2020

LEVEL-TWO VISUAL PERSPECTIVE TAKING AS SPONTANEOUS PERCEPTUAL SIMULATION

Ward, Eleanor

http://hdl.handle.net/10026.1/16524

http://dx.doi.org/10.24382/614
University of Plymouth

Copyright statement

UNIVERSITY OF PLYMOUTH

LEVEL-TWO VISUAL PERSPECTIVE TAKING AS SPONTANEOUS PERCEPTUAL SIMULATION

By

ELEANOR K A WARD

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Psychology

July 2020

Acknowledgements

Primarily my gratitude goes to my three favourite nerds. Firstly, Dr Patric Bach, saviour of experiments, first-class excel wizard, and proof-reader extraordinaire. I would never have made it this far without your patience, data enthusiasm and encouragement (and bad ass writing skills). Thank you for being the best supervisor and mentor a person could ask for. I hope your future students appreciate how fantastic you are (perhaps don't introduce them to your fancy dance moves too early on)!

Secondly, Dr Katrina McDonough – co-author, co-student, co-lunatic. For keeping me steady, and helping me through the tough times, I owe you big. Love you, mate.

Finally, my doting, encouraging, unremittingly pragmatic, and all round supporting number one nerd. Abe Kallis, partner in grime. You've held me up through this, and I am so **so** lucky to have you by my side.

My thanks also go out to Dr Giorgio Ganis, who has granted me absolute autonomy in my work. I am a far better researcher for having had the freedom to explore my ideas. Thank you for your support and encouragement. To the University of Plymouth, thank you for supporting my travels and collaborations. To Anthony et al. in the tech office – your tolerance will not be forgotten.

Finally to my friends and family who have patted me of the back, given me wine, looked after the children, fed me, funded me…I couldn't have done this without you.

Thank you all.

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Doctoral College Quality Sub-Committee.

Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment.

This study was financed with the aid of a studentship from the University of Plymouth. In addition, two bursaries were awarded by University of Plymouth in 2017 and 2018 to support participant recruitment and a third bursary was awarded to attend the Joint Action Meeting in Genova in 2019.

The research in thesis has been published in scientific journals and has been presented as talks and posters at various conferences (listed below).

Publications:

Chapter 3

Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. ***Current Biology***, *29*(5), 874-880.

Chapter 4

Ward, E., Ganis, G., McDonough, K. L., & Bach, P. (2020). Perspective taking as virtual navigation? Perceptual simulation of what others see reflects their location in space but not their gaze. ***Cognition***, *199*, 104241.

Chapter 5

Ward, E., Ganis, G., McDonough, K. L., & Bach, P. (*submitted*). Level-2 Visual perspective taking as an embodied process? Perceptual simulation of others' perspectives is not impaired by motor restriction. Preprint available at https://psyarxiv.com/84q7g/

Conference talks:

07/2019 **Ward, E.**, Ganis, G., Bach, P. Visual Perspective Taking as Perceptual Simulation in Social Interactions. Oral Presentation at InTo Brain Joint Action Meeting (JAMVIII), Genoa, Italy.

06/2019 **Ward, E.**, Ganis, G., Bach, P. Visual Perspective Taking as Perceptual Simulation. Oral Presentation at University of Plymouth Cognition Institute Conference, University of Plymouth, UK.

12/2018 **Ward, E**., Ganis, G., Bach, P. Spontaneous Vicarious Perception of the Content of Others' Visual Perspectives. Oral Presentation at VSI 2018, 4th Workshop on Virtual Social Interaction, Goldsmiths, University of London, UK.

12/2018 **Ward, E**., Ganis, G., Bach, P. Spontaneous Vicarious Perception of the Content of Others' Visual Perspectives. Oral Presentation at the Action Prediction Lab Annual Christmas Conference, University of Plymouth, UK.

06/2018 **Ward, E**., Ganis, G., Bach, P. Seeing Other Minds: Automatic Perspective Taking as Perceptual Simulation. Oral Presentation at the University of Plymouth School of Psychology Conference, University of Plymouth, UK.

Conference posters:

05/2018 **Ward, E**., Ganis, G., Bach, P. Seeing Other Minds: Automatic Perspective Taking as Perceptual Simulation. Poster Presentation at InTo Brain Conference 'MeeTo' - From Moving Bodies to Interactive Minds Conference, University of Turin, Italy.

Word count of main body of thesis: 49,856 words

Signed ………………………………

Date …………………………………

Level-two Visual perspective taking as perceptual simulation.

Eleanor K A Ward

Abstract

Visual perspective taking lies at the heart of social interactions, allowing people to readily compute what another person can see, or how they see it (Flavell et al., 1981). This ability not only links to the unique human propensity to infer others' beliefs, desires, and goals (Schurz et al., 2015), but also to more general spatial abilities such as navigation skill (Kozhevnikov et al., 2006) and action planning (Creem-Regehr et al., 2013). What is unclear is how others' perspectives are represented. While several studies have in the past *implied* that others' perspectives are represented in a similar way as our own viewpoint (e.g. Freundlieb et al., 2016; Sampson et al., 2010; Surtees et al., 2016) none have demonstrated the form that this overlapping perspective takes, and how these altercentric representations interfere with our own.

The experiments in this thesis provide the first known evidence that when taking another's perspective, representations of their visual input take a (quasi)-perceptual form, and can drive perceptual decision making in the same way as own input (e.g. Roelfsema & de Lange, 2016). Using a modified mental rotation paradigm (Shepard & Metzler, 1971), we simply asked participants to state whether alphanumeric characters appearing at varying orientations away from upright were canonical or mirror inverted (e.g. "R" vs. "Я"). In some scenes another person was present, seen sitting at 90 degrees either to the left or the right of the character. We

first confirmed the classic mental rotation effect, where response times increase gradually as a function of the item's increasing angular disparity away from upright. Crucially, we also found that the angular disparity relative to the inserted persons also produced this response pattern, and found that participants responded surprisingly quickly to items that would normally be difficult when they were facing the other person. These effects occurred spontaneously, and specifically in the presence of a human (but not a lamp), even when the other person was task irrelevant. Subsequent studies showed these effects were not sensitive to the looking direction of the other person, but instead reflected their position in space more generally. Finally, we found that a person's perceived ability to move (and therefore motorically emulate the position of the other person) did not influence these effects.

These findings show that (1) others' perspectives can indeed 'stand in' for our own input, taking the form of a perceptual simulation, which paints the (imagined) perspective of another person onto our own perception, (2) these simulations reflect what someone could – in principle – see from their position in space, and speculate that (3) VPT emerged primarily to support navigation of, and interaction with, the environment and that more sophisticated Theory of Mind may then capitalise on the epistemic insight derived through more primitive forms of VPT.

Table of Contents

List of figures

1 Chapter One – Knowing how the world looks to others

1.1 Visual Perspective Taking in social interactions

Every-day life gives rise to many situations where it is helpful to know what another person can see, or how the world looks to them (Flavell, Everett, Croft & Flavell, 1981), and humans do-so effortlessly via a process commonly referred to as visual perspective taking (VPT). Traditionally, visual perspective taking is thought to be comprised of two levels, one and two (Flavell et al., 1981). First to be acquired is level one VPT (VPT-1), which allows a person to identify whether or not another person can see something, perhaps by following their line of sight from their eyes to an object (e.g. Michelon & Zachs, 2006). Following this, a more sophisticated ability to infer *how* an object looks to another person is acquired, allowing for more complex visual and spatial judgements. This ability allows one to decide what a “6” looks like a “9” to a person opposite to oneself, or that one’s own right will be their left, for example. This ability is commonly referred to as level two VPT (VPT-2).

There are numerous everyday situations where we might benefit from understanding what another person can see, or how they see it. With regard to VPT-1, consider, for example, an attempt to cross a busy street. We must first check not only whether we can see an oncoming vehicle, but also whether the driver has noticed us (VPT-1). We use the same ability to hide objects from others, to steal from them, alert them to danger or just simply to point there attention towards objects they are looking for but cannot currently see. VPT-2 plays equally important roles. When we play a competitive sport, we can use VPT-2 to better predict what the other player will do next by using our knowledge about how the game looks to them (e.g. Creem-Regehr, Gagnon, Geuss & Stefanucci, 2013), which can lead to better performance in team sports, such as football (e.g. Mann, Farrow, Shuttleworth, Hopwood & MacMahon, 2009). Similarly, if we try to coordinate our dance moves with a partner, we may make use of VPT-2 to know how our actions appear to our partner as well as ourselves (Freundlieb, Kovacs & Sebanz, 2015). Even in simple day-to-day interactions, such as showing a picture to somebody else, we instinctively orient it so that they can easily see it (Lempers, Flavell & Flavell, 1977), or when a passer-by asks for directions, we do not describe the route from our own perspective, but do so, intuitively, from the other person's (Tversky & Hard, 2009). In each case, we use our knowledge of both *what* others can see, and *how* they see it, to facilitate these social interactions, often without conscious, effortful calculation.

It is well-established that VPT is a key developmental milestone in childhood that facilitates the transition from an egocentrically-anchored world view, to one which incorporates others' perspectives (Piaget & Inhelder, 1956) and is known to be linked to mental state attribution, or 'Theory of Mind' (ToM; Apperly & Butterfill, 2009; Batson, Early, & Salvarani, 1997; Erle & Topolinski, 2015; Tomasello, Carpenter, Call, Behne, & Moll, 2005) and also to empathy (e.g. Chartrand & Bargh, 1999; Decety & Jackson, 2004; Mattan, Rotshtein & Quinn, 2016). Piaget (1929; 1951) placed huge value in perspective taking for both social and cognitive development. Famous for his constructivist developmental theory, Piaget proposed that a child's egocentricism was key determinant of their cognitive and social development, and that children could only progress through each developmental stage through incremental forms of decentration, where children gradually progress to a point where they can hold more than one thought, or representation, in mind at one time.

Piaget and Inhelder’s (1956) Three Mountains Task was classically used to assess children’s accuracy in representing the world from a non-egocentric perspective. In this task, a child would be presented with a model of three mountains, which they were instructed to study from different angles. A doll was then placed at varying positions around the model, and the child would be asked to select, from a series of pictures, how the mountains appeared to the doll.

The age of the child reliably predicted the outcome of this task. Four year olds were unable to imagine what another person could see, generally choosing the picture that showed their own egocentric view. In contrast, at 8 years most children were able to accurately point out how the mountains looked to the doll, leading Piaget and Inhelder (1956) to conclude that children below this age remained egocentric in their thinking. However, the suggestion by Piaget and Inhelder (1956) that children as old as six or seven find it difficult to make judgements from others’ perspectives has long been overturned by growing evidence that children as young as 14 months demonstrate very basic VPT abilities in both explicit (e.g. Borke, 1975; Flavell et al., 1981; Hughes, 1975) and implicit measures (e.g. Kampis, Parise, Csibra & Kovács, 2015; Lempers et al., 1977; Poulin-Dubois, Sodian, Metz, Tilden & Schoeppner, 2007; Sodian, Thoermer & Metz, 2007).

For example, using a looking-time paradigm, Sodian et al. (2007) observed that 14 month old infants spent more time looking at pictures of a person reaching for, and grasping a novel target-object if the person in the pictures could no longer see a previously-grasped object due to the presence of an occluding screen. This indicates that these infants therefore possessed some awareness of what the other person could and could not see, and more critically, used this knowledge to rationalise the observed action. Poulin-Dubois et al. (2004) also used a preferential looking paradigm to examine the seeing-knowing relationship in children ranging in age from 14 to 24 months. At 18 months old, children showed longer looking times for unexpected outcomes (an adult finding a concealed object despite not watching it during a hiding phase). Again, this study demonstrates that young children are able to infer another's epistemic state and use this to derive predictions about how they expect them to act, and indicates an important role of VPT in understanding others' actions.

Other studies show that young children are able to use their knowledge of what others can see to actively engage in social interactions. Lempers et al. (1977) found that children as young as 18 months would reposition themselves to align with another person to whom they were showing a picture, and by the age of 3 years, most children knew to orient a picture away from themselves to appear upright to a person standing opposite themselves. This may suggest that young children possess implicit knowledge not only about what another person can see (Sodian et al., 2007) but also how this person's perspective differs from their own. This transition from knowing how to position oneself to align perspectives with another, to actively manipulating the position of an object to show it to another person may also mark the emergence of more sophisticated social abilities. If we infer a person's epistemic state via their access to information then, as visual input is one's main source of information, we cannot infer others' internal states without first representing what others can or cannot see.

In adults, VPT is more often measured using tasks where participants are asked to make explicit judgements about what another person can see, or how an object or scene looks to another person (e.g. Sampson, Apperly, Braithwaite, Andres & Bodley Scott, 2010; Kessler & Rutherford, 2010; Kessler & Thompson, 2010).

Explicit tasks are important, as they allow researchers to directly measure a person's ability to represent an object or scene from another person's point of view. One such example was a 'visibility task' used by Michelon and Zacks (2006), wherein participants were shown a scene depicting a doll, a number of objects, and several occluding screens, and were required to make judgements about what the doll could or couldn't see. Participants judged quickly and accurately what could be seen from the doll's perspective, and performance was not influenced by the position of the doll relative to the participants. When the distance between the doll and the object was increased, however, participants took longer to respond. In the same study, judging whether an object appeared to the left or the right of the doll delivered a different pattern of results, whereby response times increased as a function of the doll's increasing angular rotation away from the participant, indicating different types of VPT, and indeed different mechanisms.

Adapted versions of Piaget and Inhelder's (1956) Three Mountains Task have also been used to explicitly measure VPT in children (e.g. Borke, 1975), as have tasks that operate using the same principle (e.g. *how does this scene look to person X*?) but use more 'child friendly' stimuli (e.g. Hamilton, Brindley & Frith, 2009). Borke (1975), for example, simplified Piaget and Inhelder's (1956) experiment design by simply placing the model of the mountains on a revolving surface and instructing children to turn the model to show the perspective of a doll, rather than selecting a scene from a series of photographs. By simplifying the response method by removing the necessity to change modalities between learning and responding, this adaptation led to higher accuracy in younger children, showing that at four years old most children could accurately show how a scene would look to another person.

Other tasks explicitly instruct participants which perspective to take (e.g. self or other; Sampson et al., 2010), allowing for the measurement of behavioural differences when either one's own, or another's perspective is taken when making judgements about an object or scene. One commonly used experimental paradigm requires participants to make numerical judgements about the number of coloured disks appearing in a scene (Sampson et al., 2010). The scene depicts the back wall, and two side walls of a room, in which a human avatar is positioned facing either to the left or to the right. In each trial, coloured disks appear on either one, or both of the walls. In consistent perspective trials, the number of disks visible from the perspective of the avatar matches that of the observer (e.g. disks only appear on the wall that the avatar is facing), whilst in inconsistent trials, this number differs (e.g. disks appear on both walls, whilst the avatar only has visual access to one wall). Participants then judge how many disks they can see either from their own perspective (self), or from that of the avatar (other).

The classic finding is that responses to inconsistent trials are slower when responding from a self-perspective because what the observer can see differs to what the avatar can see, despite the avatar being irrelevant to the task. This finding highlights the ease with which we can represent the world from another perspective, that this can sometimes happen spontaneously without prompting, and indicates that other people's visual perspectives automatically overlap with our own. Several other studies have demonstrated similar interference effects (e.g. Freundlieb et al., 2016; Surtees, Apperly & Sampson, 2016), whilst the original work by Sampson et al. (2010) has been replicated many times (e.g. Furlanetto, Becchio, Samson & Apperly, 2016; Nielsen, Slade, Levy & Holmes, 2015; Ramsey, Hansen, Apperly & Samson, D, 2013; Schurz, Kronbichler, Weissengruber, Surtees, Sampson & Perner, 2015).

This finding is critical for our understanding of the overlap between our own and others' visual perspectives, and supports Piaget's (1951; 1956) proposal that decentration in childhood is marked by the ability to represent multiple perspectives at the same time. This integration of other perspectives can be beneficial to joint action (Freundlieb et al., 2016) and action understanding (Creem-Regehr et al., 2013), and has also been shown to help us to imagine, and describe routes from different perspectives (Janzen, Schade, Katz & Herrmann, 2001).

In order to coordinate our actions with another person's, we need to be able to make online predictions about our partner's action intentions. One possible way that we do this is via the knowledge we possess about what the other person can see, and how their individual vantage point will afford their interaction with their environment (Creem-Regehr et al., 2013). Therefore, in joint actions, we rapidly override our own egocentric perspective (the way we see the world *through our own eyes*) and instead adopt a shared perspective that integrates both our own and our co-actor's perspectives (Freundlieb et al., 2015). Several studies, this time using more implicit measures of VPT, have shown that we are particularly sensitive to the visual perspectives of others when engaging in joint actions, and we spontaneously override our own egocentric perspective and incorporate others' perspectives. In a negative priming task, Frischen, Loach, and Tipper (2009) showed that, in a joint action scenario, participants adopted an allocentric (non-egocentric) reference frame when inhibiting distractor targets, and were found to selectively attend to targets from the perspective of their partner, demonstrating that the observation of a co-actors actions can induce a preference for their, and not one's own, visual perspective.

Further studies have demonstrated that the presence of another person when performing joint actions can cause spontaneous shifts into their visual perspective (e.g. Freundlieb et al., 2016; Freundlieb, Kovaks & Sebanz, 2018; Surtees, Apperly & Sampson, 2016; Tversky & Hard, 2009). Freundlieb et al. (2016), used a classic stimulus response task and demonstrated that acting with another agent led to spontaneous recruitment of the other's visual perspective. Typically, spatial compatibility effects occur in stimulus response (SR) tasks where participants respond to laterally presented targets using either compatible (left button-press for left target) or incompatible (left button-press for right target) response buttons. The classic finding is that responses are faster when the laterality (left/right) of the response button maps directly on to that of the observed target, whilst incompatibly mapped responses are slower (Fitts & Seeger, 1953). Freundlieb et al. (2016) demonstrated that these effects occur spontaneously in response to *vertically* presented targets (top/bottom judgements) when these same stimuli would appear laterally (left/right) from a co-actors frame of reference. Specifically, response times for top/left mapped targets were faster than bottom/left mapped targets when this target appeared to the left *of the other person*, therefore demonstrating the incorporation of the other person's perspective into participants' egocentric frame of reference.

Note, that it was necessary in these studies for the co-actor to be an active participant in the task in order for participants to adopt their visual perspective. In contrast, other studies have shown that simply suggesting to a participant that another person might act upon a set of objects, can induce a rapid shift from an egocentric to an allocentric frame of reference when judging the spatial relationships of object arrays in their presence (Tversky & Hard, 2009), indicating that knowledge about, and not observation of, a co-actor's actions may induce spontaneous VPT in some circumstances.

Further insight into the social importance of VPT comes from studies which explore the relationship between VPT ability and individual differences in social functioning, often focusing on clinical samples of individuals with social impairments such as individuals with Autism Spectrum Disorder (ASD; e.g. Hamilton, Brindley & Frith, 2009) or schizophrenia (e.g. Langdon & Coltheart, 2001). Successful human social interaction is largely attributed to the possession of a 'Theory of Mind' (ToM; Astington & Jenkins, 1995; Frith & Frith, 2005; Watson, Nixon, Wilson & Capage; 1999), which enables us to represent others' mental states (e.g. Wimmer & Perner, 1983), therefore research investigating a direct link between VPT and Theory of Mind (ToM) has focused predominantly on individuals, including children, with Autism Spectrum Disorder (ASD; e.g. Hamilton et al., 2009; Leekam, Baron Cohen, Perrett, Milders & Brown, 1997; Warreyn, Roeyers, Oelbrandt & De Groote, 2005; Zwickel, White, Coniston, Senju & Frith, 2011). ASD is primarily characterised by impairments in social cognition (American Psychiatric Association, 2013) and poor performance in ToM tasks compared to neurotypical (NT) individuals (Baron-Cohen, Leslie & Frith, 1985; Frith, 2001; Happe, 1995). It is unsurprising, therefore, that in some studies, individuals with ASD have been shown to under-perform typically developing (TD) individuals in measures of VPT (e.g. Hamilton et al., 2009; Leekam et al., 1997; Warreyn et al., 2005), especially given the evidence for the important contribution of VPT to social and cognitive development (e.g. Piaget, 1951; Flavell et al., 1981). However, given the vastly varied clinical picture of ASD (APA, 2013), and the already mentioned discrepancies in the reported age of VPT acquisition in children (e.g. Flavell et al., 1981; Piaget & Inhelder, 1956, Sodian et al., 2007), it is equally unsurprising that, on the whole, the relationship between ASD and VPT remains unclear, with some studies (e.g. Zwickel et al., 2011) showing that individuals with ASD may indeed possess the same VPT abilities as TD individuals, and some finding that verbal mental age, rather than ASD per sé, predicts poor performance in VPT tasks (Leekam et al., 1997; for review see Pearson, Ropar & Hamilton, 2013).

There is, however, some indication in the scientific literature that other psychiatric disorders may also be associated with difficulties in taking another person's perspective. Individuals with schizophrenia, whom are similarly impaired in social cognition to those with ASD, have been shown to underperform healthy individuals in a number of studies (Eack, Wojtalik, Newhill, Keshavan & Philips, 2013; Eack, Wojtalik, Keshavan, & Minshew, 2017; Langdon & Coltheart, 2001; Langdon, Coltheart, Ward & Catts, 2001). Similarly, interference effects observed in the avatar task (Sampson et al., 2010), which have been reliably replicated on several occasions (Furlanetto et al., 2016; Nielsen et al., 2015; Ramsey et al., 2013; Schurzet al., 2015) do *not* replicate with psychopaths (Drayton, Santos & Baskin-Sommers, 2018), who typically exhibit a lack of empathy for others (Decety, Chen, Harenski & Kiel, 2013; Jolliffe & Farrington, 2004) showing that VPT may be necessary for people to empathise with others.

In summary, it is clear that VPT plays a role in the development of (Flavell et al., 1981) and sustainment (Freundlieb et al., 2018; Freundlieb et al., 2016) of successful social interactions, and also that those who lack the ability to take others' perspectives are likely impaired in other areas of social cognition (Eack et al., 2013; 2017; Hamilton et al., 2009; Langdon et al., 2001). Yet, despite the abundant evidence that VPT plays an important role in social interactions, it is still unclear which mechanisms drive some of these processes.

1.2 Types of visual perspective taking.

Flavell and colleagues (1981) proposed two qualitatively different, and sequentially acquired processes in visual perspective taking (VPT): level-one VPT (VPT-1), which computes *what* another person can see, and level-two VPT (VPT-2), which calculates *how* a scene appears to another person. As in the previous examples, VPT-1 allows us to judge whether an oncoming driver can see us approaching a road, whereas VPT-2 enables us to derive how a photograph would look to another person, and adjust its position to allow them to see it clearly. VPT-2 is also the process by which we can quickly derive that an object appearing to our left would appear to the right of a person standing opposite us.

The distinction between level-one and level-two perspective taking is well supported by research which shows that very young children possess knowledge about *what* another person can or cannot see (Flavell et al., 1981; Poulin-Dubous et al., 2007; Sodian et al., 2007), but it is not until later that children are able to explicitly express *how* a scene or object would appear to another person (e.g. Borke, 1975; Hughes, 1975; Lempers et al., 1977), demonstrating that the respective levels of perspective taking are sequentially acquired in childhood, and likely require different levels of cognitive development (Flavell et al., 1981; Piaget, 1951; Piaget & Inhelder, 1956). For example, although infants as young as 14 months show evidence of representing what another person can or cannot see to better understand why they are reaching for a particular object (Sodian et al., 2007), but it is not until they reach the age of four years that they are able to accurately report how an arrangement of model mountains would appear from someone else's perspective (Borke, 1975).

Though by adulthood, most people have already acquired the more complex ability of representing *how* a scene or object appears to others, both levels are still used under different circumstances throughout adulthood, indicating that VPT-1 and VPT-2 serve different functions, and are driven by separate mechanisms. Several studies have found a clear qualitative distinction between VPT-1 and VPT-2 in adults as well as in children (e.g. Kessler & Thompson, 2010; Michelon & Zacks, 2006; Surtees, Apperly & Sampson, 2013), this time demonstrating the different functions of VPT-1 and VPT-2, rather than emphasising sequential acquisition. Michelon and Zacks (2006), for example, demonstrated that judgements about whether or not a doll could see an object (indexing VPT-1) were unaffected by the position of the doll relative to the participant, but spatial judgements about whether an object appeared to the left or the right of the doll (indexing VPT-2) took longer the further the doll was rotated away from them. The same distinction has also been observed by Kessler and colleagues (2010) in similar studies, indicating different underlying mechanisms for each level of perspective taking.

Tasks that measure VPT-1 (e.g. Furlanetto et al., 2016; Michelon & Zacks, 2006; Nielsen et al., 2015; Sampson et al., 2010) provide evidence that VPT-1 calculates what another person can or cannot see in an often implicit and automatic manner (e.g. Sampson et al., 2010). In Sampson et al.'s (2010) avatar task, for example, numerical judgements about items in a scene were slower and less accurate if the avatar in the scene could see a different number of items than the participant, even though the avatar in the scenes was completely task irrelevant. These interference effects are attributed to an involuntary overlap of self and other perspectives, though interestingly they appear to be largely reduced when participants believe that the avatar in the scene is wearing occluding goggles, and therefore cannot 'see' the disks on the wall (Furlanetto et al., 2016).

The mechanisms that establish whether someone sees an object (or not) seem to reflect relatively straightforward line-of-sight computations (e.g. Baker, Levin, & Saylor, 2016; Michelon & Zacks, 2006; Surtees et al., 2013). Michelon and Zacks (2006) demonstrated that VPT-1 is derived from following another person's line of sight by manipulating the distance between the doll and the target object. Larger distances were reflected by longer response latencies, whilst objects positioned close to the doll were judged more rapidly. In contrast, the orientation of the doll's position, relative to the participant, did not influence response times, suggesting a more straight forward mechanism for VPT-1 than VPT-2, which is known to be influenced by the degree to which another agent is rotated away from us (e.g. Kessler & Rutherford, 2010; Kessler & Thompson, 2010; Kozhevnikov, Motes, Rasch & Blajenkova, 2006).

VPT-1 calculation therefore depends upon following another person's line of sight towards an object or scene, and when interrupted with opaque goggles, or an occluding screen, participants rapidly compute that the person cannot see the object, and do not experience interference from their perspective.

Unlike VPT-1, which serves a seemingly basic function, more sophisticated VPT-2 can be broken down further into spatial, or '*where',* judgements (e.g. Freundlieb et al., 2015; Zwickel & Muller, 2010; Tversky & Hard, 2009) and visual, or '*how'*, judgements (e.g. Freundlieb et al., 2018; Surtees et al., 2016). Spatial VPT-2, often referred to as 'spatial perspective taking' (SPT) refers to judgements made about spatial properties of objects. For instance in explicit tasks, participants may judge whether an object appears to the left or the right of another agent (e.g. Kessler & Rutherford, 2010; Kessler & Thompson, 2010; Michelon & Zacks, 2006), or in implicit tasks, people may be asked to make such judgements from their own perspective (e.g. whether an object appears to the left or right) in the presence of another person that would make these judgments differently(e.g. Freundlieb et al., 2016; Tversky & Hard, 2009; Zwickel & Muller, 2010).

