (L1-VPT). To properly capture the specific influence of group membership on the perspective-taking process, Simpson and Todd (2017) used the dot perspective-taking task (Samson et al., 2010). Referring to previous studies, they note that people are more likely to use accessible self-knowledge when making inferences about the mental states of others that are *similar* compared to *dissimilar to themselves* (Simpson & Todd, 2017; Todd, Simpson, & Tamir, 2016). This led them to predict that egocentric intrusion effects during the dot perspective-task should be stronger with an in-group compared to an out-group avatar (Simpson & Todd, 2017). In order to test this prediction they then modulated group membership of the avatar in the dot perspective-taking task through university affiliation (Experiment 1) and through a minimal group manipulation (Experiment 2).

The results revealed that across experiments, participants indeed displayed stronger egocentric intrusion effects for in-group compared to for out-group avatars (Simpson & Todd, 2017). Important to notice though, they did not find any evidence for group membership affecting the other perspective-taking component, namely the altercentric intrusion effect. Taken together, Simpson and Todd (2017) argue that their results demonstrate that shared group membership can bias L1-VPT by selectively impairing the inhibition of one's own perspective (Simpson & Todd, 2017).

Against this background, it would be interesting to see whether group affiliation only affects Level-1 perspective-taking or whether it also modulates spontaneous VSP-taking. For example, we found that participants spontaneously took into account the VSP of the confederate – but could a minimal group paradigm actually diminish or intensify these effects?

Summary of Findings and Conclusion

This dissertation investigated the phenomenon of spontaneous visuospatial perspective-taking in humans. Perspective-taking is a key component of social interactions. However, this thesis started with many open questions about whether, when and how instances of spontaneous visuospatial perspective-taking occur during social interactions. The study in Chapter 2 investigated the underlying factors as well as boundary conditions that characterize the spontaneous adoption of another person's visuospatial perspective (VSP). The results of the study showed that participants spontaneously adopted a differing VSP, given there was an intentionally acting agent alongside of them. To our knowledge, these are the first findings showing that humans adopt another person's VSP by all respects *spontaneously*; that is, in the absence of a communicative context and without being prompted to do so. At the same time, many aspects about the mechanisms underlying spontaneous VSP-taking remained unknown. Chapter 3 explored whether knowledge about another's visual access systematically modulates spontaneous VSP-taking. In two experiments we found that knowledge about another person's visual access indeed modulated the spontaneous integration of another person's VSP into one's own action planning. Specifically, our findings show that participants only adopted the other person's VSP if he had unhindered visual access to the stimuli but regardless of whether or not he performed the same task or a different task.

The results of the studies in Chapter 2 and 3 (in addition to studies from other research groups showing similar results) suggest that humans are indeed capable of spontaneously adopting each other's visuospatial perspectives. Yet, all of the evidence for spontaneous VSP-taking came from tasks where an observed agent or task partner could physically act upon objects. Thus, it was unknown whether spontaneous VSP-taking extends to mental space where another person's perspective matters for mental activities rather than for physical actions. The study in Chapter 4 investigated whether spontaneous VSP-taking occurs in mental space. In three experiments participants reliably adopted the VSP of a confederate in the context of a semantic categorization task that involved reading words. Thus, to our knowledge, these are the first results showing that people spontaneously adopt another's VSP in mental space.

Taken together, this thesis contributes to on-going debates in cognitive psychology by clarifying whether, when and how humans spontaneously adopt another's visuospatial perspective. The present work also contributes to the field of social cognition by highlighting our propensity to form representations from others' perspectives, which opens new questions about the interplay between spatial and social relations.

References

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

Atmaca, S., Sebanz, N., Prinz, W., & Knoblich, G. (2008). Action co- representation: The joint SNARC effect. *Social Neuroscience, 3,* 410– 420. http://dx.doi.org/10.1080/17470910801900908

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

Baker, L. J., Levin, D. T., & Saylor, M. M. (2016). The extent of default visual perspective taking in complex layouts. *Journal of Experimental Psychology: Human Perception and Performance, 42,* 508–516. http:// dx.doi.org/10.1037/xhp0000164

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*(1), 37–46.