One common manipulation used in explicit SPT tasks is to vary the angular distance between the participant and another agent and ask participants to make judgements about where objects are positioned from the perspective of the other person (e.g. Kessler & Rutherford, 2010; Kessler & Thompson, 2010; Kessler & Wang, 2012; Koshevnikov et al., 2006). Kessler and Thompson (2010) demonstrated that people find it harder to make spatial judgements from another person’s perspective the further they are rotated away from them. Participants were shown pictures of an avatar sitting facing a round table, the position of whom was varied between trials. Sometimes the avatar was positioned very close to the observer, and at other times the avatar was positioned further away. During the trials, a flower and a gun appeared on the table in front of the avatar, and participants were asked to make a left/right judgement about the location of one of the objects from the avatar’s viewpoint. Unlike in VPT-1 tasks, which have shown that angular disparity has no influence on people’s ability to judge whether or not another agent can see an object (e.g. Michelon & Zacks, 2006), for SPT, Kessler and Thompson (2010) found the distance between the participant and the avatar reliably predicted response duration. When the avatar was positioned further away, response times became much longer compared to when the avatar was closer to the participant. These, and other similar findings (e.g. Kessler & Rutherford, 2001), demonstrate a clear linear relationship between time taken to compute how a scene looks to another person and where this other person is positioned relative to oneself. One well supported explanation for this is that people mentally rotate themselves into the position of another person. The further a person has to rotate, the longer this transformation takes (Kessler & Rutherford, 2010; Kessler & Thompson, 2010; Kessler & Wang, 2012; Koshevnikov et al., 2006).

Despite the relatively sophisticated mental travel that it requires, implicit measures of VPT-2, or SPT, (e.g. Freundlieb et al., 2016; Tversky & Hard, 2009; Zwickel & Muller, 2010) have demonstrated a similar overlap of perspectives to that shown in VPT-1 tasks (e.g. Sampson et al., 2010). For example, Zwickel and Muller (2010) used a simple spatial judgement task to demonstrate interference effects of conflicting perspectives. In their task, participants were shown images of a face or a grey rectangle, and white dots were presented either laterally or vertically around the object. The task was simply to judge the location of the dot when it appeared on the screen. Left/right judgements were slower in the presence of a face compared to a rectangle, whilst vertical judgements were unaffected by the central stimulus. In this task, a dot presented to the right of a participant would appear to the left of the face on the screen, and this inconsistency in spatial reference frames was found to slow down responses, indicating an overlap between the self and other perspectives.

Visual VPT-2 judgments, rather than spatial judgements, reflect the visual properties of objects or scenes, for example, whether a number might appear differently from a non-egocentric perspective. Indeed, Surtees et al. (2016) recorded participants' recognition times for numbers, and found that judgements were slowed down when the same number would appear differently to another person positioned opposite the participant (e.g. "6" vs. "9"), whilst recognition judgements about numbers which had the same appearance from both perspectives (e.g. “8”) were not affected by the presence of another person. Similar interference effects were reported in a recent study by Freundlieb et al. (2018), who found that semantic word categorisation judgements were affected by the presence of another person. In this task, participants sat at 90 degrees (left or right) to a confederate and were asked to categorise a series of words presented vertically to the participant, and either upright or inverted to the confederate dependent upon where this person was sitting. Participants were particularly slow to make judgements about words that appeared upside down to the confederate, indicating that the participants’ knowledge of the relative difficulty of the task when responding from the confederate’s position interfered with their *actual* perspective, and slowed down judgements.

Such findings have been influential to our understanding of VPT, primarily informing the qualitative distinction between level-one and level-two, and the relative sophistication of each level of perspective calculation. Overall, the evidence clearly demonstrates sequential acquisition of these abilities in childhood (VPT-1 followed by VPT-2), and shows that each level supports a different function (‘what’ vs. ‘how/where’). We also know that both level-one and level-2 perspectives are represented similarly to our own, leading to interference effects in both VPT-1 (e.g. Nielsen et al., 2015; Sampson et al., 2010) and VPT-2 (e.g. Freundlieb et al., 2016; Freundlieb et al., 2018; Surtees et al., 2016) tasks, however what is less clear is which mechanisms underlie these types of perspective taking. Whilst there is evidence for a line-of-sight explanation for VPT-1 (e.g. Furlanetto et al., 2016), the mechanisms underlying VPT-2 are less well understood, and therefore require further study.

1.3 Mechanisms underlying VPT-1

Level-one VPT computes what another person can or cannot see, and this information is derived from following another person's **line of sight** in the direction of a target object (Furlanetto et al., 2016; Kessler & Rutherford, 2010; Michelon & Zacks, 2006; Sampson et al., 2010; Surtees et al., 2013). Several studies have also suggested that VPT-1 is an **automatic** process and it is widely accepted that humans can compute what others can see without cognitive effort or control (Bukowski, Hietanen & Sampson, 2015; Ramsey, Hansen, Apperly & Sampson, 2013; Sampson et al., 2010). This is indeed supported by work that demonstrates VPT-1 in very young children (e.g. Flavell et al., 1981; Poulin-Dubous et al., 2007; Sodian et al., 2007), and also in primates (e.g. Hare, Call, Agnetta & Tomasello, 2000), which demonstrate that very little cognitive control is required for this process.

There has, however, been some suggestion in the literature that implicit measures of VPT-1 may be better explained in terms of gaze cuing (e.g. Gardner, Hull & Taylor, 2018), or domain general attention orienting effects (Santiesteban, Catmur, Hopkins, Bird & Heyes, 2014). Similarities between the effects observed in implicit VPT-1 tasks (Sampson et al., 2010) and those seen in studies employing a gaze cuing paradigm (e.g. Driver, Davis, Ricciardelli, Kidd, Maxwell & Baron-Cohen, 1999; Friesen & Kingstone, 1998) have raised questions about whether VPT-1 and gaze cuing are distinct processes, and whether we automatically derive *what* other people can see, rather than simply reflexively orient our attention to *where* another person is looking.

In gaze cuing tasks, the typical finding is that the looking direction of another person causes the observer to orient their attention to look in the same direction, leading to faster identification of targets that are presented in the line of sight of this other person (e.g. Driver et al., 1999). The well-known finding is that these gaze congruency effects typically do not occur with an SOA of less than 300ms (Freisen & Kingstone, 1998; Friesen, Moore & Kingstone, 2005; Xu, Tanaka & Mineault, 2012). Bukowski et al. (2015) therefore manipulated stimulus onset asynchrony (SOA) in a variant of Sampson et al.'s (2010) task combining the classic avatar task stimuli with a simple gaze cuing paradigm (e.g. Posner, 1980) to test whether this experiment design would induce gaze cuing effects at SOA < 300ms, owing perhaps to its relative salience compared with classic cuing paradigms (Treisman & Gelade, 1980).

Participants were shown a scene identical to that used in Sampson et al.'s (2010) task. Participants were asked to validate the number of coloured disks that appeared on the wall, either within the line of sight of the avatar (congruent) or on the wall to which the avatar's back was turned (incongruent). In half of all trials, the disks were presented at the same time as the avatar (SOA = 0ms) and in the other half, disk onset was delayed by 300ms relative to the appearance of the avatar (SOA = 300ms). Response times were fastest in congruent trials following a 300ms delay, replicating prior gaze congruency thresholds (Freisen & Kingstone, 1998; Friesen et al., 2005; Xu et al., 2012), however, these gaze congruence effects could not be systematically found when the avatar and the disks appeared at the same time. The exception, however was that congruency effects did occur when participants were prompted to take the perspective of the avatar when performing the task. Whilst these results may indicate that VPT-1 does not fully satisfy the third automaticity criterion of being stimulus driven (Moors et al., 2006), this study does validate the avatar task (Sampson et al., 2010) as a measure of implicit VPT-1, rather than one that indexes attention orienting. Interestingly, Gardner, Bileviviute and Edmonds

(2018), whom in one paper (Gardner et al., 2018) supported an attention orienting view of VPT, found similar evidence for a distinction between gaze congruency and VPT-1 in another variant of the avatar task (Sampson et al., 2010), lending further support to Bukowski et al.'s (2015) findings. Taken together, these findings certainly point to spontaneous, if not automatic, computation of what another person can see, and cannot be fully explained by attentional cuing.

A further assumption is that VPT-1 is an intrinsically **social** process, and that we are more susceptible to an overlap in perspectives in the presence of another seeing agent (Sampson et al., 2010), compared to non-seeing agents (Furlanetto et al., 2016) or objects (Nielsen et al., 2015). VPT-1 is indeed believed to tap into mentalizing abilities (e.g. Hamilton et al., 2009; Schurz, Kronbichler, Weissengruber, Surtees, Samson & Perner, 2015) which allow us to represent others' beliefs, goals, desires, and crucially, knowledge. We therefore do not automatically adopt the perspective of objects because they do not possess a mental state. Bukowski et al. (2015), for example, provided strong evidence for a mentalising account of VPT, where it was only when representing the mental state of the avatar that gaze congruency effects occurred with a short SOA.

This, again, has come under scrutiny (e.g. Santiesteban et al., 2014; Cole & Millet, 2019), but the evidence for VPT -1 as a social process is extensive (see **1.1** for overview). In response to claims that VPT-1 effects occur in the presence of other objects, such as an arrow (e.g. Santiesteban et al., 2014), Nielsen et al. (2015) substituted the avatar in Sampson et al.'s (2010) task for either an arrow, or a coloured box and asked participants to respond from a 'self' perspective, or an 'other' perspective (i.e. respond from the perspective of the avatar, the arrow, or the coloured box). They found that participants were most likely to experience altercentric interference effects in the presence of a human avatar when responding from their own perspective, whilst the smallest effects were seen in the coloured box condition. Conversely, in ‘other’ trials, the largest interference of egocentric perspective was observed, showing that participants generally took longer to respond from the ‘other’ perspective, due to their own conflicting perspective, whilst in human trials this egocentric interference was much smaller. Self-reported measures of perspective taking and empathic concern (Interpersonal reactivity index; IRI, Davis 1980) were found to correlate only with interference from a human avatar, and not from the non-social objects. This study therefore demonstrates that interference of another’s perspective occurs more for social than non-social stimuli, and also by extension that VPT-1 is stimulus driven, and might therefore satisfy the criteria for automaticity (Moors et al., 2006).

To summarise VPT-1 is, to our knowledge, automatically derived in the presence of a seeing agent. By following another person’s line of sight, humans rapidly calculate what another person can or cannot see, and do so effortlessly, and without intent. What is less clear is *how* the intrusion of others’ perspectives is mentally represented, and whether the reported interference occurs on a conceptual, perceptual or epistemic level.

1.4 Mechanisms underlying VPT-2

Whilst we have a relatively clear understanding of the cues and mechanisms that drive VPT-1, the mechanisms that drive VPT-2 are less well understood. It is, however, widely assumed that VPT-2 is a cognitively effortful and controlled process (Flavell et al., 2018), unlike its level one counterpart.

One explanation of the mechanisms underlying allocentric perspective calculations is that we (mentally) reposition ourselves in space when we probe the visual perspective of another person. This type of transformation is often described as a 'self-rotation' (e.g. Kessler et al., 2010; 2010), and refers to an egocentric perspective transformation (Kozhevnikov et al., 2006) where we imagine ourselves to rotate around a central point in a scene, whilst the scene itself remains static. Related to, but qualitatively different from this are object-based transformations, where instead of imagining ourselves in a different position in space, we imagine the scene we are judging to rotate to align with our own viewing perspective (Kozhevnikov et al., 2006). To solve an explicit perspective taking task (e.g. the Three Mountains Task; Piaget & Inhelder, 1953), it is clear that either one of these strategies could be employed, but distinguishing between which transformation might happen in an implicit task (e.g. the '6' vs. '9' distinction; Surtees et al., 2016) is much harder and still actively debated (e.g. Muto, Matsushita & Morikawa, 2018).

In explicit measures of VPT-2, several studies have demonstrated a link between the angular distance from the observer to the target other person, in each case showing a reliable linear increase in time taken to make left/right judgements from the agent's perspective with increasing angular rotation away from the observer (e.g. Kessler & Rutherford, 2010; Kessler & Thompson, 2010; Kessler & Wang, 2012). One explanation is that to represent how a scene looks to another person, we must first mentally rotate our position in space into that of the target 'other', taking longer the further we have to (mentally) travel. This explanation fits with an embodied account of social cognition (e.g. Glenberg, 2010; Proffitt, 2006; Tversky & Hard, 2009), which proposes that our real or imagined actions facilitate cognitive processes, in the present case, taking the perspective of another person. This explains why, for example, the match or mismatch of our posture with that of the target person might interfere with, or improve performance in VPT-2 tasks (e.g. Kessler & Rutherford, 2010; Kessler & Thompson, 2010). Kessler and colleagues (2010; 2010) have demonstrated across a number of tasks that if an observer adopts a posture that matches that of the target 'other', then responses are faster compared to if they posture is misaligned, because the mental transformation from actual to imagined position becomes more effortful. However, in implicit VPT-2 tasks, such as Surtees et al.'s (2016) number identification task, this explanation does not hold, or at least it does not seem an efficient strategy for an individual to use if it is indeed a controlled cognitive process. People may indeed mentally rotate to the position of the person standing opposite them when making VPT-2 judgements in their presence, and as such find it more difficult than usual if, for instance, a "6" would appear as a "9" this other person. However, if such mental transformations are controlled, it is difficult to interpret *why* participants would rotate into the position of another person when (1) the other person is task irrelevant, and (2) doing so impedes their performance in the task. Similarly, in Freundlieb et al.'s (2016) spatial compatibility task, where participants appear to spontaneously recruit the perspective of their co-actor, it is not clear *why* this would happen voluntarily if the task demands do not require effortful computation of the other's perspective.

It is also unclear which social cues drive VPT-2. For example, we know that for VPT-1 calculations, a person's visual access to an object is determined by following their line of sight however, the few studies that have directly tested this for VPT-2 (Freundlieb, Sebanz & Kovaks, 2017; Mazzarella, Hamilton, Troiani, Mastromauro & Conson, 2012; Furlanetto, Cavallo, Manera, Tversky & Becchio, 2013) have delivered inconsistent results. For example, Freundlieb et al. (2017) found that a co-actor's visual access to a stimulus mediates participants' level-two computations of their visual perspective. Using the same experimental design as in their prior work (Freundleib et al., 2016), participants responded with either a left or a right button to vertically presented stimuli. These same stimuli appeared laterally from the perspective of a co-actor who was positioned at 90 degrees to the left or the right of the participant. As in prior work, the data showed spatial compatibility effects as though participants were responding from the other person's perspective, however the same effects were not present when this other person wore occluding glasses and could therefore not see the target stimuli. This may indicate that VPT-2 is computed secondarily to VPT-1 (i.e. 'what' before 'how'), or, like VPT-1, it could be that VPT-2 is similarly dependent upon line of sight following.

However, in contrast, Mazzarella et al. (2012) found that merely seeing another person look towards an object did not induce a detectable shift to their perspective. It was only through observing the other person reach for the object that caused interference from their spatial perspective. Similarly, Furlanetto et al. (2013) showed that while viewing someone gaze at an object numerically increased shifts to their spatial perspective, this did not seem to be statistically robust, and even removing gaze cues altogether (by blurring the eye region) did not affect whether people judged objects from another's or their own perspective. It is therefore unclear whether eye-gaze towards the target object drives VPT-2.   Again, with action, the picture is very similar. Whilst in Freundlieb et al.'s (2016) spatial compatibility task shifts into another's perspective only occurred if this other person was actively engaging in the task, Tversky and Hard (2009) found that people were more likely to respond from another's perspective simply because they were present when asked to make judgements about the spatial location of objects. This raises further questions about what drives VPT-2 calculations.

In sum, the specific mechanisms that drive VPT-2 are poorly understood. Whilst the evidence for an embodied account of VPT-2 (e.g. Kessler et al., 2010) is very clear, such studies employing the methods used here rely on participants' explicit allocentric judgements, therefore in the case of implicit VPT-2, there may be an alternative explanation. In particular, we need to question the long-held belief is that VPT-2 is a cognitively demanding, controlled process. Primarily, this does not fit with the results in implicit tasks (Surtees et al., 2016; Freundlieb et al., 2016; Freundlieb et al., 2018), which show seemingly *unintentional* VPT-2, which happens at the detriment to participants' performance in these tasks. Secondly, the notion that VPT-2 is an effortful process does not fit with the numerous examples of daily social interactions in which people benefit from, it seems unknowingly, taking another's perspective.

1.5 VPT as perceptual simulation

One pressing issue is to better understand exactly *how* we represent what others can see, and how they see it. While it is well supported *that* people can – often spontaneously – represent another's perspective, the mechanisms that support it are unclear. Several studies have suggested that we represent the content of others' visual perspectives in a similar way to our own (e.g. Furlanetto et al., 2016; Freundlieb et al., 2016; Freundlieb et al., 2017; Kampis et al., 2015; Sampson et al., 2010; Surtees et al., 2016). Surtees et al. (2016), for example, demonstrated that judgements about objects from one's own perspective are harder if this same object appears differently to somebody else (e.g. a “6” would appear as a “9” to a person standing opposite oneself). Similarly, Sampson et al.’s (2010) avatar task demonstrates that making judgements about a number of objects visible in a scene is more difficult when another person would judge this differently from their own perspective. Other studies have provided evidence for similar intrusion of others’ perspectives, not only when making judgements about *what* one can see (Sampson et al., 2010) but also *how* one sees the world (Surtees et al., 2016), or *where* objects appear relative to oneself (Freundlieb et al., 2016).

This interference suggests that, on some level, own and others’ views of an item are represented similarly. However, to our knowledge, it has not yet been shown what form this ‘overlap’ between multiple perspectives takes. It is possible, for example, that it takes longer to judge a “6” as a “6” when it looks like a “9” to someone else (Surtees et al., 2016) simply because our judgments become more uncertain. In other words, we may generally *know* that what they see is different to us, without actually representing what specifically the other person is seeing. In other words, the slowing of responses when own and others’ views on an item mismatch (e.g. Sampson et al., 2010, Surtees et al., 2016) could simply reflect uncertainty that results from this uncertainty, but would not indicate that participants are really representing the content of the other person’s visual perspective. Another possibility is that the slowing indexes that people are simply aware that the other person would *respond* to the stimulus using a different button to oneself. Indeed, this could explain why these effects are larger – or only observed – in joint action situations, or when the other person makes any action at all (e.g. Freundlieb et al., 2016).

The strongest explanation, which is implied in many descriptions of perspective taking, but never directly tested, is that the representation of others’ visual input takes a (quasi)-perceptual form, and that VPT ‘paints’ an image of the content of others’ perspectives onto our own perceptual system. In the above experiments, judging a “6” as a “6” would therefore become more difficult because the visual representation of the others’ view – the shape “9” – intrudes into our own perceptual processes, making it harder to recognize it as it really is. Indeed, all of the findings described above, from slowed down judgements about the magnitude of a number (Surtees et al., 2016), to judgements about the number of objects visible in a scene (Sampson et al., 2010) could be explained by a framework of perceptual intrusions of the others’ viewpoints into our own perception.

Confirming such a perceptual representation of others’ viewpoints would solve several problems with our current understanding on how others’ perspectives are represented and how they could drive behaviour in social interactions. If others’ views on the world would be represented in a (quasi-) perceptual manner, they could drive processes such as perceptual decision making and action control in the same bottom-up manner as our own input (Kampis et al., 2015; Roelfsema & de Lange, 2016), perhaps explaining the link to joint action (Creem-Regehr et al., 2013; Freundlieb et al., 2016). A (quasi-) perceptual representation would also provide direct access to others’ view the world, and how their view differs from ours, opening up the possibilities to link visual perspective taking to more sophisticated mentalizing abilities like Theory of Mind and false belief understanding in particular (Hamilton et al., 2009; Schurz et al., 2015).

On a theoretical level, such a perceptual account would link perspective taking to prior research on mental imagery. Mental imagery has been proposed to serve as a mechanism by which we re-live – or mentally manipulate – a seen object or event by way of a simulation, or ‘emulation’ (Moulton & Kosslyn, 2009). Indeed, the role of mental imagery in perceptual decision-making is well established (e.g. Shepard & Metzler, 1971). It is argued that visual simulations are pictographic, visual memories, and that their retrieval and reliving operates using the same perceptual machinery as ‘real’ visual perception (e.g. Ganis, Thompson & Kosslyn, 2004), which allow us to make specific predictions by drawing on past experiences. Mental imagery is therefore a very powerful tool, allowing us to predict action outcomes, evaluate emotional stimuli, plan the best route home, and even empathise with others.

It is noteworthy that VPT is thought to serve similar roles in social cognition (Creem-Regehr et al., 2013; Flavell et al., 1981; Hamilton et al., 2009; Piaget, 1951), and that similar brain regions, commonly known as the ‘core network’ (Buckner, Andrews-Hanna & Schacter, 2008) have been implicated in tasks measuring visual mental imagery and VPT, and also ToM (Buckner & Carroll, 2007; Hassabis & Maguire, 2007). Other studies have shown a clear overlap in neural machinery for mental imagery and visual perception (e.g. Ganis et al., 2004; Kosslyn, Thompson & Alpert, 1997), demonstrating that visual mental imagery elicits perception-like patterns of low-level processing in the primary visual cortex, indicating that visual simulations happen at the perceptual level.

Another key characteristic of mental imagery is that visual mental simulations happen sequentially (e.g. Fisher, 2006; Moulton & Kosslyn, 2009). For instance, the retrieval of one image from memory can trigger a series of further images (e.g. the unfolding of an action), or affective responses (e.g. by triggering the simulation of a feeling of joy at the birth of a child, or grief at the death of a loved one; Holmes, Mathews, Mackintosh & Dalgleish, 2008), all of which contribute to our ability to make specific event-related predictions (Moulton & Kosslyn, 2009). Shepard and Metzler’s (1971) seminal work which investigated mental rotation of objects contributed substantially to our current understanding of the sequential nature of such mental simulations. They presented participants with pairs of three-dimensional block objects at differing orientations and asked them to judge whether both were the same, or whether one object was the mirror-inverted version of the other. They found that the time taken to judge whether a rotated object was the same as its upright counterpart increased linearly with the increasing angular difference between the two objects. Since then, numerous studies have replicated this pattern with alphanumeric characters (Rusiak, Lachman, Jaskowski & van Leeuwen, 2007), abstract shapes (Shepard & Metzler, 1988), and body parts, such as hands and feet (Ionta, Fourkas, Fiorio & Aglioti, 2007).

These studies demonstrate that in order to solve MR problems, we simulate intermediate representations of objects at different orientations, with each intermediate simulation serving as the basis for the next, until we arrive at a point where we can accurately judge the characteristics of the object we are observing. Simulation therefore serves as an 'epistemic device' (Moulton & Kosslyn, 2009; p1276) that operates sequentially and can form the basis of perceptual decision making. Put simply, we perceptually represent the rotating object continuously until it appears upright on our 'mind's eye' (Farah, 1989), thereby simulating the experience of actually observing the object rotating in real life. As with a real object rotation, the further the object needs to be rotated back to upright, the longer this takes, as is reflected in the classic MR response pattern where response times increase linearly as a function of the object's increasing angular disparity away from upright (Ionta et al., 2007; Rusiak et al., 2007; Shepard & Metzler, 1971; Shepard & Metzler, 1988).

It is possible, then, that VPT allows such simulations to occur not from one's own perspective, but from that of the other person. For example, when judging the magnitude of the number “6” when standing opposite somebody who views it as a “9”, it is possible that the first representation arises from this other person’s perspective, after which this intrusion must be corrected to upright from one’s own perspective. If we spontaneously simulate the content of another person’s perspective, this mental image may override our own perceptual input, leading to the necessity to correct the mental image prior to making a judgement, which would result in the observed longer response times in Surtees et al.’s (2016) study.

As with mental imagery (Fisher, 2006; Moulton & Kosslyn, 2009) and perceptual decision making (e.g. Shepard & Metzler, 1971), other studies show that a person’s perceptual experiences, for example when observing an action, can be distorted by top-down influences (e.g. Hudson, McDonough, Edwards & Bach, 2018; McDonough, Hudson & Bach, 2019), and that our judgements can be biased by prior knowledge about a stimulus (e.g. Bruner & Goodman, 1947; Summerfield & de Lange, 2014). That what people perceive, or ‘see’ is biased by their prior expectations about an action or event is accumulating increasing support in the scientific literature. Indeed, several behavioural studies have demonstrated that social perception itself is an inferential process (e.g. Bruner & Goodman, 1947; Hudson, McDonough, Edwards & Bach, 2018; McDonough, Hudson & Bach, 2019), and what we ‘see’ is often not a veridical representation of what ‘is’. Moreover, fMRI evidence (e.g. Kok, Brouwer, van Gerven & de Lange 2013) has shown that high-level processes, such as our prior knowledge or expectations, can influence very low-level perceptual processes in their earliest stages, leading to distortions in the way we perceive an event. Therefore when another’s simulated perspective provides an entry point to our perceptual decision making (such as in a MR task; e.g. Shepard & Metzler, 1971), what we perceive may indeed reflect our knowledge about how the world looks to another person, rather than what *we* see ourselves, causing us to spontaneously make judgements from this allocentric perspective.

This view would support the notion of a shared representational space for own and others' visual perspectives (e.g. Jolicoeur & Cavanagh, 1992; Kampis et al., 2015; Piaget, 1951), Indeed, such an overlap for own and others' perspectives has been observed in infants (Kampis et al., 2015). Kampis et al. (2015) recorded infants' event-related EEG activity while they observed scenes showing an actor and an object to test whether they used their own representational system when tracking the actor's knowledge, beliefs, and visual perspective. As in other studies (e.g. Kaufman, Csibra & Johnson, 2005; Kaufman, Csibra & Johnson, 2003), the authors found gamma-band oscillatory activity for sustained object representation when objects were occluded from the infants' sight. Crucially, the same activations were found when the object was occluded from the actor's perspective, despite the infants having continued visual access to the object. The infants therefore derived that the actor could not see the object for a time, and held its representation in mind *on behalf of the actor,* indicating a shared representational space for own and others' object representations. This may indeed help to explain the overlap in perspectives that is thought to drive interference effects in implicit measures of VPT (Freundlieb et al., 2016; Sampson et al., 2010; Surtees et al., 2016) and may also explain its link to sophisticated mentalizing abilities (Furlanetto et al., 2018; Hamilton et al., 2009; Kampis et al., 2015; Piaget, 1951).

When we take another's visual perspective, the visual system is perhaps, therefore, not simply a space for processing direct visual input, but is a 'blackboard' (Roelfsema & de Lange, 2016) upon which perceptual experiences are constructed via the integration of own and (imagined) others' visual input, and our own prior expectations (e.g. Kok et al., 2013). These (quasi)perceptual experiences may then drive all processes that operate on perceptual input (Roelfsema & de Lange, 2016), facilitating perceptual decision making, but also emotional responses (Holmes et al., 2008) and event-specific predictions (Moulton & Kosslyn, 2009)

The primary aim of this thesis is to directly test whether VPT takes the form of a perceptual simulation, and whether this simulated perceptual input can facilitate faster perceptual decision making than one's own visual input. It rests on one central prediction that follows from perceptual accounts of perspective taking: if we can virtually "see" through another's eyes, then this perspective should not only interfere with own judgments (i.e. the slowing down of responses due to conflicting 'self' and 'other' perspectives; Sampson et al., 2010; Surtees et al., 2016), but would also facilitate faster responding in tasks where another agent might (in principle) have clearer visual access to a scene than oneself.

To our knowledge, no study has directly tested this prediction. Several studies show that different perspectives affect judgment of how many objects one sees (Sampson et al., 2010), where an object is positioned relative to oneself (Tversky & Hard, 2009), or how a number looks to oneself (Surtees et al., 2016). However, none of these studies were able to disentangle whether these intrusions from another's perspective to only interfere with perceptual decision making, or whether they can also be beneficial, specifically in cases where we could rely on another's view of an item that would be difficult to judge from one's own.

1.6 Thesis overview

The primary aim of this thesis is to investigate whether VPT-2 takes the form of a perceptual simulation, and whether these simulated perspectives can facilitate, rather than only interfere with perceptual decision making. All experiments presented in this thesis will employ a modified mental rotation experiment, similar to that used by Shepard & Metzler (1971), but with the critical manipulation that at times during the experiment the participants will also see another person, or non-social (control) object, at the same time as the object they are judging. These experiments will measure whether the incidental insertion of another person (or object) will disrupt the known mental rotation pattern, where responses become progressively longer with objects' increasing angular disparity away from upright (e.g. Shepard & Metzler, 1971).

In each experiment, participants will be presented with a series of images showing a table, upon which alphanumeric characters will appear in either in their canonical or mirror-inverted form. As in similar studies (e.g. Ionta et al., 2007; Rusiak et al., 2007; Shepard & Metzler, 1971; Shepard & Metzler, 1988), these items will be presented at varying degrees of rotation away from upright to the participant. In some scenes another person or object will be present, appearing at 90 degrees to the left or the right of the participant, facing the table. In others, only the table and object will be visible to the participant. Participants will simply be asked to judge whether the character is canonical or mirror inverted (e.g. "R" vs. "Я"). We will then compare whether the classic response time patterns are disturbed by the presence of another person present in the scene.

The experiments presented in chapters 2, 3 and 4 will use this method to directly test (1) whether we spontaneously adopt another person's visual perspective during perceptual decision making, (2) whether others' visual perspectives can facilitate faster recognition of items, and (3) whether VPT takes the form of a perceptual simulation. Secondary to this, further experiments will begin to unpick the cues that drive VPT by substituting the inserted person with an object (chapter 3), and manipulating the person's gaze direction (chapter 4).

To support this work, we draw on prior evidence which demonstrates that mental imagery likens to the reliving of an event by way of a mental simulation (Moulton & Kosslyn, 2009). Several studies have demonstrated the perceptual nature of mental imagery (e.g. Ganis et al., 2004), which is known to be employed when we mentally rotate an object (e.g. Shepard & Metzler, 1971), providing us with an excellent tool to test whether VPT operates using similar perceptual mechanisms. By simply comparing the mental rotation pattern of objects in the presence of another person vs. when they are not there, any disturbances in the response time pattern might then be explained by a perceptual simulation framework, where intrusions of the other person's perspective interfere with our own perceptual processes.

The central prediction here is that if we perceptually simulate the content of the other person's perspective, judgements about items oriented towards this person, and therefore (in principle) easier to judge from their perspective than our own, will be faster compared to if the characters are facing away from the other person or if this person is not present. Equally, if VPT does take the form of a perceptual simulation, we also expect to see some interference when judgements from this other person's perspective would be harder than our own, that is, if a character is inverted from their perspective. Crucially, if mental rotation response times are governed by both one's own and the others' view on an item, then response times for the *whole* mental rotation (i.e. across angular disparities) will be predicted by not only the object's orientation relative to the participant, but also its orientation relative to the other person. Indeed, such an effect would demonstrate that VPT takes a perceptual form, and that intrusions from another's visual perspective can be facilitatory as well as being a hindrance to task performance (e.g. Freundlieb et al., 2018; Sampson et al., 2010; Surtees et al., 2016), and that they can serve as input to mental rotation processes.

Chapter 2 of this thesis will first test whether we see such intrusions, and indeed whether VPT does take the form of a perceptual simulation. By using an implicit task, where the individuals inserted into the scenes are task-irrelevant, we will also observe whether any such intrusions occur spontaneously. This is important as VPT-2 is widely assumed to be a controlled and cognitively effortful process (e.g. Flavell et al 1981; Kessler et al, 2010). However, if participants experience involuntary intrusions from the other's perspective, this may indicate that VPT-2 is more spontaneous than previously thought. A further replication of the same study is also presented in this chapter.

Secondly, in this chapter we will begin to explore the cues that drive these effects. We therefore present a second experiment that addresses whether VPT intrusions occur in the presence of a 'mindless' object instead of a person. Several studies (e.g. Furlanetto et al., 2016; Nielsen et al., 2015) have shown that only humans, but not objects, can induce spontaneous VPT, supporting the link to social ability and 'mindreading' (Furlanetto et al., 2016). Indeed, in the presence of a person who possesses their own mind, and therefore thoughts, beliefs, knowledge and so on, we may simulate their visual perspective on order to probe their epistemic stance. In the case of an object that lacks such human qualities, we are less likely to attempt to understand how the world looks to it. To test this directly, we manipulate between-subjects whether participants see a person or a lamp appear to the left or the right of the table, and compare the effects between groups.

Finally in this chapter we present a third experiment, which will test this paradigm's sensitivity to different instructions, specifically by manipulating between-subjects whether we instruct participants to take the other person's perspective or not. If consciously taking another's perspective improves task performance in general, this will be reflected in an *overall* acceleration in response times, whereas if our paradigm does truly index spontaneous VPT, then we will see both increased facilitation and interference effects, simply reflecting a heightened version of intrusions recorded without perspective cuing in earlier experiments.

Whilst chapter 2 of this thesis addresses the theoretical proposal that VPT-2 takes the form of a perceptual simulation, subsequent chapters aim to test more basic assumptions that have been discussed in the earlier literature, such as which social cues drive VPT-2, or whether VPT-2 is indeed supported by mental self-rotation. **Chapter 3** therefore tests whether VPT-2 is driven by where another person is looking. People are known to derive *what* another person can see by following their line of sight (Furlanetto et al., 2016; Kessler & Rutherford, 2010; Michelon & Zacks, 2006; Sampson et al., 2010; Surtees et al., 2013) indicating a central role of gaze direction in VPT-1 calculations. However, few studies have directly tested this for VPT-2 (Furlanetto et al., 2013; Freundlieb et al., 2017; Mazarella et al., 2012), and the results from these studies have largely been inconclusive. For example one study indicates that a person's gaze towards an object plays a critical role in driving VPT-2 calculations (Freundlieb et al., 2017), whereas others conclude that gaze is only relevant in conjunction with cues to action intention (Tversky & Hard, 2009), and the data from a third indicate no contribution of eye-gaze in driving VPT-2 (Mazarella et al., 2012).

The two experiments outlined in this chapter use the same experimental paradigm as the experiments outlined in chapter 2, but for one simple manipulation. In the first experiment we vary, between-trials, whether the inserted person in the scenes is looking at the object on the table, or looking away from the table towards the corner of the room. In the second experiment we vary looking directions between-participants, thereby controlling for carryover effects (MacFie, Bratchell, Greenhoff & Vallis, 1989). This experiment also includes a third 'outward' gaze condition, in which the inserted individual looks away from the table and out towards the participant. This type of mutual gaze is known to be particularly special, in that it serves as an alerting signal, captures attention, and eliminates gaze cuing altogether (e.g. Mazzarella et al. 2012; for a review, see Hamilton, 2016). We therefore include this manipulation to (1) rule out gaze cuing as an explanation of effects observed in earlier versions of the task, and (2) to test whether the elimination of gaze-following disrupts the already reported response pattern.

Finally, **Chapter 4** will test whether VPT-2 is an embodied process, whereby we motorically simulate a whole body spatial transformation into the position of the other person in a scene. Several studies have shown that the time taken to make an explicit spatial judgement from another's perspective (e.g. "where is the flower relative to the gun") increases as a function of increasing angular distance between the observer and the other person (Kessler & Thompson, 2010; Kessler & Rutherford, 2010; Kozhevnikov et al., 2006; Surtees et al., 2013). Such studies have concluded that these judgements rely on an effortful process by which we imagine ourselves in a different position in space, and do so by way of mental self-rotation.

What is not clear is whether the same mechanism is involved in implicit VPT-2 computations. Here we draw on research that shows that embodied cognitive processes, such as emotion recognition or mental rotation can be interfered with if a person is unable to physically perform a simulated action (e.g. Cook et al., 2012; Neal & Chartrand, 2011; Parsons, 1994). To test whether VPT-2 operates using a similar mechanism, in this experiment we restrict participants' movement during half of the experiment trials. This experiment also considers whether participants' movement plays a causal role in the deployment of VPT-2 processes (e.g. Friston, 2010), or whether this movement may represent epiphenomenal 'leakage' of simulated self-rotation in space (Colton et al., 2018; Jacobson, 1930).

'Note that throughout the experiments outlined in this thesis, Bayes Factors are reported to quantify how well H_1 predicts the observed data relative to H_0 (BF_{10}). BFs are described using the common terminology applied to Bayes Factors (Wagenmakers et al., 2018). By this convention, $BF < 0.33$ provides moderate evidence in support of H_0 whilst $BF > 3$ indicates moderate evidence in favour of H_1. BFs that fall between these values are considered inconclusive, while the strength of the evidence increases the further they lay outside these boundaries (e.g. strong evidence > 10 – 30; very strong evidence > 30 – 100; extremely strong evidence >100). BFs are not reported in chapter 2 as these data were published prior to the date of submission of this thesis.

Note 2. The people in the scenes observe the items at an angle of 90 degrees to the left or right of the participant respectively. This angle is calculated from the straight line from the centre of the item to the centre of the inserted person's chest and does not change when gaze is manipulated (**Chapter 3**).

2. Chapter Two - Visual Perspective Taking as Perceptual simulation

The initial studies aimed to resolve whether representations of others' visual perspectives take the form of a perceptual simulation as has been implied in several accounts of VPT (Kampis et al., 2015; Surtees et al., 2013). These studies use a variant on a classic mental rotation task (Shepard & Metzler, 1971), and capitalise on the existing evidence that mental transformations recruit the same neural machinery as visual perception, which then drive perceptual decision making in a bottom up manner (Roelfsema & de Lange, 2016).

These studies also test (1) whether others' visual perspectives are derived spontaneously, (2) whether perspective taking occurs specifically in the presence of a human, compared to an object, and (3) whether explicit instructions to perspective take alter perceptual decision making.

The experiments in this chapter were published in Current Biology and are presented in their published form (green copy).

2.1 Spontaneous vicarious perception of the content of another's visual perspective.

Eleanor Ward, Giorgio Ganis, Patric Bach

Summary

Visual perspective taking (VPT) is a core process of social cognition, providing humans with insights into how the environment looks from another's point of view (Batson et al., 1997; Erle & Topolinski, 2015; Flavel et al., 1981; Tomasello et al., 2005). While VPT is often described as a quasi-perceptual phenomenon (Kampis et al., 2015; Surtees et al., 2013), evidence for this proposal has been lacking. Here we provide direct evidence that another's perspective can "stand in" for own sensory input perceptual decision-making. In a variant of the classic mental rotation task, participants judged whether characters presented in different orientations were canonical or mirror-inverted. In the absence of another person, we replicate the well-established positive linear relationship between recognition times and angle of orientation, such that recognition becomes slower the more an item has to be mentally rotated into its canonical orientation (Shepard & Metzler, 1971). Importantly, this relationship was disrupted simply by placing another individual in the scene. Items rotated away from the participant were recognised more rapidly not only the closer they appeared in their canonical orientation to the participant but also to this other individual, showing that another's visual perspective drives mental rotation and item recognition in a similar way as one's own. The effects were large and replicated in the three independent studies. They were observed even when the other person was completely passive, enhanced for explicit instructions to perspective-take, but reduced when the persons in the scenes were replaced with objects. The content of

another's perspective is therefore spontaneously derived, takes a quasi-perceptual form, and can stand in for own sensory input during perceptual decision-making.

Keywords: *Perspective taking; visuospatial perspective taking, mental rotation; mental imagery, imagery, perceptual decision making, perceptual simulation.*

2.2 Results

Visual perspective taking (VPT) lies at the core of the ability to make sense of other people. It allows one to derive not only which objects can be seen from another's perspective (Level 1 VPT), but also *how* these objects will look to them (Level 2 VPT) (Flavell et al, 1981; Tomasello et al., 2005). It is a phylogenetically recent, human-specific ability that forms an important milestone during development and is linked to more sophisticated mentalizing abilities, such as empathy or theory of mind (Batson et al., 1997; Erle & Topolinski, 2015; Tomasello et al., 2005). People rely on it regularly to judge how fellow drivers will respond to a difficult situation on the road, how their dance moves will be seen by others, or how to best show an object to a child so that they can recognise it easily, for example.

A recent proposal is that VPT takes a (quasi-)perceptual form, "painting" a mental image of the content of another person's viewpoint onto one's perceptual system that can stand in for one's own perception (Kampis et al., 2015; Surtees et al., 2013). In such a view, VPT not only remaps the other's spatial reference frame to one's own (e.g. that one's own left is another's right) but derives their view on an object as if one would perceive it oneself. "Seeing" the content of another's perspective in this manner could then – in a bottom-up fashion – drive all processes that operate on perceptual input (Roelfsema & de Lange, 2016) so that one's own faculties for decision making can be deployed to predict how the person will behave (Bach et al., 2014; Creem-Regehr et al., 2013; Freundlieb et al., 2016; Kovacs et al., 2010). Few, if any, studies have tested this proposal, however. While people can intentionally rotate their own body into another's perspective if so instructed (Kessler & Rutherford, 2010; Surtees et al., 2013), implicit measures only show general interference when making judgments that would be made differently from another's perspective (Sampson et al., 2010; Surtees et al., 2016; Zwickel & Muller, 2010), an effect that may index uncertainty when others would *respond* to a stimulus differently than oneself but not necessarily knowledge of how specifically they would *see* it. Moreover, these effects are only observed when sharing a task with the other individual, leading to the proposal that – while there might be a fast, automatic mechanism that derives *what* others can see, deriving *how* they see it takes longer and is under (effortful) cognitive control (Freundlieb et al., 2016; Kessler & Thompson, 2010; Surtees et al., 2016).

Here, we tested whether humans have immediate, (quasi-)perceptual access to the content of another's viewpoint. We reasoned that a different perspective might then not only interfere with own judgments but *facilitate* them, specifically if these judgments would be easier from the other's perspective. In a variant of the mental rotation task (e.g. Shepard & Metzler, 1971) (Experiment 1a, *n*=34), participants simply judged in every trial whether alphanumeric characters appearing in different orientations on a table in front of them were in canonical form or mirror-inverted (e.g. "R" vs. "Я"). Typically, decision times increase the more an item is rotated away from the participant because it needs to be mentally rotated into an orientation from which it can be judged, a process that relies on pictorial (non-abstract) item representations in early sensory cortices (Albers et al., 2013; Christophel, Cichy, Hebart & Haynes, 2015). Crucially, in 50% of trials, we inserted another person into the scene who would view the characters from either the participant's left or their right (Figure 1.1, Panel A). We hypothesized that if people have immediate, (quasi-) perceptual access to another's viewpoint, then recognition times for items rotated away from the participant should be faster if these items appear in a closer-to-canonical orientation *to this other person*, and can be better judged – or mentally rotated – from their perspective.

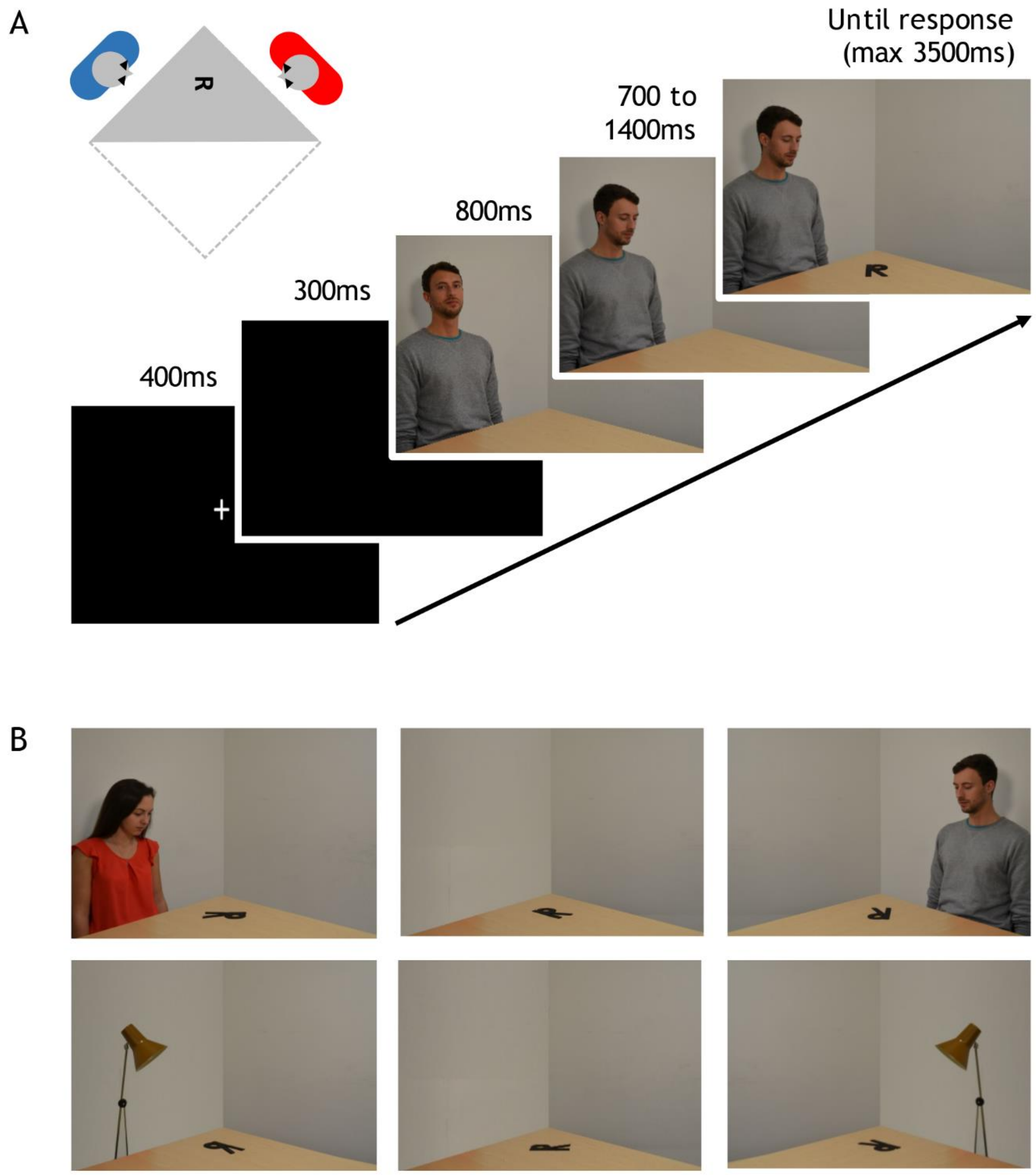

Figure 1. 1 Schematic of the trial sequence and scene setup (A) in Experiment 1a and 1b and main conditions (B). Participants judged whether alphanumeric characters were presented canonically or mirror inverted (e.g., “R” vs. “Я”), depending on whether these items appeared in the presence of another person on the left (B, upper left), a person on the right (B, upper right), in the absence of another person (B, upper middle), or in the presence of a non-human object (B, lower panels, Experiment 2, manipulated between participants). See Figure S1.3 for camera setup and measurements of all stimulus items.

2.2.1 Experiment 1a

Person location systematically biases mental rotation curves

We first confirmed that our task replicates the mental rotation effect (Shepard & Metzler, 1981). We derived a summary measure (Towards/Away-Bias) of how much faster characters are recognized the more they face the participant. For each participant and condition (No-human, Human-left, Human-right), we averaged the participant's mean response times for each character orientation, scaled by the negative of the cosine of the character orientation (see Methods). Positive values therefore indicate slower responses the more an item is oriented away from the participant, and vice versa for negative values. As expected, simple t-tests revealed slower responses for characters oriented away rather than towards participants, in all three conditions, $t(33)>7.065$, $p<.001$, $d>1.21$ (Figure 1.2A, 1.2B). This mental rotation effect was also confirmed by regressing each item's recognition time to the expected linear increase with angular disparity, revealing positive slopes in all bar one participant, mean β =1.01; $t(33)=10.5$, $p<.001$, $d=1.80$.

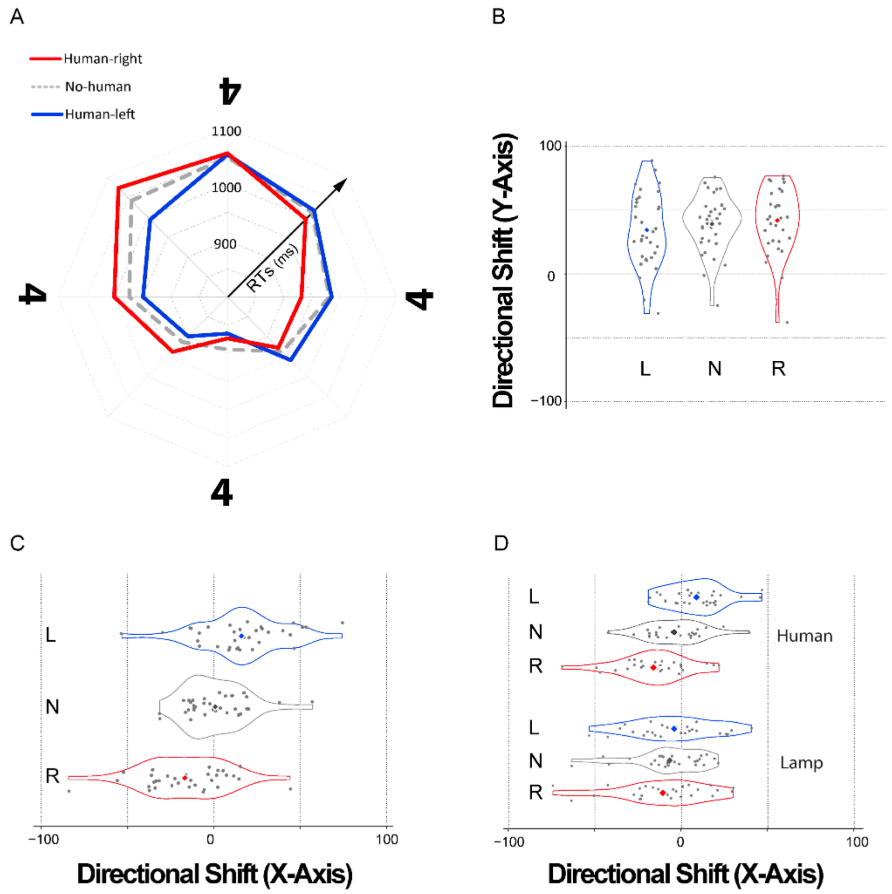

Figure 1.2 Results of Experiment 1a and 2: Person location systematically biases mental rotation curves. (A) Recognition times (ms) to correctly classify items as canonical or mirror-inverted in each of the eight orientations, depending on whether the person was absent, was sitting on the left, or was sitting on the right. B) Violin charts showing the Towards/Away-bias when the person was not present (N), was on the right (R), or was on the left (L); means are indicated by diamond symbols. (C) Violin charts showing the Left/Right-bias when the person was absent (N), was on the right (R), or was on the left (L); means are indicated by diamond symbols. (D) Results of Experiment 2: person location biases mental rotation curves more strongly than object (lamp) location. Violin charts showing the Left/Right-biases when a human (or lamp) was on the right (R), was not present (N), or was on the left (L); means are indicated by diamond symbols. See Figure S1.2 for the same data presented as line graphs, and Table S1.1 for error rate data.

The crucial test is whether shapes oriented away from the participant are easier to recognise if they appeared upright to the *other person*. We derived an analogous summary measure (Left/Right-bias, Figure 1.2A, 1.2C, see Figure S1.2A for mental rotation curves) indexing how much faster characters were recognised the more they were oriented left compared to right. For each participant and condition (No-human, Human-left, Human-right), we calculated the average recognition times, scaled by the sine of each character orientation. Positive values indicate slower recognition of right-oriented compared to left-oriented characters and vice versa for negative numbers. Directly comparing these values revealed substantial differences between Human-left and Human-right conditions, $t(33)=5.185$, $p<.001$, $d=.889$. As predicted, for persons on the left, left-oriented characters were recognised more quickly than right-oriented characters, $t(33)=3.584$, $p=.001$, $d=.614$, but vice versa for persons on the right, $t(33)=-4.074$, $p<.001$, $d=.698$. Both directional shifts differed from the (baseline) No-human condition, $t(33)>3.444$, $p<.002$, $d>.591$.

To test whether these biases reflect mental rotation from one's own *and* the other's perspective, we entered each participant's mean recognition times for all item orientations in the Human-Left and Human-Right condition into a multiple regression, with the item's angle to the participant and to the other person as two (statistically orthogonal) predictors. As expected, recognition times increased not only with the angular disparity to the participant, mean $\beta=1.06$, $t(33)=9.92$, $p<.001$, $d=1.71$, but also to the other person, mean $\beta=.41$, $t(33)=4.92$, $p<.001$, $d=.84$. Figure S1.2, top panels, shows the fit of recognition times (aggregated across all experiments) to the regression model and that they can be decomposed into two mental rotation functions from one's own and the other's perspective.

2.2.2 Experiment 1b

Replication

A replication study (Experiment 1b) with the same design (*n*=33) confirmed all findings (see Figure S1.1B). Towards/away-biases confirmed the mental rotation effect in all conditions, $t(32)>7.723$, $p<.001$, $d>1.34$. Left/Right biases differed between the Human-left and Human-right conditions, $t(32)=-4.881$, $p<.001$, $d=.85$, showing faster recognition of left-oriented than right-oriented characters in the Human-left condition, $t(32)=-4.293$, $p<.001$, $d=.747$, and vice versa for the Human-right condition, $t(32)=-2,079$, $p=.046$, $d=.362$. Both differed to the No-human (baseline) condition, $t(32)>2.993$, $p<.005$. Moreover, as before, recognition times increased linearly with the item's angle towards the participant, mean $\beta=1.14$, $t(32)=9.30$, $p<.001$, $d=1.63$, and towards the other person, mean $\beta=.40$, $t(32)=5.76$, $p<.001$, $d=1.00$.

2.2.3 Experiment 2

Person location biases mental rotation curves more strongly than object location

Having established that another's perspective speeds up recognition of characters oriented towards this perspective, (Experiment 2, *n*=54) tested whether the same was true for non-human spatial reference points, i.e. 'mind-less' objects. For half of the participants, the persons in the images were replaced with a lamp that was similarly oriented towards the items on the table as the two individuals (and therefore provided similar directional cues towards it, Figure 1.1B, lower panels). To avoid imbuing the lamp with intentionality, all motion was removed from the stimuli. Lamps and other person did not initially "look" outwards at the participant and then back at the table, but started the trial already facing the location at which the character would appear.

As before, Towards/Away-biases confirmed the mental rotation effect, in all conditions of both groups, *t*(27)>10.837, *p*<.001, *d*>2.05, for all. To test whether VPT shifts towards other persons are larger than towards objects, we entered the Left/Right biases into an ANOVA with the within-subjects factor Location (Object/Human-left, Object/Human-right) and the between-subjects factor Group (object, human). A main effect of Location replicated the known Left/Right-biases towards person locations, *F*(1,52)=23.328, *p* <.001, η_p^2=.310, which was further qualified by an interaction with Group, *F*(1,52)=7.636, *p*=.008, η_p^2=.128, revealing larger shifts in the human than the object group (Figure 1.2D, see Figure S1.1CD for mental rotation curves).

Step-down analysis of the human group fully replicated the known pattern. Left/Right-shifts differed between Human-left and Human-right conditions, *t*(27)=5.729, *p*<.001, *d*=1.08, and each differed to the baseline No-human condition, *t*(27)>2.449, *p*<.021, *d*=.457. Recognition times again reflected the item's angular difference to the participant, mean *β*=1.46, *t*(27)=11.0, *p*<.001, *d*=2.09, and to the other person, mean *β*=.31, *t*(27)=5.70, *p*<.001, *d*=1.07.

These shifts were virtually eliminated in the object group (Figure 1.2D). Left/Right biases in the Object-left, Object-right, and No-Human conditions did not differ from each other, *t*(25)<1.722, *p*>.097, *d*<.338. Moreover, recognition times only captured the item's angular difference to the participant, mean *β*=1.83, *t*(25)=16.6, *p*<.001, *d*=3.27, but not to the lamp, mean *β*=.08, *t*(25)=1.15, *p*=.262, *d*=.23, and this mean beta-coefficient was smaller than in the Human group, *t*(52)=2.72, *p*=.009.