Barrett, H. C., Broesch, T., Scott, R. M., He, Z., Baillargeon, R., Wu, D., ... Laurence, S. (2013). ESM for Early false-belief understanding in traditional non-Western societies. *Proceedings. Biological Sciences / The Royal Society, 280*(1755), 20122654. http://doi.org/10.1098/rspb.2012.2654

Birch, S. A. J., & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in Cognitive Sciences, 8,* 255–260.

Birch, S., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, 18, 382–386.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, *77*(1), B25-B31.

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,* 531–545. http://dx.doi.org/10.1007/s00221-011- 2625-z

Böckler, A., Knoblich, G., & Sebanz, N. (2012). Effects of a coactor's focus of attention on task performance. *Journal of Experimental Psychology: Human Perception and Performance, 38*(6), 1404.

Böckler, A., & Zwickel, J. (2013). Influences of spontaneous perspective taking on spatial and identity processing of faces. *Social Cognitive and Affective Neuroscience, 8,* 735–740. http://dx.doi.org/10.1093/scan/ nss061

Bratman, M. E. (1992). Shared cooperative activity. *The Philosophical Review, 101,* 327–341. http://dx.doi.org/10.2307/2185537

Bräuer, J., Call, J. & Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Anim. Cogn.* 10, 439–448.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition, 106*(3), 1465-1477.

Bugnyar, T. & Kotrschal, K. (2002). Observational learning and the raiding of food caches in ravens, Corvus corax: is it 'tactical' deception? *Anim. Behav.* 64, 185 – 195.

Bugnyar, T., Stöwe, M. & Heinrich, B. (2004). Ravens, Corvus corax, follow gaze direction of humans around obstacles. *Proc. R. Soc. Lond.* B 271, 1331 – 1336.

Bugnyar, T. & Heinrich, B. (2005). Ravens, Corvus corax, differentiate between knowledgeable and ignorant competitors. *Proc. R. Soc.* B 272, 1641 – 1646.

Bugnyar, T. (2011). Knower–guesser differentiation in ravens: others' viewpoints matter. *Proceedings of the Royal Society B: Biological Sciences*,*278*(1705), 634-640.

Bukowski, H., & Samson, D. (2017). New Insights into the Inter-Individual Variability in Perspective Taking. *Vision, 1*(1), 8.

Butterfill, S. A., & Sebanz, N. (2011). Editorial: Joint action: What is shared? *Review of Philosophy and Psychology, 2,* 137–146. http://dx.doi .org/10.1007/s13164-011-0062-3

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind and Language, 28*(5), 606–637. http://doi.org/10.1111/mila.12036

Call, J. & Tomasello, M. (1998). Distinguishing intentional from accidental actions in orangutans (Pongo *pygmaeus),* chimpanzees (Pan *troglodytes),* and human children (Homo *sapiens). Journal of Comparative Psychology,* 112, 192–206.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12,* 187–192. http:// dx.doi.org/10.1016/j.tics.2008.02.010

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the society for research in child development*, i-174.

Castelhano, M. S., Wieth, M. S., & Henderson, J. M. (2007). I see what you see: Eye movements in real-world scenes are affected by perceived direction of gaze. In L. Paletta and E. Rome (Eds.), *Attention in Cognitive Systems* (pp 252-262). Berlin: Springer.

Cho, Y. S., & Proctor, R. W. (2003). Stimulus and response representations underlying orthogonal stimulus-response compatibility effects. *Psychonomic Bulletin & Review, 10,* 45–73.

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

Clements, W., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 395*, 377–395.

Cohen, A. S., & German, T. C. (2009). Encoding of others' beliefs without overt instruction. *Cognition, 111*(3), 356–363. http://doi.org/10.1016/j.cognition.2009.03.004

Conson, M., Mazzarella, E., Donnarumma, C., & Trojano, L. (2012). Judging hand laterality from my or your point of view: Interactions between motor imagery and visual perspective. *Neuroscience Letters, 530,* 35–40. http://dx.doi.org/10.1016/j.neulet.2012.09.051

Costantini, M., Committeri, G., & Sinigaglia, C. (2011). Ready both to your and to my hands: Mapping the action space of others. *PLoS ONE, 6,* e17923. http://dx.doi.org/10.1371/journal.pone.0017923

Craft, J. L., & Simon, J. R. (1970). Processing symbolic information from a visual display: Interference from an irrelevant directional cue. *Journal of Experimental Psychology, 83,* 415–420. http://dx.doi.org/10.1037/ h0028843

Creem-Regehr, S. H., Gagnon, K. T., Geuss, M. N., & Stefanucci, J. K. (2013). Relating spatial perspective taking to the perception of other's affordances: Providing a foundation for predicting the future behavior of others. *Frontiers in Human Neuroscience, 7,* 596. http://dx.doi.org/10.3389/fnhum.2013.00596

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences, 13*(4), 148-153.