2.2.4 Experiment 3

Explicit perspective taking increases bias towards other persons

People can make perceptual judgements from another's perspective if explicitly instructed (Kessler & Rutherford, 2010; Surtees et al., 2013). Experiment 3 (*n*=52) therefore explicitly instructed half of participants to adopt the other persons' perspective (Explicit-VPT group) while giving no instructions to the other half (Implicit-VPT group), as in previous experiments. If our task truly indexes perspective taking (rather than biases in rotation direction for example), then such an explicit instruction should elicit the same, but more pronounced, biases towards the other persons, and be described by the same regression model.

Results fully supported our predictions. As before, Towards/Away-biases confirmed the mental rotation effect, in all conditions of both groups, $t(25)>5.168$, $p<.001$, $d>1.01$, for all. An ANOVA on Left/Right biases revealed the expected main effect of Person Location, $F(1,50)=120.419$, $p<.001$, $\eta_p^2=.707$, which interacted with Group, $F(1,50)=13.464$, $p=.001$, $\eta_p^2=.212$. Thus, as expected, the faster recognition of left- compared to right-oriented characters for persons sitting on the left, and vice versa for persons on the right, was even more pronounced in the Explicit-VPT than the Implicit-VPT group (Figure 1.3).

As before, our simple regression model described recognition times in terms of mental rotation from one's own and the other's viewpoint. Yet, while in the Implicit-VPT group recognition times again showed a stronger contribution of the item's angular difference to the participant, mean $\beta=1.31$, $t(25)=9.89$, $p<.001$, $d=1.93$, than the other person, mean $\beta=.40$, $t(25)=6.79$, $p<.001$, $d=1.34$, this difference was reduced in the Explicit-VPT group (Figure 1.3). It showed strong weightings of both the angular disparity to the participant, mean $\beta=.91$, $t(25)=6.182$, $p<.001$, $d=1.21$, and to the other person, mean *β*=.84, *t*(25)=8.412, *p*<.001, *d*=.1.65. Both coefficients differed between groups, *t*(50)>2.017, *p*<.049, *d*>.56.

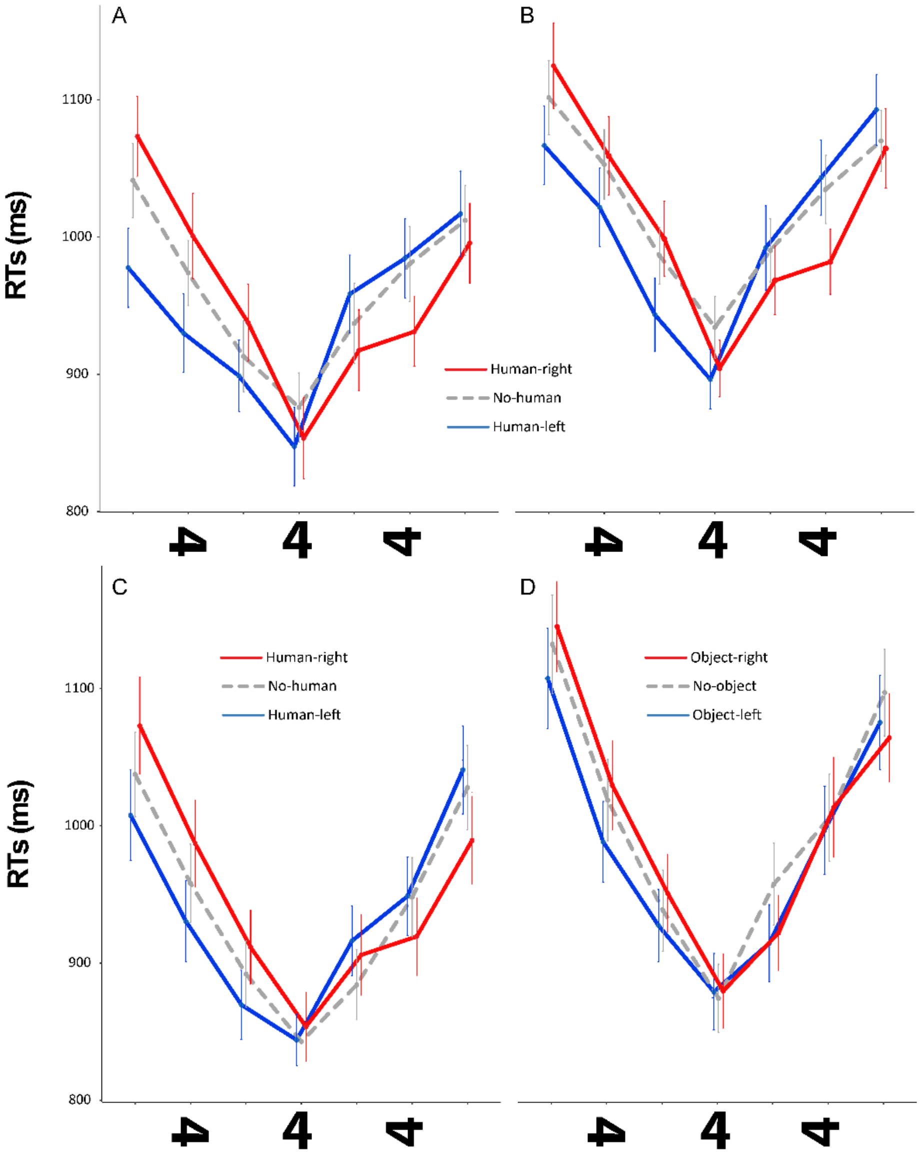

Figure 1.3. Results of Experiment 3: Explicit perspective-taking increases bias towards other persons. (A) Recognition times (ms) in the Explicit-VPT group to correctly classify items as canonical or mirror-inverted in each of the eight orientations, depending on whether the person was absent, was sitting on the left, or was sitting on the right. B) Violin charts showing the Towards/Away-bias when the person was not present (N), was on the right (R), or was on the left (L); means are indicated by diamond symbols. (C) Violin charts showing the Left/Right-bias when the person was absent (N), was on the right (R), or was on the left (L); means are indicated by diamond symbols. (D) Comparison with Implicit-VPT group: person

location biases mental rotation curves in the Explicit-VPT group more strongly than in the Implicit-VPT group. Violin charts showing the Left/Right-biases in both groups when a human was on the right (R), was not present (N), or was on the left (L); means are indicated by diamond symbols. See Figure S1.2 for the same data presented as line graphs, and Table S1.1 for error rate data.

Other's perspectives cause both facilitation and interference

Exploratory analyses with data pooled across experiments (1a, 1b, 2 human condition, 3 implict-VPT condition, see methods for details) to increase power showed that the Left/Right shifts in the Human-left and Human-right condition reflected both facilitation for characters oriented towards the other persons, $t(120)=7.27$, $p<.001$, $d=.66$ and, to a lesser degree, interference, $t(120)=5.17$, $p=.001$, $d=.47$, when characters were oriented away, relative to the No-Human baseline condition.

2.3 Discussion

We tested whether humans have direct, (quasi-)perceptual access to the content of another's perspective that can stand in for own input. We show that the classic finding from mental rotation tasks – that it takes longer to recognise an item the more it has to be mentally rotated into its canonical orientation (Shepard & Metzler, 1971) – is disrupted by other persons in the scene. In this case, recognition times increase not only with an item's angular disparity to the participant but also to the other person, such that items oriented away from participants are recognised more rapidly if they are oriented towards this person (and even more slowly when oriented away from them). Recognition times therefore reflect an integration of one's own and the others' perspectives, either from parallel processes within each judgment or across them, when participants fluently switch between own and others' perspectives. The resulting biases had a large effect size, were sensitive to instructions to take the other's perspective, but decreased when the persons were replaced with mind-less objects, even when these objects had the same directionality and faced the items as the persons did. Results therefore show that the content of another's perspective is available in (quasi-) perceptual form, so that the characters could be mentally rotated from this perspective and are recognised more rapidly if they appear in their canonical "upright" orientation from this point of view.

While VPT has been proposed before to reflect perceptual simulations of others' viewing perspective (Kampis et al., 2015; Surtees et al., 2013) direct evidence has been lacking. Prior work has only shown interference on a conceptual and/or response level, leaving open whether participants simply represent that others see a scene *differently*, without representing their specific perceptual input (Cole et al., 2016; Santiesteban et al., 2014). Moreover, in these tasks, VPT was only observed when the other person was at least somewhat relevant to one's own task, leading to the proposal that while humans are endowed with a quick (and potentially automatic) mechanism that computes *what* others can see, deriving *how* they see it is under cognitive control (Freundlieb et al., 2016; Kessler & Thompson, 2010; Sampson et al., 2010; Surtees et al., 2016) and may require effortful mental rotation of one's own body into that of the other (Kessler & Rutherford, 2010; Surtees et al., 2013).

In contrast, the perceptual biases observed here reveal that humans can rely on others' visual perspective to drive own perceptual decision making. They therefore shown that VPT takes the form of a perceptual simulation that can drive subsequent processing like actual input, allowing it to be integrated into recent perceptual accounts of imagery and working memory, for which a similar reliance on (even early) visual representations has been demonstrated (see Roelfsema & de Lange, 2016, for review]. Second, they challenge the notion that VPT2 is under cognitive control. Here, another's viewing perspective drove processing even (1) when the person stimulus was completely passive and task-irrelevant, (2) without requiring blocking person presence/location across trials or (3) asking participants to switch between self- and other-perspective, which can induce carry-over effects (Freundlieb, Kovacs & Sebanz, 2018; Freundlieb et al., 2016). Moreover, (4) effects occurred rapidly, with an upper limit provided by the time it would take to mentally rotate the item oneself (about 100 ms for orientations to $90/270^0$), and involuntarily, even in trials (5) in which the other's perspective did not help but *interfere* with item recognition (because items were oriented even further away from them).

Together, these findings provide direct evidence that humans are endowed with a mechanism that allows them to rapidly and spontaneously derive *how* others see an object. The (inferred) view of the other person takes a quasi-perceptual form that can stand in for own sensory input. In this way, the content of another's perspective can feed directly into perceptual representations and can drive, in a bottom-up manner, the processes operating on them, without explicit control, as has been demonstrated for imagery and working memory (Roelfsema & de Lange, 2016). Such spontaneous perceptual simulations of others' viewing perspective could explain not only why better perspective takers are more empathetic (Erle & Topolinski, 2015; Gronholm, Flynn, Edmonds & Gardner, 2012), why people sometimes describe object locations from another's viewpoint, or why patients with hemispatial neglect sometimes report neglected items when imagined from a different perspective (Becchio, Del Guidice, Dal Monte, Latini-Corazzini & Pia, 2011). They may also provide novel insights into more sophisticated socio-cognitive abilities in humans, such as joint action and mentalizing, and their impairment in autism, which have been shown to be related to VPT (Hamilton et al., 2009).

2.4 Method

Participants

203 (166 females) naive participants (42 in Experiment 1a, 37 in Replication Experiment 1b, 56 in Experiment 2, 68 in Experiment 3) were initially recruited via the University of Plymouth participation pool. All participants were adults (age range 18-50) and gave written informed consent according to the declaration of Helsinki. Approval was obtained from the University of Plymouth Ethics Committee. Participants received course credit as compensation. After exclusion (for criteria, see below), 34 participants (28 females; mean age: 22.8 years, range: 18-50) were considered for analysis in Experiment 1a, 33 in Experiment 1b (28 females, 21.7 years, age range: 19-50), 54 in Experiment 2 (42 Females; mean age: 20.07 years, age range: 18-34), and 52 in Experiment 3 (47 Females; mean age: 20.5 years, age range: 18-30).

Elimination criteria

In all experiments, erroneous responses were excluded from the analysis of recognition times (RTs), as well as trials with RTs longer than 2000ms, or shorter than 150ms. Participants with error rates in excess of 20% across all conditions were not considered for analysis (Experiment 1a, n=8; Experiment 1b, n=4; Experiment 2, n=2; Experiment 3, n=16).

Power analysis

Power analyses were conducted in G*Power (Version 3.1) [27], assuming a power of .80, using the sensitivity analysis function. We do not report theoretical power based on previously reported effect sizes as this neglects uncertainty around these effect size measurements [28-29]. Instead, we report effect sizes that can in principle be detected with our experimental parameters (i.e. given required power, participant numbers and type of test). These analyses were conducted for the crucial comparison of Left/Right-biases in the Human-left and Human-right conditions, which measures the extent that person presence biases mental rotation towards either 90^0 or 270^0.

For initial Experiment 1a, a final sample size of *n*= 34 provides .80 power to detect effect sizes in the predicted direction (one-sided) of at least Cohen's *d*=.43 and effects in either direction of Cohen's *d=.*49*.* As measured effect sizes from this experiment (and in additional piloting in our lab) robustly exceeded this criterion (*d*=.89), we replicated the study with similar sample sizes. For the replication Experiment 1b, the final sample size of *n*=33 allows us to detect effects in the predicted direction of at least Cohen's *d*=.44 and effects in either direction with Cohen's *d*=.50. Assuming that the object condition eliminates the effects in reported Experiment 1a and 1b, and that the between-participant standard deviation stays the same, anticipated effect sizes for the crucial between-groups comparison should again be larger (*d*=.89, when derived from effect sizes of Experiment 1, *d*=.85 when derived from Experiment 2). For Experiment 2, the final sample size of *n*=54 allows us to detect between-group differences of *d*=.69 (one-sided) and .77 (two-sided). Similarly, for Experiment 3, the final sample size of *n*=54 allows us to detect between-group differences of *d*=.69 (one-sided) and .79 (two-sided).

Experiment 1a and 1b - Apparatus, stimuli & procedure

All experiments were conducted in behavioural testing lab space of the University of Plymouth. The experiments were administered using Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Stimuli were presented on a 19" LED computer monitor (Resolution: 1900x1200; Refresh rate: 60Hz). Responses were made on a standard computer keyboard with UP, DOWN, and SPACE keys as active response keys. Red and green stickers were positioned on the DOWN and UP keys, respectively.

Participants sat upright facing the screen at a distance of approximately 60cm and were given written and verbal instructions. They were given examples of the rotated items that would appear on the screen and completed eight training trials that were identical to the main experiment (Figure 1.1). Each trial (total trials = 520) started with a fixation cross displayed for 400ms, followed by 300ms blank screen. The subsequent stimulus sequence included three frames, presented without inter-stimulus interval, creating the impression of apparent motion (Wertheimer, 1912). The first frame, measuring 33.4 by 23.5 degrees of visual angle, was presented for 800 ms. In 50% of the trials, it showed a view onto a corner of a square table in a grey room. In the other 50%, it showed a person sitting behind the same square table, gazing outwards at the participant. The person (either male or female) sat either on the left or right side of the table (12.5% of trials, each).

The second frame in the sequence was identical to the first for the No-human trials. In the trials with the person, (s)he now looked down towards the middle of the table. Then, after a random interval between 700ms and 1400ms, the third frame was presented. This was identical to frame 2, but now one of 64 possible items appeared on the table, at the location on the table the on-screen person was gazing at. This item was one of four alphanumeric characters (4, Z, P, or R), presented either in the canonical version or mirror-inverted about their vertical axis, in one of eight orientations (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°, with 0° denoting the upright canonical orientation and angles increasing in a counter-clockwise fashion) relative to the participant. The characters always appeared in the same position on the table, half-way between the outward corner of the table and its centre, such that the persons to the left and right would gaze at the table from roughly 90° and 270°, respectively (perpendicular to the viewpoint of the participant, see Figure 1.1A inlay, Figure S1.3 for measurements of the scene setup), as at these angles the character's angular disparities from the participant and the other person were statistically independent across conditions. Character rotation occurred around the character's centre point.

Participants were asked to judge whether each character was presented in its canonical or mirror-inverted form. The third frame remained on the screen until a response was made to a maximum duration of 3500ms. Participants responded using their right hand by pressing the green key to indicate a canonical item and the red key to indicate a mirrored item. Response times were measured relative to item onset.

Following this task, participants were asked to report via free text any particular strategies they used for their judgements. Some reported that they noticed themselves sometimes taking the perspective of the other persons: Experiment 1a, *n*=5; Experiment 1b, *n*=5; Experiment 2, *n*=6; Experiment 3, *n*=4. No participants in the object condition reported taking the perspective of the lamp. All effects reported in the main text remain if these participants are removed from the analyses (see Robustness analyses below).

For a demonstration of the design (albeit with slightly different timing), please see the movie here: https://goo.gl/eAV1eg

Experiment 2 - Apparatus, stimuli & procedure

Participants completed a total number of 584 trials, following the same apparatus and procedure as in Experiment 1a, with the following exceptions. First, for half of the participants, the person in the picture was replaced with a lamp of similar height and orientation towards the character. The lamp was chosen as it possesses similar characteristics as other people, such as a clear front and back, and the potential to "look" in a certain direction on the table (see Figure 1.1B, lower panels). Second, the letter Z was removed from the item list because participants alerted us that it could be mistaken for an N when rotated 90 degrees. Third, the first (outward gaze towards the participant) frame (Figure 1.1A) was removed from the sequence. In Experiment 1a, the transition from Frame 1 (person gazing towards the participant) to Frame 2 (person gazing at the table) resulted in apparent motion and the impression of the other person directing attention to the item, which drew attention to the person, might signal internal states, and would therefore not be appropriate to the inanimate and mind-less lamp stimulus. In order to keep the trial duration consistent with that of the first experiment, the original second image in the sequence was presented for an additional 800ms. Fourth, to increase power for the crucial comparison between person/object on the left and right, the No-human trials were reduced to 33%. The remaining two thirds of trials were divided equally between the person/lamp appearing on the left and the right (33% each). To ensure that person and lamp trials were of equal variability, each participant in the Human group only saw only one of the two individuals (the male or the female), counterbalanced between participants.

Experiment 3 - Apparatus, stimuli & procedure

Participants completed a total number of 584 trials, following the same apparatus and procedure as in Experiment 2 (human condition), with the exception that half of the participants were simply instructed to take the perspective of the actors when they appeared on the screen, and to respond from their own perspective when there was no actor present. The exact instruction given was '*You will see some people appear on the screen. When this happens, I would like you to respond from their visual perspective. If there are no people visible on the screen, you should respond from your own perspective'.* The other half of participants did not receive any specific instructions to perspective-take. In fact, the presence of persons in some trials was not mentioned.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data (pre-)processing and analysis were conducted in Microsoft Excel (2010), SPSS (version 23). Figures 1.2 and 1.3 were created using R (ggplot2). Power analyses were conducted in G*Power (Version 3.1; Erdfelder, Faul, & Buchner, 1996).

Dependent measures

Dependent measures were the recognition times (measured from item onset) for each character orientation (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315), depending on condition (No-human, Human-left, Human-right). Analogous analyses of error rates were also conducted to rule out speed/accuracy trade-offs. In all experiments, error rates numerically followed the pattern of the main recognition times but did not show statistically reliable differences (Table S1.1).

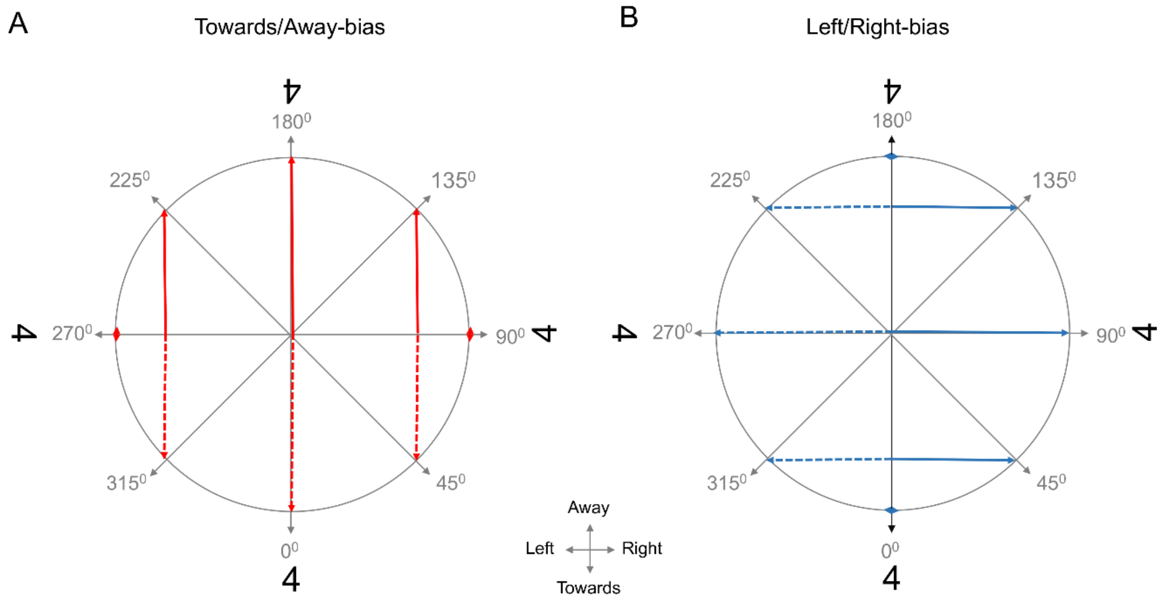

Figure 1.4. The contribution of each character orientation to the summary measures Towards/Away-bias (A) and Left/Right-bias (B). Each character's contribution to the Towards/Away-biases and Left/Right-biases are defined by the negative of the cosine and the sine of the character rotation angle, respectively. The red arrows show the positive (filled arrows) and negative (dotted arrows) contributions of each character orientation to the Towards/Away-bias. The blue arrows show the positive (filled arrows) and negative (dotted arrows) contributions to the Left/Right-bias.

Toward/Away-bias and Left/Right-bias

To quantify the recognition times when the characters either faced the participant (i.e. was seen in its canonical orientation from the perspective of the participant) or the other person in the scenes, we derived two analogous and statistically independent summary measures. The first summary measure (Toward/Away-bias) indexes to what extent characters were recognized faster the more they were facing towards the participant (0°) rather than away from them (180°), separately for each participant and each condition (No-human, Human-left, Human-right). This measure therefore quantifies the mental rotation effect [18]. The second summary measure (Left/Right-bias) indexes how much faster characters were recognized when oriented towards the left (270°) rather than right (90°), or vice versa.

The contribution of each character orientation to the two summary measures is derived by treating each participant's recognition time for this character orientation as a vector in a coordinate system, with the recognition time providing the distance from the origin and the rotation angle the polar angle. A character orientation's contribution to the Toward/Away-bias is then derived simply from the recognition times multiplied with the negative of the cosine of the orientation angle. As a result, characters contribute negatively the more they face the participant (315°, 0°, 45°; Figure 1.3A, red filled arrows) and positively they more they are oriented away from them (225°, 180°, 135°; Figure 1.3A, red dotted arrows). Similarly, the contribution of a character's orientation to the Left/Right-bias was calculated as the recognition time multiplied with the sine of the orientation angle. Character orientations contribute positively the more they face to the right (45°, 90°, 135°; Figure 1.3B, blue dotted arrows) and negatively the more they face to the left (225°, 270°, 315°; Figure 1.3B, blue filled arrows). This procedure effectively maps the changes evident in the radar plots for each angle onto two orthogonal and statistically independent summary measures, so that they can be compared across conditions (either within- or between-participant), without accruing alpha inflation due to multiple testing, which would result if each of the eight angles were compared separately.

By averaging these values, separately for each summary measure, participant and condition (No-human, Human-left, Human-right), we are able to calculate, first, whether characters were recognized faster the more they appear in the canonical orientation to the participant (negative values on the Toward/Away-bias) compared to when they are oriented away (positive values), reflecting the expected mental rotation effect. Similarly, they allowed us to calculate to what extent characters were recognized faster the more they are oriented leftwards and would appear in their canonical orientation to a person sitting to the left (positive values on the Left/Right-bias) rather than rightwards, appearing in their canonical orientation to a person sitting on the right (negative values). We were then able to determine if this left/right bias changed depending on whether another person was presented in the scenes and on whether the person was on the left or on the right.

The crucial comparison is the difference between the Left/Right-biases in the Human-left and Human-right conditions, but additional comparisons of interest are also reported. Note that the direct comparison of the Human-left and Human-right conditions is statistically identical to the comparison of how much person presence shifts mental rotation performance in the Human-left and Human-right conditions relative to the No-Human baseline (i.e. how much person presence shifts recognition times away from 0^0 towards either 90^0 or 270^0), as this would involve subtracting the same baseline value from each of the two conditions for each participant, and would therefore not affect the absolute difference between them. In Experiment 1a and 1b, differences between these conditions were assessed with repeated measures t-tests. Effect sizes were measured in Cohen's *d*. In Experiment 2, a mixed-factors ANOVA with the within-subjects factor Location (left, right) and between-subjects factor Group (human, object) was used to assess whether the difference in Left/Right-biases towards a person on the left or right was larger than the difference in perceptual shifts towards an inanimate object (a lamp) in the same position. In Experiment 3, the same ANOVA model was used to whether the difference in Left/Right-biases towards a person on the left or right increased when participants were explicitly instructed to take the other person's perspective compared to when they received no instruction, as in the previous experiments.

Distinguishing between facilitation and interference

Additional analyses were conducted on the pooled data across experiments (Exp. 1a, Exp. 1b, Human condition of Experiment 2) to verify whether the measured Left/Right-biases towards person locations reflect only facilitation (i.e. faster recognition of items oriented towards another person) or interference as well (slower recognition of items oriented away from the other person). To this end, we used the same logic as for the overall perceptual shifts and again scaled average recognition times by the sine of each character orientation, but separately for the trials in which the items were facing towards the other person (e.g. for a person on the right: 45^0, 90^0, 135^0) and away from the other person (e.g. 315^0, 270^0, 225^0), separately for the Human-Left and Human-Right conditions, across all participants of the three experiments to increase power (Experiment 1a, 1b, human group of Experiment 2, implicit group of Experiment 3). We were then able to separately assess, with paired t-tests, whether recognition times were generally slower than would be expected from the baseline No-human condition for items facing away from the other person (i.e. measuring interference), and whether they were faster than expected when facing towards the other person (measuring facilitation).

Across-participant regression analyses

In prior work, the mental rotation effect is sometimes characterised in terms of separate linear regressions of an items' recognition time to its angular disparity relative to the participant, for each participant separately [18]. The results reveal linear increases with increasing angular disparity for the large majority of participants. Here, we used this analysis model to test whether an item's recognition times can be described, on a single participant basis, as a linear increase of the character's angular disparity *both* to the participant and to the other person. To this end, we entered each participant's item mean recognition times for each character orientation in the Human-left and Human-right condition as dependent variable in a multiple regression, with the item's angular disparity to the participant and to the other person as two statistically independent predictors. This analysis provides regression coefficients for both predictors – angular disparity to participant and other person – for each participant separately. We report mean across-participant regression coefficients for each of these two predictors and compare them with t-tests against zero.

The fit of model and observed data can be seen in Figure S1.2, separately for the aggregated data across all experiments in which VPT was induced implicitly in the top row (Exp. 1a, 1b, Exp. 2 human group, Exp. 3 Implicit-VPT group) and when it was induced explicitly in the bottom row (Exp. 3). To increase power, the data of the Human-right condition was "flipped" and collapsed onto the data for the (mirror-symmetrical) Human-left condition (so that recognition times for 270^0 when a person sitting on the left map onto 90^0 for a person sitting on the right). The left panels (A, D) show the data predicted from the average across-participant regression coefficients and intercepts, for the No-Human "baseline" condition, and the Human-left condition. The middle panels (B, E) show the observed data for the No-human condition and the Human-left condition. Finally, the right panels (C, F) show the residuals when *only* the expected mental rotation function from the participant's own perspective is regressed out. As can be seen, these residuals show a mental rotation function from the perspective of the other person (sitting at 270^0), showing graphically that, in the Human-left and Human-right conditions, an item's recognition times across angles can be decomposed into mental rotation functions from one's own and the other person's perspective.

Robustness analyses

We confirmed that our effects did not depend on specific (groups of) participants. We first verified that our analysis still holds when all participants were excluded that explicitly reported noticing taking the perspective of the other individuals in the scene. This was the case, without substantially reducing overall effect sizes (Experiment 1a, excluded n=4, $t(29)=5.052$, $p<.001$, $d=.92$; Experiment 1b, excluded n=3, *t*(29)=4.316, *p*<.001, *d*=.79; Experiment 2, human group, n=5, *t*(22)=5.167, *p*<.001, *d*=1.08; no participants reported VPT in Experiment 2, object group; Experiment 3, implicit group, excluded n=4, *t*(21)=6.634, *p*<.001, *d*=1.41).

Second, we verified that our effects are not over-inflated due to some participants being older than usual (>35 years, Experiment 1a, *n*=5; Experiment 1b, *n*=2; Experiment 2, *n*=0; Experiment 3, *n*=0). While there was a weak correlation with the difference in Left/Right-bias between the Human-left and Human-right condition and participant age across all participants (*r*=.294, *p*=.001), re-running these tests with older participants excluded revealed overall identical effect sizes. Indeed, when we exclude all participants older than 35, the overall effect size in Experiment 1a decreases (from *d*=.89. to *d*=.80) but increases in Experiment 1b (from *d*=.85 to *d*=.94).

2.5 Supplemental Information

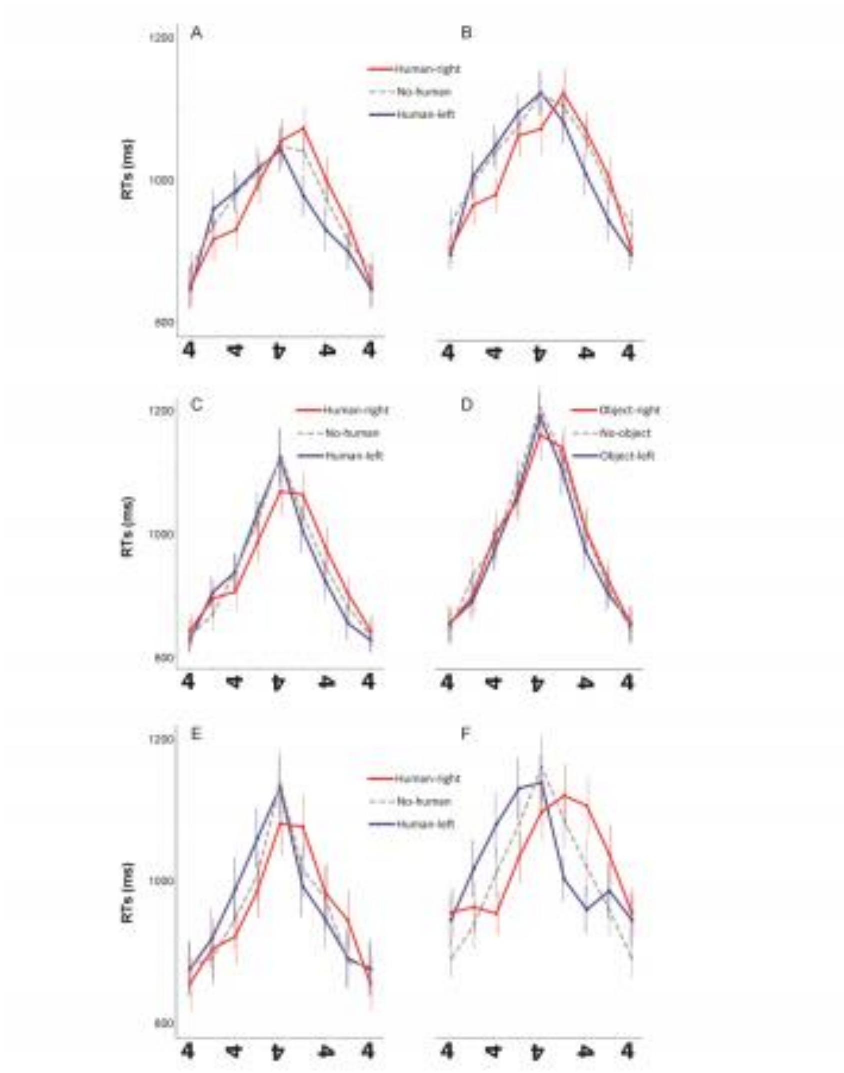

***Figure S1*.1. Results of Experiments 1a, 1b, 2 and 3 in line graph form.** Relates to Figure 1.2 (Person location systematically biases mental rotation curves) and Figure 3 (**Explicit perspective-taking increases Left/Right-bias)**. (A) Results of Experiment 1a. (B) Results of Experiment 1b. (C) Results of Experiment 2, Human Group. (D) Results of Experiment 2, Object group. (E) Results of Experiment 3, Implicit-VPT group. (F) Results of Experiment 3, Explicit-VPT group. In each plot, data points show recognition times (RTs) for all eight character orientations when the other person was not present (No-Human, grey lines), was on the right (Human-Right, red lines), or was on the left (Human-Left, blue lines). Note that, following symmetry conventions in these plots in the literature, the data points for 0^0 angular disparity are identical with the data points for 360^0 angular disparity.

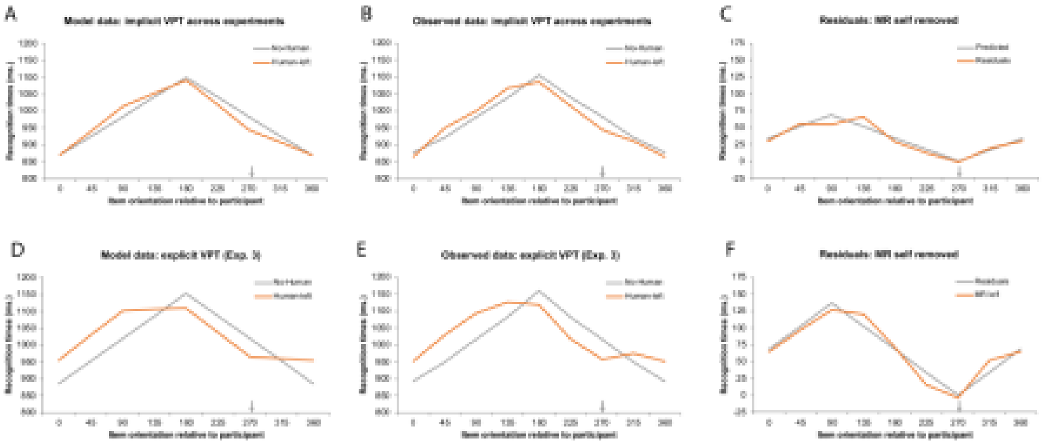

***Figure S1.2*. Fit of regression model and observed data.** Relates to Figure 1.2 and 1.3. The top panels show the aggregated data across all experiments in which VPT was induced implicitly (Exp. 1a, 1b, Exp. 2 human group, Exp. 3 Implicit-VPT group). The bottom row shows the data when VPT was induced explicitly (Exp. 3). For all panels, the data of the Human-right condition was “flipped” and collapsed onto the data for the (mirror-symmetrical) Human-left condition, with the small arrow indicating the viewpoint of the other person at 270^0. The left panels (A, D) show the data predicted from the average across-participant regression coefficients and intercepts, for the No-Human “baseline” condition, and the Human-left condition. The middle panels (B, E) show the observed data for the No-human condition and the Human-left condition. The right panels (C, F) show the residuals when only the expected mental rotation function from the participant’s own perspective is regressed out, leaving a mental rotation function centred on the other person’s perspective (270^0).

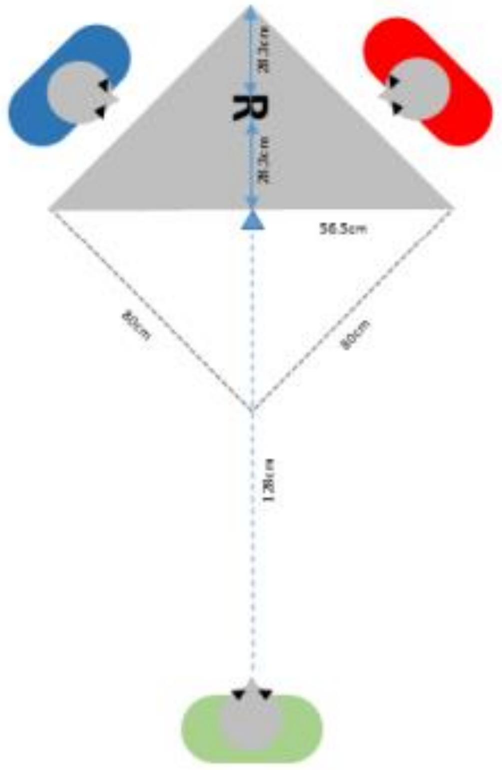

***Figure S1.3*. Scene Dimensions.** Measurements for the setup of table, person locations (red, blue) and observer/camera position (green) from which all experimental stimuli were derived. Relates to Figure 1.1.

Table S1.1 Means (M) and Standard Deviations (SD) for the Left/Right- and Towards/Away-biases in Error rates in Experiment 1a, 1b, Experiment 2, and Experiment 3. Relates to Figure 1.2 and Figure 1.3 and Figure S1.1. Forward/Away and Left/Right-biases were calculated analogously as for the recognition times. **p*<.05. **p<01, ****p*<.005.

	Y-Axis			X-Axis		
Experiment	Human-Left M(SD)	Human-Right M(SD)	No-human M(SD)	Human-Left M(SD)	Human-Right M(SD)	No-human M(SD)
1a	.014 (.025)**	.011(.017)***	.019(.016)***	.002(018)	-.004(018)	.002(015)
1b	.020(.020)***	.015(.023)***	.014(.014)***	.006(019)	-.002(017)	-.002(012)
2 Human	.021(.015)***	.024(.02)***	.027(.025)***	.001(012)	-.003(018)	-.001(015)
2 Object	.027(.025)***	.027(.026)***	.023(.025)***	-.001(018)	-.001(016)	-.002(013)
3 Implicit	.019(.015)***	.02(.021)***	.017(.017)***	.005(.011)*	-.002(.016)	.001(.017)
3 Explicit	.016(.024)***	.012(.022)*	.021(.024)***	.015(.022)***	-.008(.023)	-.004(.012)

2.6 References

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., and de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. Curr. Biol. 23(15), 1427-1431.

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. Psychological Science, 28(11), 1547-1562.

Bach, P., Fenton-Adams, W., and Tipper, S.P. (2014). Can’t touch this: the first-person perspective provides privileged access to predictions of sensory action outcomes. J Exp Psychol Hum Percept Perform. 40(2), 457.

Batson, C. D., Early, S., and Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imaging how you would feel. Personality and Social Psychology Bulletin. 23(7), 751-758.

Becchio, C., Del Giudice, M., Dal Monte, O., Latini-Corazzini, L., and Pia, L. (2011). In your place: neuropsychological evidence for altercentric remapping in embodied perspective taking. Social cognitive and affective neuroscience, 8(2), 165-170.

Christophel, T. B., Cichy, R. M., Hebart, M. N., and Haynes, J. D. (2015). Parietal and early visual cortices encode working memory content across mental transformations. NeuroImage. 106, 198-206.

Cole, G. G., Atkinson, M., Le, A. T., and Smith, D. T. (2016). Do humans spontaneously take the perspective of others?. Acta Psychol (Amst). 164, 165-168.

Creem-Regehr, S. H., Gagnon, K. T., Geuss, M. N., and 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.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. Behavior Research Methods, Instruments, & Computers, 28(1), 1-11.

Erle, T. M., & Topolinski, S. (2015). Spatial and empathic perspective-taking correlate on a dispositional level. Social Cognition. 33(3), 187-210.

Flavell, J. H., Everett, B. A., Croft, K., and Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. Dev Psychol. 17(1), 99.

Freundlieb, M., Kovács, Á. M., and Sebanz, N. (2016). When do humans spontaneously adopt another's visuospatial perspective? J Exp Psychol Hum Percept Perform. 42(3), 401.

Freundlieb, M., Kovács, Á. M., and Sebanz, N. (2018). Reading Your Mind While You Are Reading—Evidence for Spontaneous Visuospatial Perspective Taking During a Semantic Categorization Task. Psychological Science. 29(4), 614-622.

Freundlieb, M., Kovács, Á.M., and Sebanz, N. (2016). When do humans spontaneously adopt another's visuospatial perspective?. J Exp Psychol Hum Percept Perform. 42(3), 401.

Gronholm, P. C., Flynn, M., Edmonds, C. J., and Gardner, M. R. (2012). Empathic and non-empathic routes to visuospatial perspective-taking. Conscious Cogn. 21(1), 494-500.

Hamilton, A. F., Brindley, R., and Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. Cognition 113, 37–44.

Kampis, D., Parise, E., Csibra, G., and Kovács, Á. M. (2015). Neural signatures for sustaining object representations attributed to others in preverbal human infants. Proc. R. Soc. B. 282(1819), 20151683.

Kessler, K., and Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. Frontiers in Psychology. 1, 213.

Kessler, K., and Thomson, L. A. (2009). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. Cognition. 114.

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

Lakens, D., & Albers, C. (2017). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. Journal of Experimental Social Psychology, 74, 187-195.

Roelfsema PR, and de Lange FP (2016). Early visual cortex as a multi-scale cognitive blackboard. Annual Reviews in Vision Science. 2,131-151

Samson, D., Apperly, I.A., Braithwaite, J.J., Andrews, B.J., and Bodley Scott, S.E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. J Exp Psychol Hum Percept Perform. 36(5), 1255.

Santiesteban, I., Catmur, C., Coughlan Hopkins, S., Bird, G., and Heyes, C. (2014). Avatars and arrows: implicit mentalizing or domain-general processing? J Exp Psychol Hum Percept Perform. 40, 929–937.

Shepard, R. N., and Metzler, J. (1971). Mental rotation of three-dimensional objects. Science. 171(3972), 701-703.

Surtees, A. D. R., Apperly, I. A., & Samson, D. (2013). The use of embodied self-rotation for visual and spatial perspective-taking. Frontiers in Human Neuroscience. 7, 698

Surtees, A., Samson, D., and Apperly, I. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. Cognition. 148, 97-105.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. Behav Brain Sci. 28, 675–691.

Zwickel, J., and Müller, H.J. (2010). Observing fearful faces leads to visuo-spatial perspective taking. Cognition. 117(1), 101-105

3. Chapter Three - Gaze Direction as a cue to VPT

The studies in this chapter aimed to resolve whether perceptual simulations of others' perspectives are specifically driven by where the actor in the scenes was looking, or whether VPT reflects a more general spatial navigational ability. Again, these studies use a variant on a classic mental rotation task (Shepard & Metzler, 1971).

The experiments in this chapter were submitted to Cognition and are currently undergoing a first round of revisions. Experiments presented in their submitted form (green copy). A pre-print of this chapter is available at https://psyarxiv.com/hfnrw, and will be updated once resubmitted to Cognition.

3.1 Perspective taking as virtual navigation? Perceptual simulation of how others see the world reflects their location in space but not their gaze.

Eleanor Ward*, Giorgio Ganis, Katrina McDonough, Patric Bach

Abstract

Other peoples' (imagined) visual perspectives are represented perceptually in a similar way to our own, and can drive bottom-up processes in the same way as own perceptual input (Ward, Ganis & Bach, 2019). Here we test directly whether visual perspective taking is driven by where another person is looking, or whether these perceptual simulations represent their position in space more generally. Across two experiments, we asked participants to identify whether alphanumeric characters, presented at one of eight possible orientations away from upright, were presented normally, or in their mirror-inverted form (e.g. “R” vs. “Я”). In some scenes, a person would appear sitting to the left or the right of the participant. We manipulated either between-trials (**Experiment 4a**) or between-subjects (**Experiment 4b**), the gaze-direction of the inserted person, such that they either (1) looked towards the to-be-judged item, (2) averted their gaze away from the participant, or (3) gazed out towards the participant (Exp. 2 only). In the absence of another person, we replicated the well-established mental rotation effect, where recognition of items becomes slower the more items are oriented away from upright (e.g. Shepard and Meltzer, 1971). Crucially, in both experiments and in all conditions, this response pattern changed when another person was inserted into the scene. People spontaneously took the perspective of the other person and made faster judgements about the presented items in their presence if the characters were oriented towards upright to them. The gaze direction of this other person did not influence these effects. We propose that visual perspective taking is therefore a general spatial-navigational ability, allowing us to calculate more easily how a scene would (in principle) look from another position in space, and that such calculations reflect the spatial location of another person, but not their gaze.

Keywords: *visual perspective taking; perceptual simulation; gaze cuing; navigation; mental rotation; mental imagery*

Introduction

Visual perspective taking (VPT) allows people not only to derive what others can see, but also how they see it (Flavell, Everett, Croft, & Flavell, 1981). Doing so helps us navigate our social environment (Kozhevnikov, Motes, Rasch & Blajenkova, 2006), engage in joint action (Freundlieb, Kovács, & Sebanz, 2016) and work out what others will do next (Bach, Fenton-Adams, & Tipper, 2014; Creem-Regehr, Gagnon, Geuss, & Stefanucci, 2013; Kovacs, Teglas, & Endress, 2010). Deriving how the world appears to other people may also form a foundation for more sophisticated abilities to reason about others mental states, such as their beliefs, emotions, and goals (Batson, Early, & Salvarani, 1997; Erle & Topolinski, 2015; Tomasello, Carpenter, Call, Behne, & Moll, 2005; Mattan, Rotshtein & Quinn, 2016). Yet, the mechanisms underlying the ability to perspective take, and the stimulus features that trigger shifts to another's perspective are largely unclear.

Recent research has started to conceptualize perspective taking as a form of perceptual simulation, which "paints" or "inserts" another's view of the world onto one's own perceptual processes, as if it were one's own perceptual input (Kampis, Parise, Csibra, & Kovács, 2015; Surtees, Apperly, & Samson, 2013; Ward, Ganis, & Bach, 2019). As soon as another's perspective is represented in such a (quasi-)perceptual manner, it could drive one's own action and decision making processes just like own input, explaining the developmental link between VPT and higher level mentalizing abilities (Hamilton, Brindley, & Frith, 2009) and its link to joint action (Freundlieb, Kovács & Sebanz, 2016; Tversky & Hard, 2009).

Direct evidence for such a proposal had been lacking however. Consistent with a similar representational format of one's own and others' perspectives, it has been shown that own judgments of a stimulus are harder if another person around us would make the same judgments differently. People find it harder, for example, to judge how many objects they can see from their own perspective when another actor sees a different number (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010). Similarly, it is harder to identify a character as a “6” if it looks like a “9” to another person (Surtees, Samson, & Apperly, 2016), or to judge whether an object is to the left or the right to oneself when this object is in another location for another person (Tversky and Hard, 2009; Zwickel, 2009; Zwickel and Müller, 2010). Yet, these studies leave open whether the interference happens on the perceptual, conceptual, response or even metacognitive level, perhaps indexing simply the uncertainty when a person becomes aware that others would judge – or respond to – the same stimulus differently from another perspective, without representing how specifically they see it.

A recent series of studies from our lab (Ward et al., 2019) provided the first direct evidence that another’s visual perspective is perceptually represented. We reasoned that if people have (quasi-)perceptual access to the content of another’s' perspective, then this perspective should not only interfere with one’s perceptual judgements (Sampson et al., 2010; Surtees et al., 2016), but should also *facilitate* them, precisely if this other perspective would offer a better view on an object than one has oneself. To test this, we let participants take part in a variant of the classic mental rotation task (Shepard & Melzer, 1971) in which they simply judged whether alphanumeric characters presented at varying orientations were ‘normal’, or mirror-inverted (e.g. “R” v. “Я”). The classic finding is that these judgements become progressively slower with increasing angular disparity away from upright (Cooper, 1975; Shepard & Melzer, 1971) because people must mentally rotate each item back into its canonical orientation before judging it. We find that this classic pattern can be dramatically altered simply by inserting another person into the scenes that views the items from another perspective. Items oriented away from participants were identified more quickly if they appear upright to the other person, as if people could simply identify it from their alternative perspective. Conversely, they judged it more slowly if the item was even further rotated away from the other person. Regression analyses showed that recognition times across letter orientations can be predicted from the angular disparity of the item not only to the participant, but also to the other person, suggesting that participants mentally rotated the items from both their own and the other's perspective.

These data showed that people can represent the *content* of another's (imagined) viewing perspective in a form that can drive perceptual judgements in the same bottom-up manner as our own perception (e.g. Roelfsema & de Lange, 2016), allowing us to make judgments from others' perspectives that would be more difficult from our own. Importantly, these shifts to the others' perspective occurred spontaneously, even when the persons in the scene were completely task-irrelevant. They were enhanced when participants were explicitly instructed to take the others' perspective, but were eliminated when the other person was replaced by an inanimate object (Ward et al., 2019), suggesting that these perceptual shifts capture another's ability to see (or at least mentally represent) the world.

This paradigm provides us with a unique window into the process that underlie visual perspective taking. Here, we use this task to test which cues are used to drive spontaneous shifts into another's visual perspective. One important candidate is another person's gaze. Gaze has long been associated with mentalizing and perspective taking, exemplified perhaps by the folk-psychological interpretation of perspective taking as "seeing through another's eyes". People are highly sensitive to others' gaze, to the extent that one's own spatial attention is involuntarily directed to the object another person is looking at, creating shared attention (for a review, see

Frischen, Bayliss and Tipper, 2007). Importantly, this effect is at least partially driven by mental state attributions to the gazing person (Morgan, Freeth & Smith, 2019), and seeing others gaze at an object transfer their emotional state (e.g. a smile or frown) to it, rendering it more/less appealing (e.g., Bayliss, Paul, Cannon & Tipper, 2006; for review, Becchio, Bertone & Castiello, 2008). In addition, a person's gaze direction signals how they intend to act upon an object (Ambrosini, Costantini & Sinigaglia, 2011), promotes later (automatic) imitation of these actions (Wang & Hamilton, 2012), and increases activity in brain regions related to Theory of Mind and mentalizing (e.g., Mosconi, Mack, McCarthy & Pelphrey, 2005; Williams, Waiter, Perra, Perret & Whiten, 2005). If the associated processes rely on embodying the content of another's visual perspective, then visual perspective taking, too, should be highly sensitive to where another is looking.

Another possibility is however that spontaneous visual perspective taking is separate to these mentalizing processes and primarily driven by the presence – and spatial location – of another person in the scene. In other words, people might be putting themselves generally into the shoes of others, without truly looking through their eyes. It is well supported that people embody another's location in space. Several studies have shown, for example, that it takes longer to explicitly judge where an object is from another's perspective the more this person is oriented away from us, as we mentally rotate our own bodies into the other's perspectives, with this transformation taking longer the more we are oriented away from them (Kessler & Thompson, 2010; Kessler & Rutherford, 2010; Kozhevnikov et al., 2006; Surtees et al., 2013). Other studies link perspective taking not to Theory of Mind and mentalizing, but to more fundamental navigational abilities (Allen, Kirasic, Dobson, Long & Beck, 1996; Hegarty & Waller, 2004; Kozhevnikov et al., 2006) and the ability to decide from which other locations one can interact with the environment (Gunalp, Moossaian & Hegarty, 2019; Kessler & Thompson, 2010).

Strikingly, there is very little evidence about whether another's eye gaze is an important cue for taking another's visual perspective. Most studies have demonstrated only that another person's eyes contribute to working out *what* this person can or cannot see (Level 1 perspective taking), not *how* they see it (Level 2 perspective taking). These studies have shown that this form of Level 1 perspective taking relies on a process that "mentally scans" (Kosslyn, Ball, & Reiser, 1978) the line of sight between another's eyes and the object they can potentially see (e.g., Michelon & Zacks, 2006) and that, based on this information, people work out whether their own judgment and others differs (e.g., Surtees et al., 2013). To our knowledge only two studies have investigated how others' eye gaze contributes to Level 2 perspective taking, people's ability to work out *how* objects look to the other person, with both providing non-conclusive results. Mazzarella, Hamilton, Troiano, Mastromauro and Conson (2012) found that seeing another person reach for an object induces spatial interference effects from the others' perspective, so that participants were less likely to judge object location from their own perspective. However, merely seeing the other person shift their eyes towards the object (and not reaching for it) did not induce a detectable shift to their perspective. Similarly, Furlanetto, Cavallo, Manera, Tversky and Becchio (2013) showed that while viewing someone gaze at an object numerically increased shifts to their spatial perspective, this did not seem to be statistically robust, and even removing gaze cues altogether (by blurring the eye region) did not affect whether people judged objects from another's or their own perspective.

Our paradigm provides an ideal means to disentangle these two possibilities. By measuring how much faster items are identified the more they are oriented towards other people in the scenes (and slower the more they are oriented away from them), our paradigm provides a quantitative measure of participants' tendency to spontaneously take another's visual perspective. If another's gaze is critical to trigger these VPT processes, then these recognition biases should only be seen for an actors that gazes at the items but be eliminated if the actor is not looking directly at them, but gazes elsewhere. If, on the other hand, VPT represents a mental transformation from one's own position in space into that of another, then the gaze direction of this other person will not influence recognition times, and *all* conditions will show shifts towards the incidentally presented people in the scenes, even if they gaze away from the relevant items. Here, we test these predictions in two experiments.

3.2 Experiment 4a – manipulating gaze between trials

Experiment 1 tests how another's gaze affects perspective taking, using a within-subject design. Participants were shown a series of scenes depicting a table in front of them. In each trial, an alphanumeric character appeared on the table, presented either normally, or in its mirror-inverted form, at varying rotations away from upright. Participants had to simply judge, with a button press, whether the items were mirrored or normal ("R" vs. "Я"). In some trials, another person was seen sitting to the left or the right of the table. Our previous studies have shown that people spontaneously make use of another person's perspective when making these judgements. When letters are facing another person, it was easier to judge whether they are mirrored or normal than it would be if this other person was not there or sitting in the opposite direction, indicating that people judged the letters through this person's eyes. Here we manipulated – between trials – the direction in which the person was looking, in order to test whether people's tendency to spontaneously represent how an object looks to another person is affected by whether this person is actually looking at the object. In different trials, the person on the screen could either (1) look at the item on the table, as in our prior work, or (2) look away from it to the wall in the room. If perspective taking takes the others' gaze into account, then the ability to recognize an object from another's perspective should be enhanced – or only found – when the person is actually looking at the to-be-judged item, but less so if looking at the participant or the room's back wall. If, however, perspective taking emerges from a general rotation of one's own body into the spatial location of another then it should happen irrespective of where the actor is looking.

3.2.1 Method

Participants

Thirty-seven (31 females) naive participants were initially recruited via the University of Plymouth participation pool. All participants were adults (age range 18-35) and gave written informed consent according to the declaration of Helsinki. Approval was obtained from the University of Plymouth Ethics Committee. Participants received course credit as compensation. As in Ward et al., (2019), participants with more than 20% error rates across all conditions were not considered for analysis (n=5). The remaining thirty-two participants (27 females; mean age: 20.1 years, range: 18-35) provide enough power to detect effects in the range of d =.45. Prior work on this paradigm (Ward et al., 2019) has revealed that effect sizes are substantially larger (.747 < d < 1.08 for the main perspective taking effect).

Apparatus, stimuli and procedure

The study was conducted in behavioural testing lab space of the Action Prediction lab of the University of Plymouth. It was administered using Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Stimuli were presented on a 19” LED computer monitor (Resolution: 1900x1200; Refresh rate: 60Hz). Responses were made on a standard computer keyboard with UP, DOWN, and SPACE keys as active response keys. Red and green stickers were positioned on the DOWN and UP keys, respectively.

Participants sat upright facing the screen at a distance of approximately 60cm and were given written and verbal instructions. Before the experiment proper commenced, they were given examples of the rotated items that would appear on the screen and completed eight training trials that were identical to the main experiment. The experiment proper consisted of 520 trials. Each trial (Figure 2.1) started with a fixation cross displayed for 400ms, followed by 300ms blank screen. The subsequent stimulus sequence included three frames, each measuring 33.4 by 23.5 degrees of visual angle. The first frame was presented for 800 ms. In one third of the trials, it showed a view onto a corner of a square table in a grey room without a person being present (No-Person trials). The remaining trials showed a person (either male or female) sitting at the same square table, gazing outwards at the participant. The person sat either on the left side of the table (Person-left trials) or on its right side (Person-right trials), so that they were positioned at roughly 270 or 90 degree angle to the position at which the rotated items would appear, relative to the participant.