Csibra, G., & Volein, A. (2008). Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology, 26*(1), 1-11.

Dally, J. M., Emery, N. J. & Clayton, N. S. 2006. Food-caching western scrub-jays keep track of who was watching when. *Science* 312, 1662–1665.

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews, 3*, 71–100.

Decety, J., & Lamm, C. (2006). Human empathy through the lens of social neuroscience. *Scientific World Journal, 6*, 1146–1163.

Decety, J., & Ickes, W. (2009). The social neuroscience of empathy. MIT Press.

Decety, J., & Lamm, C. (2009). Empathy vs. personal distress. In J. Decety & W. Ickes (Eds.), The Social Neuroscience of Empathy. Cambridge: MIT Press.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371.

Deligianni, F., Senju, A., Gergely, G., & Csibra, G. (2011). Automated gaze-contingent objects elicit orientation following in 8-month-old infants. Developmental psychology, 47(6), 1499.

Dennett, D. C. (1978). Beliefs about beliefs [P&W, SR&B]. *Behavioral and Brain Sciences, 1*(4), 568–570.

D'Entremont B, Hains SMJ, Muir DW. A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior and Development*. 1997;20:569–572.

Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science, 13,* 331–338. http://dx.doi.org/10.1111/j.1467-7687.2009.00888.x

Driver IV, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition, 6*(5), 509-540.

Elekes, F., Varga, M., & Király, I. (2016). Evidence for spontaneous Level-2 perspective taking in adults. *Consciousness and Cognition, 41,* 93–103. http://dx.doi.org/10.1016/j.concog.2016.02.010

Emery, N. J. & Clayton, N. S. 2001. Effects of experience and social context on prospective caching strategies by scrub jays. *Nature* 414, 443 – 446.

Emery, N. J. & Clayton, N. S. 2004. The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science* 306, 1903 – 1907.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87,* 327–339.

Fauconnier, G. (1994). Mental spaces: Aspects of meaning construction in natural language. Cambridge University Press.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Fitts, P. M., & Deininger, R. L. (1954). S-R compatibility: Correspondence among paired elements within stimulus and response codes. *Journal of Experimental Psychology, 48,* 483-492.

Fitts, P. M., & Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology, 46*, 199-210.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young childrens knowledge about visual-perception: Further evidence for the Level-1–Level-2 distinction. *Developmental Psychology, 17,* 99 –103. http://dx.doi.org/10.1037/0012-1649.17.1.99

Flombaum, J. & Santos, L. (2005). Rhesus monkeys attribute perceptions to others. *Current Biol.* 15: 447-452.

Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2016). When do humans spontaneously adopt another's visuospatial perspective? *Journal of Experimental Psychology: Human Perception and Performance, 42,* 401– 412. http://dx.doi.org/10.1037/xhp0000153

Freundlieb, M., Sebanz, N., & Kovács, Á. M. (2017). Out of your sight, out of my mind: knowledge about another person's visual access modulates spontaneous visuospatial perspective-taking. *Journal of Experimental Psychology: Human Perception and Performance*.

Frischen, A., Loach, D., & Tipper, S. P. (2009). Seeing the world through another person's eyes: Simulating selective attention via action observation. *Cognition, 111,* 212–218. http://dx.doi.org/10.1016/j.cognition .2009.02.003

Furlanetto, T., Becchio, C., Samson, D., & Apperly, I. (2016). Altercentric interference in Level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Journal of Experimental Psychology: Human Perception and Performance, 42,* 158–163.

Furlanetto, T., Cavallo, A., Manera, V., Tversky, B., & Becchio, C. (2013). Through your eyes: Incongruence of gaze and action increases spontaneous perspective taking. *Frontiers in Human Neuroscience, 7,* 455. http://dx.doi.org/10.3389/fnhum.2013.00455

Galati, A., Michael, C., Mello, C., Greenauer, N. M., & Avraamides, M. N. (2013). The conversational partner's perspective affects spatial memory and descriptions. *Journal of Memory and Language, 68,* 140–159. http://dx.doi.org/10.1016/j.jml.2012.10.001

Garnham, W. A., & Ruffman, T. (2001). Doesn't see, doesn't know: Is anticipatory looking really related to understanding or belief? *Developmental Science, 4*(1), 94–100. http://dx.doi.org/10.1111/1467-7687.00153.