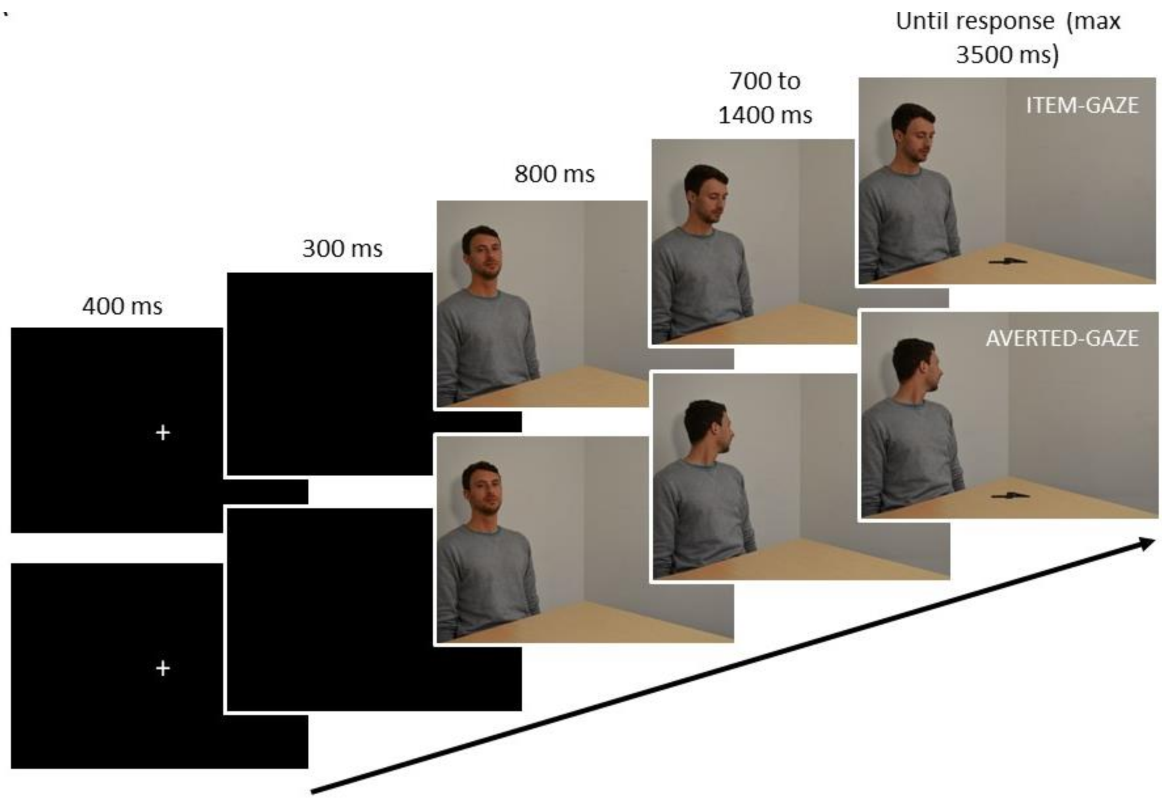

Figure 2.1. **Example stimuli used in Experiment 4a.** Looking direction (Item-gaze or Averted-gaze) varied between trials.

The second frame in the sequence was identical to the first for the No-Person trials. In the other trials, the person looked down towards the middle of the table in half the trials, where the item would appear, and in the other half they looked away towards the corner of the room. Because this second frame was presented without inter-stimulus interval, it created the impression of apparent motion (Wertheimer, 1912), so that the person seemed to shift their gaze from the participant towards the table or towards the back of the room. Then, after a random interval between 700ms and 1400ms, the third frame was presented. This was identical to frame 2, but now one of 48 possible items appeared on the table, at the location on the table the on-screen person was gazing at. This item was one of three alphanumeric characters (4, P, or R), presented either in the canonical version or mirror-inverted about their vertical axis, in one of eight orientations (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°, with 0°

denoting the upright canonical orientation and angles increasing in a counter-clockwise fashion) relative to the participant. The characters always appeared in the same position on the table, half-way between the outward corner of the table and its centre, such that the persons to the left and right would gaze at the table from roughly 90° and 270°, respectively (perpendicular to the viewpoint of the participant). Character rotation occurred around the character's centre point.

Participants were asked to judge whether each character was presented in its canonical or mirror-inverted form (e.g. "R" vs "Я"). Participants responded using their right hand by pressing the green key to indicate a canonical item and the red key to indicate a mirrored item. Response times were measured relative to item onset. The third frame remained on the screen until a response was made to a maximum duration of 3500ms.

Analysis

Data (pre-)processing and analysis was identical to Ward et al., (2019) and conducted in Microsoft Excel (2010) and JASP (2018). Violin plots were created using Raincloud Plots (Version 1; Allen, Podiaggi, Whitaker et al., 2019). Power analyses were conducted in G*Power (Version 3.1; Erdfelder, Faul, & Buchner, 1996).

Dependent measures were the recognition times (measured from item onset) for each character orientation (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315), depending on Person location (No-Person, Person-left, Person-right) and Gaze direction (Item-gaze, Averted-gaze). As in our original study (Ward et al., 2019), the recognition times for each angle and condition were mapped onto two orthogonal and statistically independent summary measures, so that they can be compared across conditions, without accruing alpha inflation due to multiple testing, which would result if each of the eight angles were compared separately. These two summary measures are derived by treating each participant's recognition time for this character orientation as a vector in a polar coordinate system, with the recognition time providing the distance from the origin and the rotation angle the polar angle.

The first summary measure (Toward/Away-bias) quantifies the mental rotation effect (Shepard & Meltzer, 1971) and indexes to what extent characters are recognized faster the more they are facing towards the participants (0°) rather than away from them (180°), separately for each participant and condition (No-Person, Person-left Item-gaze, Person-left Averted-gaze, Person-right Item-gaze, and Person-right Averted-gaze). A character orientation's contribution to the Toward/Away-bias is derived simply from the recognition times multiplied with the negative of the cosine of the orientation angle. As a result, characters contribute negatively the more they face the participants (315°, 0°, 45°) and positively they more they are oriented away from them (225°, 180°, 135°). Positive values on this score therefore indicate faster recognition of letters oriented towards rather than away from the participant.

The second summary measure (Left/Right-bias) indexes how much faster characters were recognized when oriented towards the left (270°) rather than right (90°), and therefore allows us to quantify whether people present in these location induce shifts in how quickly these letters can be identified. It was derived analogously to the Towards/Away-bias. The contribution of a character's orientation to the Left/Right-bias was calculated as the recognition time multiplied with the sine of the orientation angle. Character orientations contribute negatively the more they face to the right (45°, 90°, 135°) and positively the more they face to the left (225°, 270°, 315°). A zero value on the Left/Right-bias therefore indicates that items oriented left and right are identified equally quickly, while positive numbers indicate faster identification for left- compared to right-oriented letters, and vice versa for right-oriented letters.

By averaging these values, separately for each summary measure, participant and condition (No-Person, Person-left-item, Person-left-averted, Person-right-item, Person-right-averted), we can calculate, first, whether characters are recognized faster the more they appear in the canonical orientation to the participant compared to when they are oriented away (positive values on the Toward/Away-bias), reflecting the expected mental rotation effect. Similarly, they allow us to calculate to what extent characters are recognized faster the more they are oriented leftwards and would appear in their canonical orientation to a person sitting to the left (positive values on the Left/Right-bias) rather than rightwards, appearing in their canonical orientation to a person sitting on the right (negative values). We were then able to determine if this difference in left/right biases changed depending on whether the other person was looking at the object or not. Note that the direct comparison of the Person-left and Person-right conditions is statistically identical to the comparison of how much the presence of person shifts mental rotation performance in the Person-left and Person-right conditions relative to the No-Person baseline (i.e. how much person presence shifts recognition times away from 0 towards either 90^0 or 270^0), as this would involve subtracting the same baseline value from each of the two conditions for each participant, and would therefore not affect the absolute difference between them.

Across-participant regression analyses

In prior work, the mental rotation effect is typically also characterised in terms of a linear regressions of an items' recognition time to its angular disparity relative to the participant, for each participant separately (Shepard & Meltzer, 1971). The results reveal linear increases with increasing angular disparity for the large majority of participants. Here, we used this analysis model to test whether an item's recognition times can be described, on a single participant basis, as a linear increase of the character's angular disparity *both* to the participant and to the other person. To this end, we entered each participant's item mean recognition times for each character orientation in each condition (No-Person, Person-item, Person-averted) as dependent variable in a multiple regression analysis, with the item's angular disparity to the participant and to the other person as two statistically independent predictors. This analysis provides regression coefficients for both predictors – angular disparity to participant and other person – for each participant and condition separately. We report mean across-participant regression coefficients for each of these two predictors and compare them with t-tests against zero, and against each other.

3.2.2 Results

Erroneous responses (7% on average) were excluded from the analysis of recognition times (RTs), as well as trials with RTs longer than 2000ms, or shorter than 150ms.

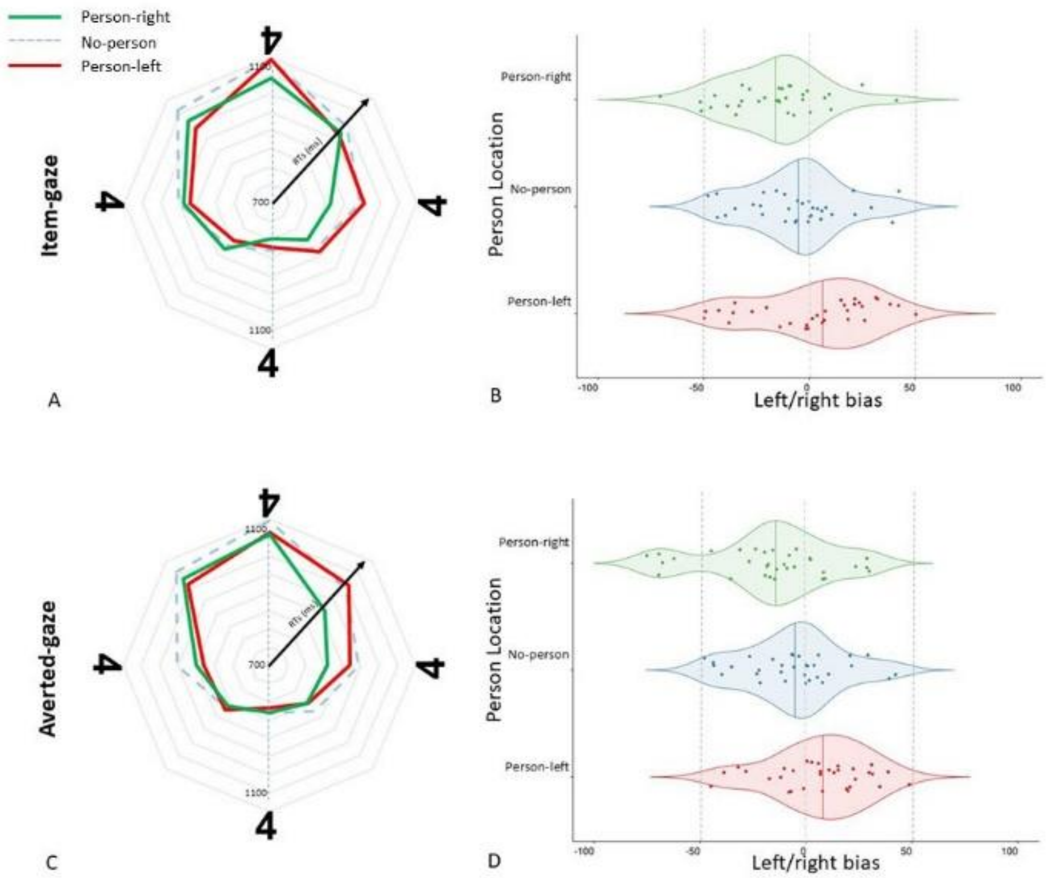

Figure 2.2. Results of Experiment 4a. **(A)** Mean recognition times (ms) for mirror-inverted/canonical judgements for Item-gaze trials when the actor is inserted on the right and the left showing faster recognition times for items oriented upright to the actor. (B) Left/right bias when the inserted person is looking at the table showing faster recognition times for items appearing upright to the actor. (C) Mean recognition times (RTs) for mirror-inverted/canonical judgements for Averted-gaze trials (D) Left/right bias when then inserted person appears to look away from the table (towards to wall).

Mental rotation

We first determined whether our data replicated the classical mental rotation effect (Shepard & Metzler, 1971). To this end, we first compared the overall (across conditions) towards/away bias, indexing, in milliseconds, how much more slowly items are identified the more they are oriented away from the participant compared to towards them, with a simple t-test against zero. The towards/away bias across all conditions was positive, *M*=56.34; *SD*=28.05, *t*(31)=11.36, *p*<.001, *d*=2.01, BF_{10}=5.797e+09, showing – unsurprisingly – that items are identified more quickly the more they are oriented towards rather than away from participants. This mental rotation effect was also confirmed by regressing each item's recognition time to the expected linear increase with angular disparity, as in prior research, (Shepard & Metzler, 1971; Ward et al., 2019), revealing positive slopes in all participants, mean β =1.4; *t*(31)=14.60, *p*<.001, *d*=2.6, BF_{10}=3.181e+12.

We then verified that this overall mental rotation effect was not affected by person presence. As the actors would sit at 90 and 270 degree angle to the participant, and their location was therefore orthogonal to the towards/away axis, we did not expect that the presence/location of the other person would affect the overall mental rotation effect. Indeed, a 2x3 ANOVA on the towards/away-biases across conditions with the factors Gaze (item, averted) and Location (person left, person right, no person) did not reveal any main effects or interactions, *F*<2.68, *p*>.08, for all. When the six conditions were analysed separately, slower recognition of turned away items was present in all conditions, *t*(31) > 8.29, *p* <.001, *d* >1.5 for all.

Perspective taking

The main question was whether people would spontaneously take the perspective of the other person in the scenes, such that items were recognized faster when oriented towards compared to away from them, and whether this effect, in turn, was determined by where this person was looking. We therefore calculated, for each participant and condition separately, the left/right-bias (for details, see analysis-section), indexing, in ms., how much faster items were identified the more they were oriented to the left than to the right. The resulting left/right-bias summary scores were then entered into a 2 x 2 repeated measures ANOVA with within-subjects factors of Location (Person-left, Person-right) and Gaze (Averted-gaze, Item-gaze). This analysis revealed decisive evidence for a main effect of Location, $F(1,31)=27.921$, $p<.001$, $\eta p^2=.474$, $BF_{10}=4379.04$. As in our prior work (Ward et al., 2019), left/right-biases were more negative (indexing faster recognition of rightwards- than leftwards oriented letters) when someone was sitting on the right, and more positive (indexing faster recognition of leftwards- than rightwards oriented letters) when someone was sitting on the left. Strikingly, this analysis provided moderate evidence against a main effect of Gaze, $F(1,31)=.545$, $p=.466$, $\eta p^2=.017$, $BF_{10}=.203$, and against an interaction of Location and Gaze, $F(1,31)=.291$, $p=.593$, $\eta p^2=.009$, $BF_{10}=.216$, showing that the location of the actor, and not where they were looking, was the critical cue to induce perspective taking.

Table 2.1 Means (M) and Standard Deviations (SD) for the Left/Right- and Towards/Away-biases in Error rates in Experiment 4a and Experiment 4b. Relates to Figure 2.1 and Figure 2.4. Forward/Away and Left/Right-biases were calculated analogously as for the recognition times. *p<.05. **p<01, ***p<.005.

		Toward/away bias			Left/right bias		
Exp.	Looking direction	Person-left M (SD)	Person-right M (SD)	No-Person M (SD)	Person-left M (SD)	Person-right M (SD)	No-Person M (SD)
1.	Item	.021 (.027)***	.022 (.027)***	.021 (.024) ***	-.0001 (.024)	-.005 (.02)	.0004 (.02)
	Averted	.024 (.028) ***	.025 (.026)***	.021 (.024) ***	.003 (.021)	-.006 (.022)	.0004 (.02)
2.	Item	.028 (.026) ***	.024 (.026) ***	.022 (.03) ***	.004 (.011)	-.001 (.017)	.004 (.015)
	Averted	.028 (.027) ***	.031 (.032) ***	.022 (.025) ***	-.0019 (.014)	-.0023 (.015)	-.003 (.016)
	Participant	.034 (.029) ***	.025 (.024) ***	.033 (.022) ***	-.007 (.02)	-.01 (.021)	-.005 (.019)

Indeed, when the difference in left/right-bias between Person-left and Person-right conditions was compared separately for each gaze condition, the predicted differences were apparent for both Item-gaze trials $t(31)=-3.287$, $p=.003$, $d=.58$, $BF_{10}=14.55$, and Averted-gaze trials $t(31)=5.044$, $p<.001$, $d=1.01$, $BF_{10}=1145.37$, with effect sizes, if anything, being larger in the away trials.

Regression analyses

As in our previous studies, we tested whether recognition times could be described as independent linear increases depending on an item's angular disparity to the participant as well as the other person, by using both disparities as orthogonal predictors in a simple regression model, for each participant and condition separately (and then comparing them against zero). Overall, these revealed very strong evidence for independent contributions of both the angular disparity to the participant, mean β =1.39, $t(31)=11.09$, $p<.001$, $d=1.96$, $BF_{10}=3.278e+09$, and to the other person, mean β =.27, $t(31)=4.95$, $p<.001$, $d=.87$, $BF_{10}=919.8$, showing that recognition times can be described by independent mental rotation functions form one's own and the other person's perspective.

To test how these relationships were affected by the other person's gaze, regression coefficients for each participant and condition were entered into a 2 x 2 repeated measures ANOVA with Gaze (Item, Averted) and Perspective (Self, Other) as within-subject factors. This analysis only revealed a main effect of Perspective, $F(1,31)=68.96$, $p<.001$, $\eta p^2=.69$, $BF_{10}=2.118e+18$ showing that, as expected, angular disparity towards the participants determined recognition times to a stronger extent that angular disparity to the other person. Replicating the main analysis, the analysis provided moderate *against* a contribution of gaze, neither revealing a main effect of Gaze, $F(1,31)=.267$, $p=.609$, $\eta p^2=.009$, $BF_{10}=.24$, nor an interaction of Gaze and Perspective, $F(1,31)=.044$, $p=.835$, $\eta p^2=.001$, $BF_{10}=.193$. In Item-gaze trials, both the angular disparity away from upright to the participant, mean $\beta=1.39$, $t(31)=9.55$, $p<.001$, $d=1.7$, $BF_{10}=1.044e+08$, and to the actor, mean $\beta=.25$, $t(31)=3.273$, $p=.003$, $d=.58$, $BF_{10}=13.98$ contributed to recognition times. This pattern was the same in the Averted-gaze trials, with the stronger contribution again coming from the angular disparity to the participants, mean $\beta=1.42$, $t(31)=10.81$, $p<.001$, $d=1.9$, $BF_{10}=1.799e+09$, and the smaller contribution coming for the angular disparity to the other person in the scene, mean $\beta=.30$, $t(31)=5.25$, $p<.001$, $d=.91$, $BF_{10}=2002.84$.

3.3 Experiment 4b – manipulating gaze between participants

Experiment 4b confirmed that the mere presence of a person is sufficient to induce a shift into their perspective. As in our prior research (Ward et al., 2019), we found that oriented items were recognized more rapidly if they appeared closer to their canonical orientation to another person in the scenes. Moreover, recognition times could be well-described by mental rotation functions centred on both one's own and the other person's perspective. Strikingly, however, these shifts into the others' perspective were independent of the other person's gaze and occurred irrespective of where they are looking at the to-be-recognized item or away from it. This suggests that taking another's visual perspective is based on their location in space, but not their gaze at (or away) from the relevant items.

One reason for this lack of a difference might be that within-participant experiments such as ours suffer from influential companion (e.g., Poulton, 1992) and carry-over effects (MacFie, Bratchell, & Greenhoff, 1989), for example when a participant's memory of the actor's gaze (e.g. at the item) in a previous trail influenced their processing in the current trial (e.g. at the wall). In Experiment 2*,* we therefore varied gaze *between* participants to rule out such an explanation. In addition, we added a third condition, in which the on-screen actors looked outwards at the participant. Prior work suggests that such gaze at the participant plays a special role in social cognition. It serves as an alerting signal, captures attention, and eliminates gaze cuing altogether (e.g. Mazzarella et al. 2012; for a review, see Hamilton, 2016). It may therefore provide the perhaps strongest manipulation of gaze. If visual perspective taking takes into account what another actually sees through their ideas – rather than what they *could* see from their position in space – then it should now be disrupted.

3.3.1. Method

Participants

Ninety-one (61 females) naive participants were initially recruited via the University of Plymouth participation pool. All participants were adults (age range 18-35) and gave written informed consent according to the declaration of Helsinki. Approval was obtained from the University of Plymouth Ethics Committee. Participants received course credit as compensation. After exclusion (>20% error), 80 participants (51 females; mean age: 19.9 years, range: 18-30) were considered for analysis, with at least 26 participants for each condition. This provides enough power to detect effects in the range of d = .5 for our critical comparison of left/right-bias across person locations. Results of the previous study (Ward et al., 2019) and as well as Experiment 1 show that effects are likely larger.

Apparatus, stimuli and procedure

Participants completed 584 trials following the same apparatus and procedure as in Experiment 1. Between participants, we varied whether the actor always looked out towards the participant (participant-gaze), looked directly at the item on the table (item-gaze), or looked away from the item towards the corner of the room (averted-gaze). In each group, an equal number (1/3rd) of trials showed a person on the left, a person on the right, or no person. The starting frame in each trial showed the table, with either the person present or absent. Depending on the participant's group allocation, the person present in the scenes either looked directly at the future item location, at the corner of the room or outwards towards participant. This frame was displayed for a randomized time between 1500 and 2220ms. It was directly followed by the appearance of the item on the centre of the table for 3500ms (while the person remained present), or until a response was made (Figure 2.3).

3.3.2. Results

As in Experiment 1, erroneous responses (10% on average) were excluded from the analysis of recognition times (RTs), as well as trials with RTs longer than 2000ms, or shorter than 150ms.

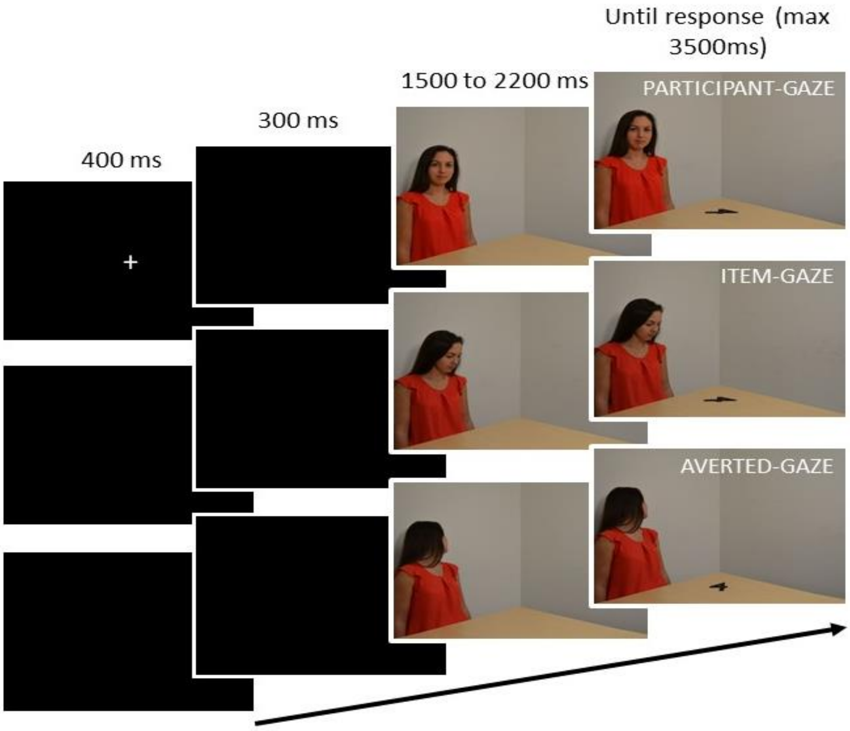

Figure 2.3. **Example stimuli used in Experiment 4b.** Participants viewed only one of the three possible looking directions in a between-subjects design.

Mental rotation

We again first verified that our data would replicate the mental rotation effect (Shepard & Metzler, 1971). As in Experiment 1, a simple t-test comparing toward/away bias across all conditions against zero confirmed that recognition times were generally slower the more items were rotated away from participants, $t(79)=20.69$, $p<.001$, $d=2.31$, $BF_{10}=2.223e+30$. This mental rotation effect was also confirmed by regressing each item's recognition time to the expected linear increase with angular disparity, revealing positive slopes in all bar one participant, mean $\beta = .85$; $t(79)=54.96$, $p<.001$, $d=6.15$, $BF_{10}=9.342e+60$.

To confirm that this overall mental rotation effect was not affected by person presence, we ran a 3x3 ANOVA on the towards/away-biases across conditions with the factors Gaze (item, averted, participant) and Location (person left, person right, no person) as in Experiment 1. As before, this did not reveal any main effects or interactions, $F(2/4,154/77)<2.47$, $p>.088$, $BF<.404$ for all. The mental rotation effect was present in all conditions, $t(25/26) > 9.17$, $p <.001$, $d > 1.8$, for all.

Perspective taking

We then tested whether the task would replicate the shift to the other person's perspective and whether this shift was modulated by their gaze. Left/right-biases from all participants were entered into a 2 x 3 repeat measures ANOVA with within-subjects factors of Location (Person-left, Person-right) and between-subject factors of Gaze (Averted-gaze, Item-gaze, Participant-gaze). It again revealed a main effect of Location, $F(1,77)=29.472$, $p<.001$, $\eta p^2=.277$, $BF_{10}=15474.14$. As can be seen in Figure 2.3, left/right-biases were more negative when someone was sitting on the right and more positive when someone was sitting on the left. Thus, letters were identified more quickly when pointing leftwards when someone was also sitting on the right, but identified more quickly when oriented rightwards when someone was sitting on the left. However, as in Experiment 1, there was neither a main effect of Gaze, $F(2,77)=1.189$, $p=.310$, $\eta p^2=.030$, BF_{10} =.230 , nor an interaction of Location and Gaze, $F(2,77)=2.11$, $p=.128$, $\eta p2=.052$. BF_{10} = .554. Pairwise comparisons of the critical Left/Right actor locations for all conditions confirmed that regardless of where the actor was looking, recognition times were faster when items that would usually be difficult to recognise were easier when appearing upright from this person's position, $t(26)=2.347$, $p=.027$, $d=.5$, $BF_{10}=2.055$, $t(26)=2.730$, $p=.011$, $d=.54$, $BF_{10}=4.239$, and $t(25)=4.173$, $p<.001$, $d=.89$, $BF_{10}=94.93$ for Item, Participant, and Averted-gaze, respectively.