German, T. P., & Hehman, J. A. (2006). Representational and executive selection resources in 'theory of mind': Evidence from compromised belief-desire reasoning in old age. *Cognition, 101*(1), 129-152.

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

Hare, B., Call, J. & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour,* 61, 139–151.

Hare, B., Addessi, E., Call, J., Tomasello, M. & Visalberghi, E. (2003). Do capuchin monkeys, *Cebus apella,* know what conspecifics do and do not see? *Anim. Behav.* 65: 131-142.

Hare B, Call J, Tomasello M (2006). Chimpanzees deceive a human competitor by hiding. *Cognition* 101:495–514.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. *Current directions in psychological science, 2*(3), 96-100.

Heinrich, B. & Pepper, J. R. 1998. Influence of competitors on caching behaviour in common ravens, Corvus corax. *Anim. Behav.* 56, 1083–1090.

Herrmann, E., Call, J., Hernàndez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science, 317*(5843), 1360–6. http://doi.org/10.1126/science.1146282.

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science, 9,* 131–143. http://dx.doi.org/10 .1177/1745691613518076

Hirata, S. & Matsuzawa, T. (2001). Tactics to obtain a hidden food item in chimpanzee pairs (Pan *troglodytes*). *Animal Cognition,* 4, 285–295.

Hobson, R. P. (1989). On sharing experiences. *Development and Psychopathology, 1*(3), 197-203.

Hood BM, Macrae CN, Cole-Davies V, Dias M. Eye remember you: The effects of gaze direction on face recognition in children and adults. *Developmental Science*. 2003;6:67–71.

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. *Anim. Behav.* 69: 11-18.

Kessler, K., & Thomson, L. A. (2010). The embodied nature of spatial perspective taking: Embodied transformation versus sensorimotor interference. *Cognition, 114,* 72– 88. http://dx.doi.org/10.1016/j.cognition .2009.08.015

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in compre- hension. *Psychological Science, 11,* 32–38. http://dx.doi.org/10.1111/ 1467-9280.00211

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89,* 25–41. http://dx.doi.org/10.1016/S0010- 0277(03)00064-7

Klatzky RL (1998) Allocentric and egocentric spatial representations: definitions, distinctions, and interconnections. In: Freksa C, Habel C, Wender KF (eds) Spatial cognition—an interdisciplinary approach to representation and processing of spatial knowledge. Springer, Berlin, pp 1–17

Koriat, A., & Norman, J. (1985). Reading rotated words. *Journal of Experimental Psychology: Human Perception and Performance, 11*(4), 490.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility—A model and tax- onomy. *Psychological Review, 97,* 253–270. http://dx.doi.org/10.1037/ 0033-295X.97.2.253

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830-1834.

Langton, S. R., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in cognitive sciences, 4*(2), 50-59.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476–490.

Lozano, S. C., Hard, B. M., & Tversky, B. (2007). Putting action in perspective. Cognition, 103(3), 480-490.

Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive Architecture of Belief Reasoning in Children and Adults: A Primer on the Two-Systems Account. *Child Development Perspectives, 0*(0), 1–6. http://doi.org/10.1111/cdep.12183

Mattan, B., Quinn, K. A., Apperly, I. A., Sui, J., & Rotshtein, P. (2015). Is it always me first? Effects of self-tagging on third-person perspective- taking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 1100–1117. http://dx.doi.org/10.1037/xlm0000078

Mazzarella, E., Hamilton, A., Trojano, L., Mastromauro, B., & Conson, M. (2012). Observation of another's action but not eye gaze triggers allocentric visual perspective. *The Quarterly Journal of Experimental Psychology, 65,* 2447–2460. http://dx.doi.org/10.1080/17470218.2012.697905

Meltzoff, A. N., & Brooks, R. (2008). Self-experience as a mechanism for learning about others: A training study in social cognition. *Developmental Psychology, 44,* 1257–1265. http://dx.doi.org/10.1037/a0012888

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & Psychophysics, 68,* 327–337. http://dx.doi.org/10 .3758/BF03193680

Moore C, Corkum V. Social understanding at the end of the first year of life. *Developmental Review*. 1994;14:349–372.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?. S*cience*, *308*(5719), 255-258.

Perner, J. (1991). *Understanding the representational mind*. The MIT Press.

Perner, J., Brandl, J. L., & Garnham, A. (2003). What is a perspective problem? Developmental issues in belief ascription and dual identity. *Facta Philosophica, 5,* 355–378.

Povinelli, D. J. & Eddy, T. J., (1996a). What young chimpanzees know about seeing. *Monographs of the Society for Research in Child Development,* 61, 1–152.

Povinelli, D. J. & Eddy, T. J., (1996b). Factors influencing young chimpanzees recognition of 'attention'. *Journal of Comparative Psychology,* 110, 336–345.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 4*, 515–526.

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition, 117,* 230–236. http://dx.doi.org/10.1016/j.cognition.2010.08.003

Ramsey, R., Hansen, P., Apperly, I., & Samson, D. (2013). Seeing it my way or your way: Frontoparietal brain areas sustain viewpoint- independent perspective selection processes. *Journal of Cognitive Neu- roscience, 25,* 670–684. http://dx.doi.org/10.1162/jocn_a_00345

Ricciardelli, P., Bricolo, E., Aglioti, S. M., & Chelazzi, L. (2002). My eyes want to look where your eyes are looking: Exploring the tendency to imitate another individual's gaze. *Neuroreport, 13*(17), 2259-2264.

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science, 26,* 867–891. http://dx.doi.org/10.1016/j.humov.2007.07.002

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1255–1266. http://dx.doi.org/10.1037/a0018729

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance, 40,* 929–937. http://dx.doi.org/10.1037/a0035175

Santiesteban, I., Shah, P., White, S., Bird, G., & Heyes, C. (2015). Mentalizing or submentalizing in a communication task? Evidence from autism and a camera control. *Psychonomic Bulletin & Review, 22,* 844–849. http://dx.doi.org/10.3758/s13423-014-0716-0

Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of interpersonal coordination. In A. Fuchs & V. K. Jirsa (Eds.), *Coordination: Neural, behavioral and social dynamics* (pp. 281–308). Berlin, Germany: Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-540-74479- 5_14

Schober, M. F. (1995). Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about location? *Discourse Processes, 20,* 219–247. http://dx.doi.org/10.1080/01638539509544939

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

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition, 88,* B11–B21. http://dx.doi.org/ 10.1016/S0010-0277(03)00043-X

Sebanz, N., Knoblich, G., & Prinz, W. (2005). How two share a task: Corepresenting stimulus-response mappings. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1234–1246. http://dx.doi.org/10.1037/0096-1523.31.6.1234

Senju A, Hasegawa T. Direct gaze captures visuospatial attention. *Visual Cognition*. 2005;12:127–144.

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

Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychological Science, 22,* 878 – 880. http://dx.doi.org/10.1177/ 0956797611411584

Shelton, A. L., & McNamara, T. P. (2004). Spatial memory and perspec- tive taking. *Memory & Cognition, 32,* 416–426. http://dx.doi.org/10 .3758/BF03195835

Simon, J. R., Small, A. M., Ziglar, R. A., & Craft, J. L. (1970). Response interference in an information processing task: Sensory versus perceptual factors. *Journal of Experimental Psychology, 85*(2), 311.

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus-response compatibility: An integrated perspective* (pp. 31– 86). Amsterdam, the Netherlands: North-Holland. http://dx.doi.org/10.1016/ S0166-4115(08)61218-2

Simpson, A. J., & Todd, A. R. (2017). Intergroup visual perspective-taking: Shared group membership impairs self-perspective inhibition but may facilitate perspective calculation. *Cognition*, 166, 371-381.

Sjolund, L. A., Erdman, M., & Kelly, J. W. (2014). Collaborative inhibition in spatial memory retrieval. *Memory & Cognition, 42,* 876–885. http:// dx.doi.org/10.3758/s13421-014-0407-0

Strijkers, K., Bertrand, D., & Grainger, J. (2015). Seeing the same words differently: The time course of automaticity and top–down intention in reading. *Journal of Cognitive Neuroscience*.