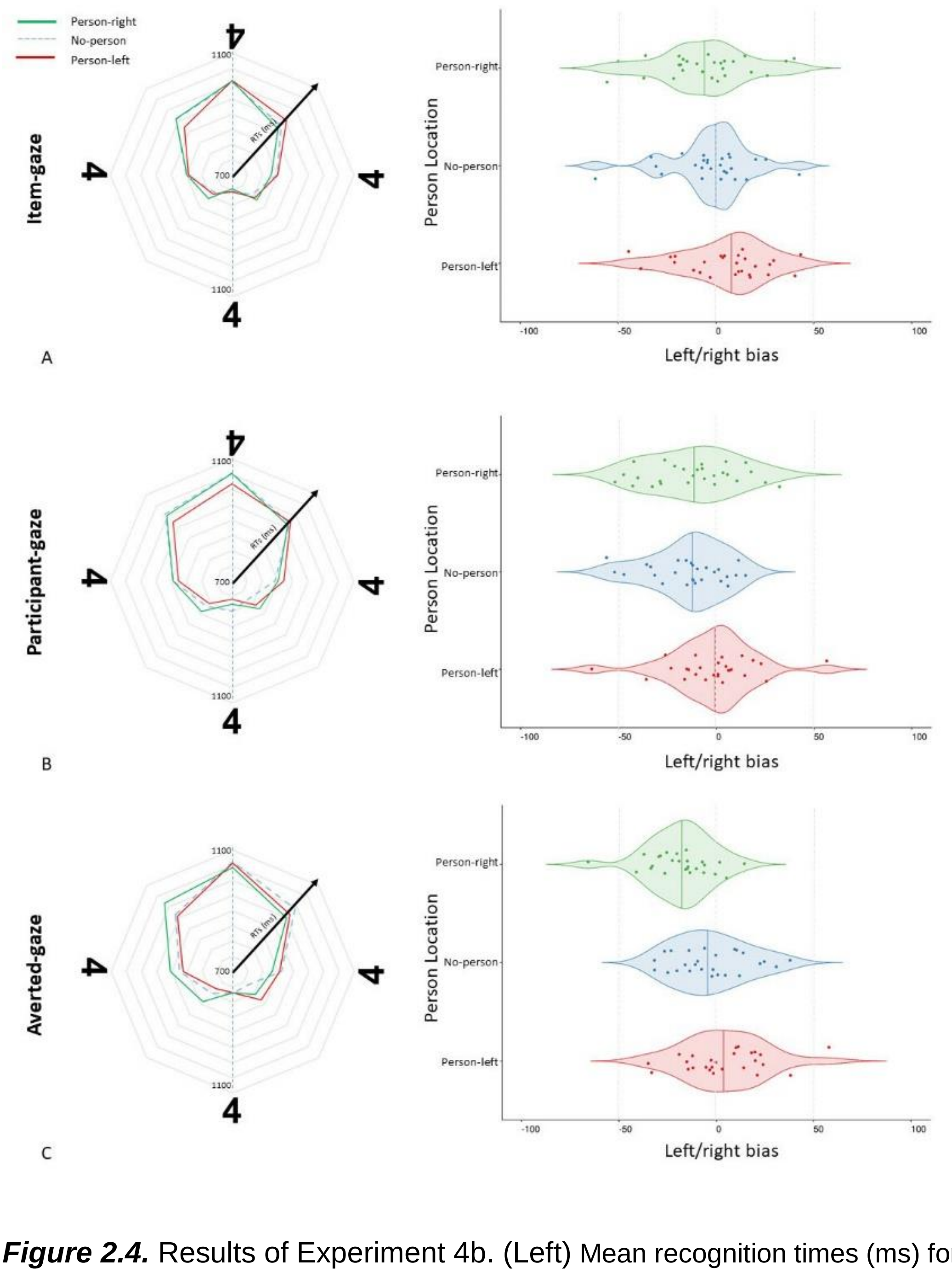

Figure 2.4. Results of Experiment 4b. (Left) Mean recognition times (ms) for mirror-inverted/canonical judgements and (Right) Left/right bias when the inserted person is looking at the table showing faster recognition times for items appearing upright to the actor for (A) Item-gaze, (B) Participant-gaze, and (C) Averted-gaze.

Regression analyses

As in Experiment 1, we tested whether recognition times can be described as independent linear increases depending on an items angular disparity to the participant as well as the other person. Mean beta coefficients for all gaze conditions were entered into a simple regression model, with the angular disparity of the character relative the actor and the participant as independent predictors of recognition times. To test for differences between gaze conditions, mean Beta for all participants were entered into a 3x2 ANOVA with the between-subjects factor, Gaze (item, participant, averted), and within-subjects factor, Perspective (own, other). As in Experiment 1, a 2 x 3 ANOVA revealed only a main effect of Perspective, $F(1,77)=285.99$, $p<.001$, $\eta p2=.788$, $BF_{10}=1.587e+34$, showing that the angular disparity of an item to one's own perspective determined recognition times to a stronger extent than to the other's perspective. There was no main effect of Gaze, $F(2,79)=.231$, $p=.794$, $BF_{10}=.099$, and no interaction of Perspective and Gaze, $F(2,77)=.23$, $p=.795$, $\eta p2=.006$, $BF_{10}=.130$. Step-down analyses confirmed that, in each condition, recognition times could be described by a linear combination of an item's angular disparity to the participant (Item-gaze, $\beta=1.45$, $t(26)=10.207$, $p<.001$, $d=1.95$, $BF_{10}=7.276e+7$; Averted-gaze, $\beta=1.57$, $t(26)=13.697$, $p<.001$, $d=2.64$, $BF_{10}=6.344e+7$; Participant-gaze, mean $\beta=1.63$, $t(25)=10.352$, $p<.001$, $d=2.03$, $BF_{10}=3.399e+10$) and to a lesser extent to the other person (Item-gaze, mean $\beta=.13$, $t(26)=2.234$, $p=.034$, $d=.429$, $BF_{10}=1.684$; Averted-gaze, $\beta=.13$, $t(26)=2.699$, $p=.012$, $d=.52$, $BF_{10}=68.782$; participant-gaze, $\beta=.28$, $t(25)=4.032$, $p<.001$, $d=.79$, $BF_{10}=3.982$).

3.3.3 General Discussion

Using our recently developed task (Ward et al., 2019), we tested whether spontaneous shifts into another's visual perspective depend on the other person looking at the to-be-judged item. We replicated the finding that people spontaneously represent – perceptually simulate – the content of another's viewing perspective and use it to recognize items that would be more difficult to recognize from their own perspective. Thus, participants identified leftwards oriented items oriented more quickly when another person was sitting to the right than on the left, and vice versa for rightwards oriented items. Moreover, as in our original study (Ward et al., 2019), recognition times were well described by mental rotation functions based on the item's orientation away from the participant and the other person. Our results therefore confirm that participants spontaneously shift their perspective into the location of another person and represent how an item would look from this position. Once represented in such a quasi-perceptual manner, the content of another's viewing perspective can therefore drive perceptual decision making in a similar manner as one's own perceptual input, and drive mental rotation processes.

Here we asked whether these spontaneous acts of visual perspective taking only reflect the other person's location in space or whether they are also driven by the person's gaze at the task-relevant objects (or away from them). We therefore manipulated, in two experiments, whether the other person was directly looking at the item, towards the wall, or towards the participant (Experiment 2 only). Neither experiment revealed an influence of gaze on visual perspective taking. The signature perspective taking effect – that items are identified more quickly when appearing upright to another person – was robustly present irrespective of whether this person was looking at the item, towards the wall, or directly at the participant, and irrespective of whether looking direction was manipulated between-trials in Experiment 1 and between participants in Experiment 2, preventing any obscuring influence of carry over or influential companion effects (MacFie et al., 1989; Poulton, 1982).

These data show that spontaneous visual perspective taking is primarily driven by the presence of another person and their location in space, but does not take into account where they are looking. This insensitivity to another's gaze is surprising under the (folk-psychological) assumptions that visual perspective taking is based on a sophisticated mechanism that allows us to "see" through another's eyes (e.g., Furlanetto, Becchio, Samson, & Apperly, 2016). It may also be surprising given the crucial role others' gaze is assumed to play in social perception generally. Viewing another's gaze provides immediate information about *whether* another person can see an item (e.g., Gardner et al., 2018), shifts one's own spatial attention towards the gazed at object (Frischen et al., 2007), and transfers the actor's emotional evaluation (e.g., when smiling, frowning) to the object so that one's own and others judgments align (Becchio et al., 2008). Our data suggest that these effects, insofar as they imply any sort of mental state attribution, are separate from the perceptual shifts into the eyes of the other person, and therefore reflect different mechanisms (see Cole, Atkinson, Le, & Smith, 2016, for a similar argument for gaze cuing).

Instead, the reliance on another's position in space suggests that spontaneous visual perspective taking relies on a more general mechanism that calculates how an item would, in principle, look from another's vantage point, irrespective of whether this person is currently looking at the item. It is consistent with the observation that visual perspective taking, when explicitly instructed, relies on mentally rotating one's own body into the position of the other person (e.g., Kessler & Rutherford, 2010; Surtees et al., 2013). In such views, rather than gaze, the ability to mentally transform one's own position in space into that of another is integral for deriving how the world looks to this other person.

These findings align visual perspective taking with fundamental navigational abilities, particularly the ability to navigate an imagined environment, rather than with more sophisticated Theory of Mind and mentalizing abilities. Several authors have made such a case, reporting for example a correlational link between perspective taking and navigational skills, or that people are just as ready to judge items from the perspective of another person or from the perspective of a location they could, potentially, occupy (Gunalp et al., 2019; Hegarty & Waller, 2004; Kessler & Thompson, 2010; Kozhevnikov et al., 2006). Similarly, it is well-established that a key node of the theory of mind network, the Temporoparietal junction (TPJ) is also involved in visual perspective taking (e.g., Zacks, Rypma, Gabrieli, Tversky and Glover, 1999; for meta-analysis, see Schurz, Aichhorn, Martin & Perner, 2013) and self-other distinction (David et al., 2006; Lamm, Bukowski, & Silani, 2016) perhaps on the basis of the spatial difference vector between self and other (e.g., Jeannerod, 2007). However, the TPJ, and posterior parietal regions in general, are also implicated in imagined navigation (Committeri, Piccardi, Galati & Guariglia, 2015), specifically the translation of allocentric hippocampal place representations into representation of a location into egocentric space (Boccia, Sulpizio, Palermo, Piccardi, Guariglia, & Galati, 2017). The perhaps most dramatic example for this relationship is the link between the TPJ and out-of-body-experiences, in which patients subjectively "see" the world (and themselves) from outside their own body.

The TPJ is the one region typically impaired in patients that report these experiences and stimulation of the TPJ in healthy individuals, via transcranial magnetic stimulation (TMS), can induce them (Blanke et al., 2005). We therefore speculate that the ability to visually perspective-take might emerge from an evolutionary ancient mechanism that allows us to calculate what we can see – and what/how we can interact with – from different positions in space, therefore serving both epistemic and action functions. More sophisticated abilities for mentalizing and Theory of Mind can then capitalise on the emerging insights that other perspectives might provide one with very different knowledge of the same scenes (Frith & Frith, 2007; Kessler & Thompson, 2010; Mundy & Newell, 2007), giving rise to a common neuronal network for Theory of Mind, visual perspective taking, self-other distinction and (imagined) navigation.

A question for future research is how complete this simulation of different locations is. In our experiments, the other person's posture remained consistent across conditions, and their head was averted to the corner of the room, or outwards towards the participant. What, however, would happen if not only the actor's head, but their whole body was turned to face away from the table in our task? Such studies would allow one to disentangle if visual perspective taking is primarily based on locations in space, or whether it is embodied, in the sense that it is derived from a mental model of oneself in a target posture (Amorim, Isableu, & Jarraya, 2006). If so, this would predict that one's own ability to embody another's posture through our own behaviour might influence these effects. Aligning one's posture with that of another person has been shown to make judgements from their perspective easier (Kessler & Thompson, 2010; Kessler & Rutherford, 2010). In our own task, could participants' inadvertent imitation of their virtual partner's posture then explain why mental rotation can be performed with such ease from their perspective, even without any instruction to do so?

3.3.4 Conclusion

This study provides the first evidence that perceptual shifts into another person's perspective are unaffected by where this person is looking. It therefore confirms that people perceptually represent the content of another's perspective in a (quasi)perceptual manner, but that these perceptual simulations are based on their location in space, but do not take into account whether this person can actually 'see' target items. These data link perspective taking to more fundamental navigational skills and the ability what one could, in principle, see from another's vantage point, irrespective of what this person is currently looking at.

3.3.5 References

Allen, G. L., Kirasic, K. C., Dobson, S. H., Long, R. G., & Beck, S. (1996). Predicting environmental learning from spatial abilities: An indirect route. *Intelligence*, *22*(3), 327-355.

Ambrosini, E., Costantini, M., & Sinigaglia, C. (2011). Grasping with the eyes. *Journal of neurophysiology*, *106*(3), 1437-1442.

Amorim, M. A., Isableu, B., & Jarraya, M. (2006). Embodied spatial transformations:" Body analogy" for the mental rotation of objects. *Journal of Experimental Psychology: General*, *135*(3), 327.

Bach, P., Fenton-Adams, W., and Tipper, S.P. (2014). Can't touch this: the first-person perspective provides privileged access to predictions of sensory action outcomes. *J Exp Psychol Hum Percept Perform*. 40(2), 457.

Batson, C. D., Early, S., and Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imaging how you would feel. *Personality and Social Psychology Bulletin*. 23(7), 751-758.

Bayliss, A. P., Paul, M. A., Cannon, P. R., & Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic bulletin & review*, 13(6), 1061-1066.

Becchio, C., Bertone, C., & Castiello, U. (2008). How the gaze of others influences object processing. *Trends in Cognitive Sciences*, *12*(7), 254-258.

Blanke, O., Mohr, C., Michel, C. M., Pascual-Leone, A., Brugger, P., Seeck, M., ... & Thut, G. (2005). Linking out-of-body experience and self processing to mental own-body imagery at the temporoparietal junction. *Journal of Neuroscience*, *25*(3), 550-557.

Boccia, M., Sulpizio, V., Palermo, L., Piccardi, L., Guariglia, C., & Galati, G. (2017). I can see where you would be: patterns of fMRI activity reveal imagined landmarks. *Neuroimage*, *144*, 174-182.

Cole, G. G., Atkinson, M., Le, A. T., and Smith, D. T. (2016). Do humans spontaneously take the perspective of others?. *Acta Psychol* (Amst). 164, 165-168.

Committeri, G., Piccardi, L., Galati, G., & Guariglia, C. (2015). Where did you “left” Piazza del Popolo? At your “right” temporo-parietal junction. Cortex, 73, 106-111.

Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology*, *7*(1), 20-43.

Creem-Regehr, S. H., Gagnon, K. T., Geuss, M. N., and 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.

David, N., Bewernick, B. H., Cohen, M. X., Newen, A., Lux, S., Fink, G. R., ... & Vogeley, K. (2006). Neural representations of self versus other: visual-spatial perspective taking and agency in a virtual ball-tossing game. *Journal of Cognitive Neuroscience*, *18*(6), 898-910.

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.

Erle, T. M., & Topolinski, S. (2015). Spatial and empathic perspective-taking correlate on a dispositional level. *Social Cognition*. 33(3), 187-210.

Flavell, J. H., Everett, B. A., Croft, K., and Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Dev Psychol*. 17(1), 99.

Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2016). When do humans spontaneously adopt another's visuospatial perspective? *Journal of experimental psychology: human perception and performance*, *42*(3), 401.

Freundlieb, M., Sebanz, N., & Kovács, Á. M. (2017). Out of your sight, out of my mind: Knowledge about another person's visual access modulates spontaneous visuospatial perspective-taking. *Journal of experimental psychology: human perception and performance*, *43*(6), 1065.

Frischen, Alexandra, Andrew P. Bayliss, and Steven P. Tipper. (2007)"Gaze cueing of attention: visual attention, social cognition, and individual differences." *Psychological bulletin* 133.4 (2007): 694.

Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology*, *17*(16), R724-R732.

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*(2), 158.

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.

Gunalp, P., Moossaian, T., & Hegarty, M. (2019). Spatial perspective taking: Effects of social, directional, and interactive cues. *Memory & cognition*, 1-13.

Hamilton, A. F., Brindley, R., and Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition* 113, 37–44.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, *32*(2), 175-191.

JASP Team (2018). JASP (Version 0.10.0)[Computer software].

Jeannerod, M. (2007). Being oneself. Journal of Physiology-Paris, 101(4-6), 161-168.

Kampis, D., Parise, E., Csibra, G., and Kovács, Á. M. (2015). Neural signatures for sustaining object representations attributed to others in preverbal human infants. *Proc. R. Soc. B*. 282(1819), 20151683.

Kessler, K., and Rutherford, H. (2010). The two forms of visuo-spatial perspective-taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology,. Psychol. 1*, 1–12.

Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88.

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

Kozhevnikov, M., Motes, M. A., Rasch, B., & Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *20*(3), 397-417.

Lamm, C., Bukowski, H., & Silani, G. (2016). From shared to distinct self–other representations in empathy: evidence from neurotypical function and socio-cognitive disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150083.

MacFie, H. J., Bratchell, N., Greenhoff, K., & Vallis, L. V. (1989). Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of sensory studies*, *4*(2), 129-148.

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*(12), 2447-2460.

Michelon, P., and Zacks, J. M. (2006). Two kinds of visual perspective-taking. *Percept. Psychophys.* 68, 327–337.

Morgan, E. J., Freeth, M., & Smith, D. T. (2018). Mental State Attributions Mediate the Gaze Cueing Effect. Vision, 2(1), 11.

Mosconi, M. W., Mack, P. B., McCarthy, G., & Pelphrey, K. A. (2005). Taking an "intentional stance" on eye-gaze shifts: a functional neuroimaging study of social perception in children. *Neuroimage*, *27*(1), 247-252.

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current directions in psychological science*, *16*(5), 269-274.

Poulton, E. C. (1982). Influential companions: Effects of one strategy on another in the within-subjects designs of cognitive psychology. *Psychological Bulletin*, *91*(3), 673.

Roelfsema PR, and de Lange FP (2016). Early visual cortex as a multi-scale cognitive blackboard. *Annual Reviews in Vision Science*. 2,131-151.

Samson, D., Apperly, I.A., Braithwaite, J.J., Andrews, B.J., and Bodley Scott, S.E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *J Exp Psychol Hum Percept Perform*. 36(5), 1255.

Schurz, M., Aichhorn, M., Martin, A., & Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. *Frontiers in human neuroscience*, 7, 712.

Shepard, R. N., and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*. 171(3972), 701-703.

Surtees, A. D. R., Apperly, I. A., & Samson, D. (2013). The use of embodied self-rotation for visual and spatial perspective-taking*. Frontiers in Human Neuroscience*. 7, 698.

Surtees, A., Samson, D., & Apperly, I. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, *148*, 97-105.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav Brain Sci*. 28, 675–691.

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, *110*(1), 124-129.

Wang, Y., & Hamilton, A. F. D. C. (2012). Social top-down response modulation (STORM): a model of the control of mimicry in social interaction. *Frontiers in human neuroscience*, 6, 153.

Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. *Current Biology*, *29*(5), 874-880.

Williams, J. H., Waiter, G. D., Perra, O., Perrett, D. I., & Whiten, A. (2005). An fMRI study of joint attention experience. *Neuroimage*, *25*(1), 133-140.

Zacks, J., Rypma, B., Gabrieli, J. D. E., Tversky, B., & Glover, G. H. (1999). Imagined transformations of bodies: an fMRI investigation. *Neuropsychologia*, *37*(9), 1029-1040.

Zwickel, J. (2009). Agency attribution and visuospatial perspective taking. *Psychonomic Bulletin & Review*, *16*(6), 1089-1093.

Zwickel, J., and Müller, H.J. (2010). Observing fearful faces leads to visuo-spatial perspective taking. *Cognition*. 117(1), 101-105.

4. Chapter Four - Is Implicit VPT-2 embodied?

This final experimental chapter uses the same mental rotation paradigm to directly test whether participants' movement plays a causal role in their tendency to spontaneously shift into the perspective of the other person. Prior work (e.g. Kessler & Thompson, 2010) has conceptualised VPT as being supported by a motorically driven, mental body transformation into the position of another person. To test this directly, we simply restricted participants' movement during half of the experimental trials, and measured whether this affected participants' tendency to spontaneously derive how the object of interest appeared to the inserted persons in the scene.

4.1 Is Implicit Level-2 Visual perspective taking embodied? Spontaneous perceptual simulation of others' perspectives is not impaired by motor restriction.

Eleanor Ward*, Giorgio Ganis, Katrina L. McDonough, Patric Bach

School of Psychology, University of Plymouth, Drake Circus, Devon, UK, PL4 8AA

*Corresponding author/lead contact: Eleanor.ward@plymouth.ac.uk

Abstract

Embodied accounts of VPT suggest that making judgements from another's perspective becomes less effortful if one's own body position aligns with that of the other person (e.g. Kessler & Thompson, 2010) indicating a causal role of body posture in visual perspective taking. Here we test whether movement has a causal role in perspective taking. We used our previously established task that shows that other peoples' visual perspectives are represented perceptually in a similar way to one's own and can drive bottom-up processes in the same way as own perceptual input (Ward, Ganis & Bach, 2019; Ward, Ganis, McDonough & Bach, under review). We asked participants to identify whether alphanumeric characters, presented at one of eight possible orientations away from upright, were presented normally, or in their mirror-inverted form (e.g. "R" vs. "Я"). In some scenes, a person would appear sitting to the left or the right of the participant. In half of the trials, participants' movement was restricted with a standard chin rest, and in the remaining half, participants could move freely when making their judgements. In the absence of another person, we replicated the well-established mental rotation effect, where recognition of items becomes slower the more items are oriented away from upright (e.g. Shepard and Meltzer, 1971). Crucially, in all conditions, this response pattern changed when another person was inserted into the scene. People spontaneously took the perspective of the other person and made faster judgements about the presented items in their presence if the characters were oriented towards upright to them. Participants' ability to move did not influence these effects. The results therefore rule out active physical movement as a causal explanation of visual perspective taking, and instead argue that subtle postural readjustments when making judgements from an other's perspective are a bodily consequence of the mental transformations of a person's *actual* to *imagined* position in space.

Keywords: *visual perspective taking; perceptual simulation; navigation; mental rotation; mental imagery; active inference*

Introduction

Humans effortlessly take others' perspectives and derive what they can or cannot see, or how a scene looks to them (Flavell, Everett, Croft & Flavell, 1981). This everyday skill allows people to give a passer-by directions so they can plan a route from their own perspective, or work out whether an oncoming driver has noticed them before safely crossing a road, for example. These abilities have been argued to underlie the ability to coordinate actions with others (Freundlieb, Kovács, & Sebanz, 2016), and may form the basis of more sophisticated social abilities such as reasoning about others' beliefs, desires, and goals (Batson, Early, & Salvarani, 1997; Erle & Topolinski, 2015; Tomasello, Carpenter, Call, Behne, & Moll, 2005; Mattan, Rotshtein & Quinn, 2016).

Recent work has conceptualised perspective taking as a form of perceptual simulation, which inserts the content of another's perspective onto one's own perceptual processes, as if it were one's own perceptual input (Kampis, Parise, Csibra, & Kovács, 2015; Surtees, Apperly, & Samson, 2016; Ward, Ganis, & Bach, 2019). Such a (quasi-)perceptual representation could then drive one's own action and decision making processes just like own input, explaining the developmental link between visual perspective taking and higher level mentalizing (Batson et al., 1997; Erle & Topolinski, 2015; Hamilton, Brindley & Frith, 2009; Tomasello et al., 2005; Mattan, et al., 2016) and its link to joint action (Freundlieb et al., 2016).

A recent series of studies from our lab provided the first direct evidence that people represent others' perspectives in a similar way to their own (Ward et al., 2019; Ward, Ganis, McDonough & Bach., 2020), and that these (imagined) other-perspectives can drive perceptual decision making processes in the same way as their own perceptual input, similar to other perceptual simulation processes (e.g. see

Roelfsema & de Lange, 2016, for a review). Prior studies had already provided evidence for an overlap between one's own and others representations of the world, so that stimulus judgments become harder if another person around us would make the same judgements differently from their perspective (e.g., Sampson, Apperly, Braithwaite, Andrews, & Bodely Scott, 2010; Surtees, Samson, & Apperly, 2016; Tversky and Hard, 2009; Zwickel and Müller, 2010; Zwickel, White, Constantin, Senju & Frith, 2010). Yet, these studies left open whether this interference happens on a perceptual level or a conceptual/response level, or whether it simply indexes the uncertainty when a person becomes aware that others would judge the same stimulus differently from their perspective.

To reveal whether people have direct perceptual access to the content of another's perspective, we tested whether another's viewpoint allows people to make perceptual judgments that would be difficult from their own perspective. We adapted the established mental rotation task, in which participants simply report, as quickly as possible, whether alphanumeric characters at various orientations are presented in their canonical or mirror-inverted form (e.g. "R" vs. "Я"). The well-known finding is that these judgements increase linearly the more the characters are rotated away from upright (Shepard & Melzer, 1971), because people first must mentally rotate them back into their canonical orientation before being able to judge them. Here, we used this task to test whether people would spontaneously make these judgments *from the perspective of the other person*, so that they can rapidly judge items that are oriented away from themselves, if they appear upright to the other person. Indeed, participants recognized the items more quickly when an incidentally inserted other person would have a more upright view of the to-be-judged character than them, whilst judgements that would be more difficult from this other perspective became slower. Moreover, regression analyses showed that recognition times across letter orientations increased linearly with the angular disparity of the item not only to the participant's viewpoint, but also to the other person's viewpoint, suggesting that participants could mentally rotate the items from *their own or the other's* perspective.

These data provided direct evidence that people can mentally represent the content of another's viewing perspective in a form that can "stand in" for own visual input and drive subsequent perceptual judgements and mental rotation processes. Importantly, these shifts to the others' perspective occurred spontaneously, even when the persons in the scene were completely task-irrelevant. Further studies showed that the same effects were not present when the person was substituted for an inanimate object (i.e. a lamp, Ward et al., 2019), but increased substantially when participants were asked explicitly to take the other person's perspective. More recent work (Ward et al., *under review*) shows that these shifts into the other's perspective are not sensitive to where this person looks but reflect their location in space and which perspectives this vantage point would, in principle, afford onto the to-be-judged item (irrespective of where the person actually looks).

An interesting anecdotal observation was that, within these tasks, participants would sometimes inadvertently shift their *actual* position towards the other person's, angling their head slightly leftwards if another person appeared to the left of the items on the screen, and rightwards if the person appeared to the right. This observation fits with the view that perspective taking is an 'embodied' process (e.g. Kessler & Rutherford, 2010; Kessler & Thompson, 2010; Kessler & Wang, 2012), in which people mentally rotate themselves into the position of the other person. Studies have shown for example that explicit perspective taking (i.e. consciously judging how a scene would appear to another person with a different view) takes longer the more another person is rotated from one's own perspective (Kozhevnikov , Motes, Rasch & Blajenkova, 2006; Surtees et al., 2013; Kessler & Rutherford, 2010). Similarly, when people physically align their posture with that of another person, judgements from this *other*-perspective become easier, whilst adopting a misaligned posture makes it harder to take this other person's perspective.

Whilst there have been a number of distinct claims made about the processes involved in embodied cognition – and what embodied cognition can mean – (e.g. Wilson, 2002), the assumption tested in this work is that of the interaction of cognition and action. Most relevant to this work is the notion that 'off-line' cognition is grounded in the body, and that, given the evolutionary emphasis on human survival, high-level cognitive functions are likely based generally on more primitive functions such as motor processing (Muto, Matsushita & Morikawa, 2018). This view holds that even when we are not actively performing an action, activity in the mind remains grounded in mechanisms that evolved to afford interaction with the environment (Wilson, 2002). It is possible in our task, therefore, that when we (mentally) move into another person's perspective, we draw on mechanisms that evolved to facilitate actual action, such as mechanisms that would actively help us navigate – and move – through space. Participants' overt movements in earlier versions of this task might therefore reflect an epiphenomenal 'leakage' from the mental transformation of people's actual to the imagined other-position, similar to other bodily consequences of motor imagery (Colton, Bach, Whalley & Mitchel, 2018; Bach, Allami, Tucker, & Ellis, 2014; Jacobson, 1930; Vargas, Olivier, Craighero, Fagida, Duhamel & Sirigu, 2004).