Soechting JF, Flanders M (1992) Moving in three-dimensional space: frames of reference, vectors, and coordinate systems. *Annu Rev Neurosci 15*:167–191

Sodian, B., Thoermer, C., & Metz, U. (2007). Now I see it but you don't: 14- month- olds can represent another person's visual perspective. *Developmental Science*, *10*(2), 199-204.

Song, H. J., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental psychology*, *44*(6), 1789.

Sterck EHM, Watts DP, van Schaik CP (1997). The evolution of female social relationships in nonhuman primates. *Behav Ecol Sociobiol* 41:291–309.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological science*, *18*(7), 580-586.

Surtees, A., Apperly, I., & Samson, D. (2013). Similarities and differences in visual and spatial perspective-taking processes. *Cognition, 129,* 426 – 438. http://dx.doi.org/10.1016/j.cognition.2013.06.008

Surtees, A. D. R., Samson, D., & Apperly, I. A. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, 148, 97– 105. http://doi.org/10.1016/j.cognition.2015.12.010

Surtees, A., Apperly, I., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition, 150,* 43–52. http://dx.doi.org/10.1016/j.cognition.2016.01.014

Surtees, A. D., Noordzij, M. L., & Apperly, I. A. (2012). Sometimes losing your self in space: Children's and adults' spontaneous use of multiple spatial reference frames. *Developmental Psychology, 48,* 185–191. http://dx.doi.org/10.1037/a0025863

Todd, A. R., Simpson, A. J., & Tamir, D. I. (2016). Active perspective taking induces flexible use of self-knowledge during social inference. *Journal of Experimental Psychology: General,* 145, 1583–1588.

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. First language, 4(12), 197-211.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child development*, 1454-1463.

Tomasello, M. (1995). Joint attention as social cognition. *Joint attention: Its origins and role in development*, 103-130.

Tomasello, M., Call, J. & Hare, B. (1998). Five primate species follow the visual gaze of conspecifics. *Animal Behaviour,* 55, 1063–1069.

Tomasello, M. (1999). The human adaptation for culture. *Annual review of anthropology, 28*(1), 509-529.

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

Tsai, C. C., Kuo, W. J., Hung, D. L., & Tzeng, O. J. (2008). Action co-representation is tuned to other humans. *Journal of Cognitive Neu- roscience, 20,* 2015–2024. http://dx.doi.org/10.1162/jocn.2008.20144

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition, 110,* 124–129. http://dx.doi.org/ 10.1016/j.cognition.2008.10.008

Vogeley, K., & Fink, G. R. (2003). Neural correlates of the first-person- perspective. *Trends in Cognitive Sciences, 7,* 38–42. http://dx.doi.org/ 10.1016/S1364-6613(02)00003-7

von Grünau, M., & Anston, C. (1995). The detection of gaze direction: A stare-in-the-crowd effect. *Perception, 24*(11), 1297-1313.

Vuilleumier, P., George, N., Lister, V., Armony, J., & Driver, J. (2005). Effects of perceived mutual gaze and gender on face processing and recognition memory. *Visual Cognition, 12*(1), 85-101.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta- analysis of theory- of- mind development: the truth about false belief. *Child development*, *72*(3), 655-684.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103-128.

Zwickel, J. (2009). Agency attribution and visuospatial perspective taking. *Psychonomic Bulletin & Review, 16,* 1089–1093. http://dx.doi.org/10 .3758/PBR.16.6.1089

Appendix

Table A1. Word Lists (Chapter 4, Exp. 1-3). Font: Cambria, Regular, 58; all capital letters.

ÁLLATOK (Animals)	**NÖVÉNYEK (Fruit & Vegetables)**
antilop	áfonya
bölény	ananász
cinke	bodza
poloska	brokkoli
cinege	cékla
gorilla	cukkini
hiéna	datolya
hiúz	egres
hörcsög	füge
katica	gesztenye
kenguru	kapor
keselyű	karfiol
koala	kelbimbó
lajhár	kókusz
leopárd	kökény
vidra	köles
orrszarvú	mandarin
panda	mangó
papagáj	menta
pingvin	padlizsán
pocok	citromfű
polip	spenot
sündisznó	retek
rozmár	kakukkfű
szöcske	sóska
teknős	sütőtök
vaddisznó	szamóca
vakond	szeder
viziló	torma
gazella	újhagyma
zebra	zeller
zsiráf	zöldborsó