Here we ask whether these bodily movements are not be simply bodily signs of a mental perspective transformation, but whether they play a *causal* role in driving the shift to the other person's perspective. There are two ways in which overt body movements could support judgments from the other person's perspective. First, several recent proposals from the field of embodied cognition argue that people actively use their own body and the environment to support cognitive judgments (e.g., Glenberg, 2010; Proffitt, 2006; for perspective taking, see Tversky & Hard, 2009). In our case people could have used the bodily movement to trigger "embodied" processes that allow them to picture the world from another's perspective. When people grow up, they develop highly automatic processes that allow them to predict the perceptual consequences of their actions (i.e. "forward models", Blakemore, Frith & Wolpert, 1999; Miall & Wolpert, 1996), such that they can predict, before the action is completed, which visual (e.g., Hughes & Waszak, 2011), auditory (Kunde, Koch, & Hoffmann, 2004), or tactile sensation it will produce (e.g., Morrison, Tipper, Fenton-Adams & Bach, 2013; Bach, Fenton-Adams & Tipper, 2014). In mental rotation tasks, it has been shown for example that manual rotations consistent with the speed and direction of mental rotations facilitate faster judgements. These movements appear to directly support the mental movement, as restricting these movements or asking participants to make different movements interferes with imagery of finger movements (e.g., Vargaset al., 2004) or mental rotation processes (Wohlschläger & Wohlschläger, 1998). Similar links have been observed for emotion judgments and restrictions of one's own facial musculature, and abstract mathematical relationships and hand gestures (Cook, Yip & Goldin-Meadow, 2010; Neal & Chartrand, 2011; Parsons, 1994). In our task, therefore, people could make overt movements towards the other person's location for the same purpose: to trigger processes that predict the perceptual consequences of how the world would look if these movement had been completed.

A second possibility is that the body movements reflect *active* attempts to effectively sample the scenes from the other person's perspective. Recent proposals from the domain of predictive processing argue that perception is not a passive process, but a process of "active inference" in which people constantly move their bodies (Friston, Daunizeau & Kiebel, 2009) and their eyes (e.g., Parr & Friston, 2017) to most effectively sample the information that they require for the task, or to fulfil their prior expectations and avoid 'surprising' states (Friston, 2010). In our task, the presence of a person on the left or the right, might have triggered body movements so that people's own perspective – and the perceptual input they receive – aligned more closely with that of the other person. For our task, this raises the possibility, therefore, that the measured shift into the other's perspective does not reflect changes to participants' *mental* representation of perceptual input, but an actual change in the perceptual input they receive, so that they can actually see the item better in orientations that aligns with the other person's location.

One effective way to test whether the subtly body movements of participants play a causal role in perspective taking is by comparing performance in conditions in which these movements are possible and conditions in which they restricted. As noted above, restricted movement tasks have been used to test embodied emotional processes (Neal & Chartrand, 2011), mental rotation processes (Moreau, 2013), and even mathematical reasoning processes (Cook et al., 2012). We gave participants the same mental rotation task as in our previous studies (Ward et al., 2019; *2020*) and asked them to report whether alphanumeric characters appearing on a table in front of them in different orientations were presented normally or were mirror-inverted (e.g. “R” vs. “Я”). In some of the trials, a person appeared in the scenes and looked at items from either the left or the right of the table. This allows us to measure how much faster items are identified when they face the other person, compared to facing away from them. The crucial manipulation was that in half of the trials, participants’ movement was restricted using a chinrest; they could therefore not adjust their own body movement to either actively sample the scenes from the others’ perspective or to trigger “embodied” perspective taking processes. If movements are causal in creating the shifts to the others’ perspective, then restricting participants’ movement should disrupt perspective taking, and the response time benefits for items easy to recognize for the other person would be reduced or eliminated. If, however, the movements are simply epiphenomenal ‘leakage’ of mental rotations into the other person’s body (e.g. Colton et al., 2018), then preventing these movements should have no effect.

A secondary goal of the current study was to explore which individual differences determine the tendency to spontaneously take another’s visual perspective. Prior work (e.g. Langdon, Coltheart, Ward & Catts, 2001) indicates that individuals with schizophrenia are impaired in visual perspective taking. We therefore tested whether participants’ individual ability to judge from the other’s perspective (as measured in our task) is related to their tendency to experience psychosis-like states, using the Schizotopy Questionnaire (STQ; Claridge & Brocks, 1984). Similarly, we tested whether prior findings that link perspective taking to more sophisticated mentalizing abilities and social interaction (e.g. Batson et al., 1997; Erle & Topolinski, 2015; Hamilton et al., 2009; Tomasello et al., 2005; Mattan, et al., 2016) can be demonstrated in our task. We therefore also gave the Interpersonal Reactivity Index (IRI; Davies, 1983) and the Autism Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin & Clubley, 2001) to all participants, to ascertain to what extent the presence of autistic-like traits and social interaction abilities predict spontaneous perspective-taking.

4.2 Experiment 5

4.2.1 Method

Participants

Sixty-nine (59 females, 1 non-binary gender) naive participants were initially recruited via the University of Plymouth student participation pool. All participants were adults (age range 18-35) and gave written informed consent according to the declaration of Helsinki. Approval was obtained from the University of Plymouth Ethics Committee. Participants received course credit as compensation. After exclusion (error >20%), the remaining sixty-one participants (51 females, 1 non-binary gender; mean age: 20.5 years, range: 18-30) provide 80% power to detect effects in the range of d =.32. Prior work on this paradigm (Ward et al., 2019) has revealed that effect sizes are substantially larger (.747 < d < 1.08 for the main perspective taking effect). For correlations with measures of individual differences, the sixty-one participants also provide 80% power to detect correlation coefficients in the range of r = .25 (two-tailed) or r = .23 (one-tailed).

Apparatus, stimuli and procedure

All experiments were conducted in behavioural testing lab space of the University of Plymouth. The experiments were administered using Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Stimuli were presented on a 19” LED computer monitor (Resolution: 1900x1200; Refresh rate: 60Hz). Responses were made on a standard computer keyboard with UP, DOWN, and SPACE keys as active response keys. Red and green stickers were positioned on the DOWN and UP keys, respectively. A standard chinrest was provided for participants, fixed with a screw clamp central to the computer monitor at a distance of 60cm, and a height of 30 cm from the desk surface.

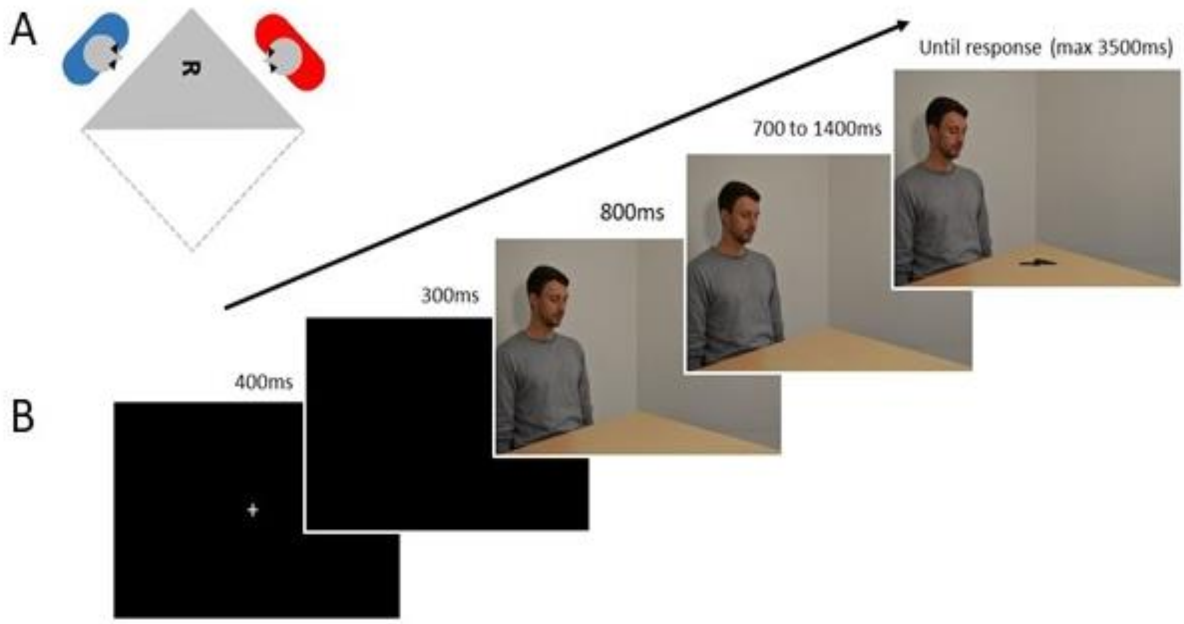

***Figure 3.1*. (A) Scene set up and, (B) schematic of the trial sequence.** Panel A shows the position of the inserted persons relative to the character on the table, producing a viewing angle of approx. 90 degrees relative to the participant. Panel B shows the timing of the trial sequence. First participants viewed a fixation cross and 300ms blank screen. The next scene showed a male (pictured above) or female actor positioned either to the left (as shown above) or the right of the table. After a random period of 1500ms to 2200ms, an alphanumeric character appeared on the table either in its canonical or mirror-inverted form. Participants responded with a button press to indicate whether they though the letter was normal or mirrored. In half of all trials, participants’ movement was restricted using a standard chin rest.

Participants sat upright facing the screen at a distance of approximately 60cm and were given written and verbal instructions. They were given examples of the rotated items that would appear on the screen and completed eight training trials that were identical to the main experiment (Figure 3.1). Each trial (total trials = 572) started with a fixation cross displayed for 400ms, followed by 300ms blank screen. The subsequent stimulus sequence included two frames, measuring 33.4 by 23.5 degrees of visual angle, presented without inter-stimulus interval. The first frame was presented for 1500 to 2200 ms. In one third of the trials, it showed a view onto a corner of a square table in a grey room. The remaining trials showed a person sitting behind the same square table, gazing at the centre of the table. The person could either be male or female and sat either on the left or right side of the table on an equal number of trials.

The second frame in the sequence was identical to the second frame, but now one of 48 possible items appeared on the table, at the location on the table the on-screen person was gazing at. This item was one of three alphanumeric characters (4, P, or R), presented either in the canonical version or mirror-inverted about their vertical axis, in one of eight orientations (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°, with 0° denoting the upright canonical orientation and angles increasing in a counter-clockwise fashion) relative to the participant. The characters always appeared in the same position on the table, half-way between the outward corner of the table and its centre, such that the persons to the left and right would gaze at the table from roughly 90° and 270°, respectively (perpendicular to the viewpoint of the participant), as at these angles the character's angular disparities from the participant and the other person were statistically independent across conditions. Character rotation occurred around the character's centre point.

The third frame remained on the screen until a response was made to a maximum duration of 3500ms. Participants were asked to judge whether each character was presented in its canonical or mirror-inverted form. Participants responded using their right hand by pressing the green key to indicate a canonical item and the red key to indicate a mirrored item. Response times were measured relative to item onset.

The trials were divided into four blocks of 144 trials each. Half were completed using a chin rest (height 30 cm from desk, 60 cm from screen) in order to restrict motion, and the remaining half of trials were completed without a chin rest, in an ABAB order, counterbalanced across participants. The presented stimuli were pseudorandomised across blocks, such that all possible combinations of actor-location/item/presentation/orientation were shown in both the headrest and no-headrest condition throughout the experiment. In both conditions the viewing distance from head to screen was approximately 60cm, as in all previous experiments.

Quantification and Statistical analysis

Data (pre-)processing and analysis was identical to Ward et al., (2019) and conducted in Microsoft Excel (2010) and JASP (2018). Violin plots were created using Raincloud Plots (Version 1; Allen, Podiaggi, Witaker et al., 2019). Power analyses were conducted in G*Power (Version 3.1; Erdfelder, Faul, & Buchner, 1996).

Dependent measures were the recognition times (measured from item onset) for each character orientation (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315), depending on person location (No-person, Person-left, Person-right) and movement condition

(Free-movement, No-Movement). Analogous analyses of error rates were also conducted to rule out speed/accuracy trade-offs. In both conditions, error rates numerically followed the pattern of the main recognition times but did not show statistically reliable differences (Table 3.1).

To quantify changes in recognition times when the characters either faced the participant (i.e. was seen in its canonical orientation from the perspective of the participant) or the other person in the scenes, we derived two analogous and statistically independent summary measures, as in our previous work (Ward et al., 2019). The first summary measure *Toward/Away-bias* indexes to what extent characters were recognized faster the more they faced towards the participant (0°) rather than away from them (180°), separately for each participant and each condition (No-person Free-movement, Person-left Free-movement, Person-left No-movement, Person-right Free-movement, Person-right No-movement and No-person No-movement). This measure therefore quantifies the mental rotation effect (Shepard & Melzer, 1981). The second summary measure (Left/Right-bias) indexes how much faster characters were recognized when oriented towards the left (270°) rather than right (90°), or vice versa.

The contribution of each character orientation to the two summary measures was derived by treating each participant's recognition time for this character orientation as a vector in a coordinate system, with the recognition time providing the distance from the origin and the rotation angle the polar angle. A character orientation's contribution to the Toward/Away-bias was then derived simply from the recognition times multiplied with the negative of the cosine of the orientation angle. As a result, characters contribute negatively the more they face the participant (315°, 0°, 45°) and positively they more they are oriented away from them (225°, 180°, 135°).

Similarly, the contribution of a character's orientation to the Left/Right-bias was calculated as the recognition time multiplied with the sine of the orientation angle. Character orientations contribute positively the more they face to the left (45°, 90°, 135°) and negatively the more they face to the right (225°, 270°, 315°). This procedure effectively maps the changes evident in the radar plots for each angle onto two orthogonal and statistically independent summary measures, so that they can be compared across conditions without accruing alpha inflation due to multiple testing, which would result if each of the eight angles were compared separately.

By averaging these values, separately for each summary measure, participant and condition (No-person, Person-left, Person-right), we are able to calculate, first, whether characters were recognized faster the more they appear in the canonical orientation to the participant (negative values on the Toward/Away-bias) compared to when they are oriented away (positive values), reflecting the expected mental rotation effect. Similarly, they allowed us to calculate to what extent characters were recognized faster the more they were oriented leftwards and would appear in their canonical orientation to a person sitting to the left (positive values on the Left/Right-bias) rather than rightwards, appearing in their canonical orientation to a person sitting on the right (negative values). We were then able to determine if this left/right bias changed depending on whether another person was presented in the scenes and on whether the person was on the left or on the right.

The crucial comparison is the difference between the Left/Right-biases in the Person-left and Person-right conditions, which describes how much faster letters are recognized when rotated left than right, depending on whether the other person is sitting to the left or right. Note that the direct comparison of the Person-left and Person-right conditions is statistically identical to the comparison of how much person presence shifts mental rotation performance in the Person-left and Person-right conditions relative to the No-Person baseline (i.e. how much person presence shifts recognition times away from 0˚ towards either 90˚ or 270˚), as this would involve subtracting the same baseline value from each of the two conditions for each participant, and would therefore not affect the absolute difference between them.

Across-participant regression analyses.

In prior work, the mental rotation effect is sometimes characterised in terms of separate linear regressions of an items' recognition time to its angular disparity relative to the participant, for each participant separately (Shepard & Melzer, 1971). The results reveal linear increases with increasing angular disparity for the large majority of participants. Here, we used this analysis model to test whether an item's recognition times can be described, on a single participant basis, as a linear increase of the character's angular disparity *both* to the participant and to the other person. To this end, we entered each participant's item mean recognition times for each character orientation in the Person-left and Person-right condition as dependent variable in a multiple regression, with the item's angular disparity to the participant and to the other person as two statistically independent predictors. This analysis provides regression coefficients for both predictors – angular disparity to participant and other person – for each participant separately. We report mean across-participant regression coefficients for each of these two predictors and compare them with t-tests against zero.

Individual differences measures

All participants were given three paper questionnaires after the computer task. First, the *Autism Quotient* (AQ; Baron-Cohen, Wheelwright, Skinner, Martin & Clubley, 2001) consists of 50 questions assessing social skills (e.g., I enjoy social occasions), attention to detail (e.g., I often notice small sounds when others do not), attention switching (e.g., I prefer to do things the same way over and over again), communication (e.g., I enjoy social chit-chat), and imagination (e.g., I find making up stories easy). The overall score gives a measure of autistic traits, where numerically high scores indicate high levels, and scores at the lower end indicate lower levels of autistic traits. Responses are recorded using a four-point Likert scale, with the options Definitely Agree, Slightly Agree, Slightly Disagree, and Definitely Disagree. Prior validation studies (e.g., Baron-Cohen et al., 2001b and Wakabayashi et al., 2006) show moderate to high internal consistency (Cronbach's alpha = .63–.77) and high test–retest reliability (r = .70), in both autistic and neurotypical samples.

Second, the *Interpersonal Reactivity Index* (IRI; Davis, 1983) is a 28 item questionnaire measuring empathy, comprised of the four subscales measuring perspective taking (e.g. I sometimes try to understand my friends better by imagining how things look from their perspective), empathic concern (e.g. I am often quite touched by things that I see happen.), fantasy (e.g. I daydream and fantasize, with some regularity, about things that might happen to me) and personal distress (e.g. When I see someone who badly needs help in an emergency, I go to pieces). Responses are made on a 5-point Likert scale ranging from "does not describe me well" to "describes me very well". Numerically high scores indicate high levels of empathy, whilst lower scores indicate low levels of empathy. Prior validation studies show moderate to high internal consistency (Cronbach's alpha = .73–.83; DeCorte, Buysse, Verhofstadt, Roevers, Ponnet & Davis, 2007) and high test–retest stability

(Intraclass correlation coefficients =.71-.86; Gilet, Mella, Studer, Grühn, Labouvie-Vief, 2013).

Finally, the *Schizotypy Questionnaire* (STQ; Claridge & Brocks, 1984) is a short measure of schizotypal personality traits, and consists of two scales, corresponding to the distinction made in the Diagnostic and Statistical Manual of Mental Disorders, Third Edition (DSM-III; American Psychiatric Association, 1980) between schizotypal personality disorder (STA scale) and borderline personality disorder (STB scale). Simple 'yes/no' responses are made to questions targeting schizophrenic-like features (e.g. Do you ever suddenly feel distracted by distant sounds that you are not normally aware of?), and borderline-personality traits (e.g. Do you at times have an urge to do something harmful or shocking?), scoring 1 for 'yes' responses, and 0 for 'no' responses. Numerically high scores indicate higher levels of Schizotypy, whilst lower scores indicate lower levels. Here we are interested specifically in schizotypal traits, therefore only responses for questions 1-37 in the STA part of the STQ are collected and reported in this study.

We were interested in whether either of these individual difference measure predicts people's spontaneous tendency to take the other person's perspective. This tendency is indexed by the difference between left/right-biases when a person is sitting on the left compared to when they are sitting on the right, reflecting how much faster/slower item recognition the more items are oriented towards/away from the other person. We therefore calculated this difference for participant separately by subtracting the mean left/right-bias value for a person sitting on the right from the value for a person sitting on the left. Scores for the AQ (Baron-Cohen et al., 2001), the IRI (Davis, 1983), and the STQ (Claridge & Brocks, 1984) were then entered as predictor variables, and perspective taking scores as the dependent variable, into a multiple linear regression model.

4.2.2 Results

As in our prior work (Ward et al, 2019; Ward et al., under review), erroneous responses (8% on average) were excluded from the analysis of recognition times (RTs), as well as trials with RTs longer than 2000ms, or shorter than 150ms.

Mental rotation

We first confirmed that our data replicate the known mental rotation effect (Shepard & Meltzer, 1971), where RTs increase linearly with the item's angular disparity to the participant. We first derived the overall (across conditions) towards/away bias, indexing in milliseconds, how much more slowly items are identified the more they are rotated away from the participant compared to towards them, and compared it with a simple t-test against zero. This towards/away bias was positive in all conditions, M=54.14; SD=22.09, $t(60)$=19.14, p<.001, d=2.45, BF_{10}=2.124e+24, showing, unsurprisingly, that items are identified more quickly the more they are oriented towards the participants. We further confirmed this mental rotation effect by regressing each item's recognition time to the expected linear increase with angular disparity, as in prior research (Shepard & Melzer, 1971; Ward et al., 2019), revealing positive slopes in all bar one participant, mean β =1.5; $t(60)$=23.34, p<.001, d=2.99, BF_{10}=3.124e+28.

We then verified that this overall mental rotation effect was not affected by person presence and the chinrest manipulation. As the actors would sit at 90 and 270 degree angle to the participant, and their location was therefore orthogonal to the towards/away axis, we did not expect that person presence/location would affect the overall mental rotation effect. Indeed, a 2x3 ANOVA on the towards/away-biases across conditions with the factors Movement (No-movement, Free-movement) and Location (person left, person right, no person) did not reveal any significant main effects or interactions, $F<1$, $BF_{10}<.155$ for all. When conditions were analysed separately, decisive evidence of slower recognition of turned away items was present in all conditions, $t(60) > 12.99$, $p <.001$, $d >1.7$, $BF_{10}>1.188e+16$, for all.

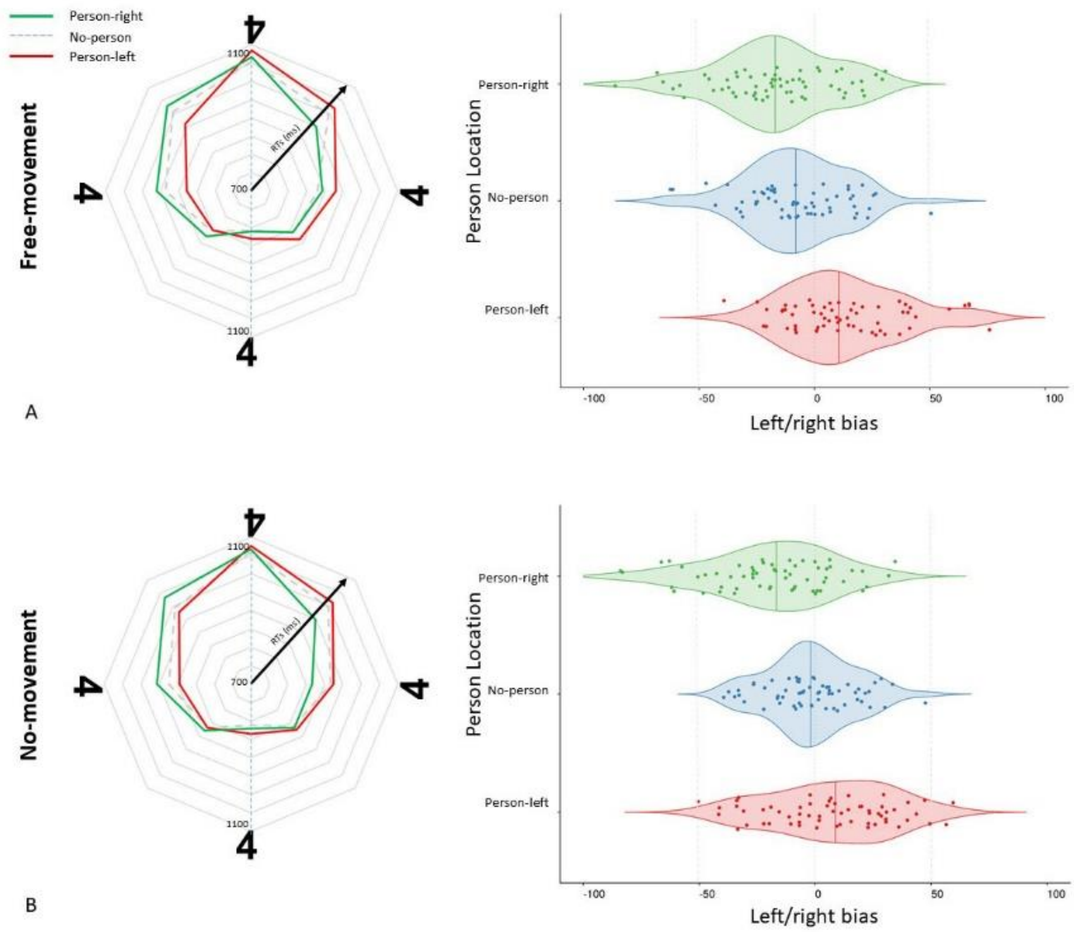

***Figure 3.2.* Results for both free-movement and no-movement conditions: Movement restriction does not impede visual perspective taking.** (A) *Free-movement condition*. Left panel: Mean recognition times (ms) to correctly classify items as canonical or mirror-inverted in each of the eight orientations depending on whether the person was absent, sitting on the left, or sitting on the right. Right**:** Violin charts showing the Left/right bias when a person was

sitting on the right (top), the person was absent (middle), and the person was sitting on the left (bottom). (B) *No-movement condition*. Left: Mean recognition times (ms) for mirror-inverted/canonical judgements as described for (A). Right: Left/right bias when participants' movement is restricted using a chin rest, as described for (A).

Perspective taking

The main question was whether people would spontaneously take the perspective of the other person in the scenes, such that items were recognized faster when oriented towards compared to away from them, and whether this effect, in turn, was determined by whether participants were able to physically align their posture with the actors in the scenes. We therefore derived, for each participant and condition separately, the Left/Right-bias, which indexes how much faster left-oriented items are identified compared to right-oriented items. Here, positive values indicate faster recognition times for left-oriented items (upright to person sitting on the left) and negative values indicate faster recognition of items oriented to the right (upright to a person seated on the right).

These left/right biases were entered into a 2 x 2 ANOVA, with Location (person-left, person-right) and Movement (headrest, no-headrest) as within subject factors. Replicating our prior work, this analysis revealed decisive evidence of a main effect of Location, $F(1,60)=50.556$, $p<.001$, $\eta p^2=.457$, $BF_{10}=2.240e+14$. As in our prior work (Ward et al., 2019), left/right-biases were more negative (indexing faster recognition of rightwards- than leftwards oriented letters) when someone was sitting on the right, and more positive (indexing faster recognition of leftwards- than rightwards oriented letters) when someone was sitting on the left confirming that the presence of another person facilitates faster judgements when items are seen as upright from the position of this other person.

The predicted interaction of Location and Movement, $F(1,60)=.689$, $p=.41$, $\eta p^2=.011$, $BF_{10}=1.196$, was not significant, indicating that people spontaneously simulate the visual perspectives of the inserted persons, even when they are unable to physically align themselves with their position in space. Direct comparisons for left/right bias revealed reliable differences for person-left and person-right locations in both the free-movement condition, $t(60)=6.81$, $p<.001$, $d=.87$, $BF_{10}=2.367e+6$, and in the no-movement condition, $t(60)=5.07$, $p<.001$, $d=.65$, $BF_{10}=3916.17$. Next to this, the analysis only revealed an unpredicted and theoretically irrelevant main effect of Movement, so that recognition times were generally faster in the free movement condition, but Bayesian analyses revealed this effect to be negligible, $F(1,60)=4.82$, $p=.031$, $\eta p^2=.074$, $BF_{10}=.381$.

Table 3.1. Means (M) and Standard Deviations (SD) for the Left/Right- and Towards/Away-biases in Error rates in all conditions. Relates to Figure 3.2. Forward/Away and Left/Right-biases were calculated analogously as for the recognition times. *$p<.05$. **p<001.

	Toward/away bias			Left/right bias		
Condition	Person-left *M (SD)*	Person-right *M (SD)*	No-person *M (SD)*	Person-left *M (SD)*	Person-right *M (SD)*	No-person *M (SD)*
All	-.023 (.02)**	-.02 (.02)**	-.021 (.02)**	-.005 (.015)*	.002 (.016)	-.001 (.014)
Free-movement	-.022 (.02)**	-.017 (.02)**	-.019 (.03)**	-.006 (.019)*	.000 (.022)	.002 (.021)
No-movement	-.024 (.02)**	-.022 (.02)**	-.024 (.02)**	-.005 (.021)*	.003 (.021)	-.004 (.017)

Regression analysis

As in our previous studies, we tested whether recognition times could be described as independent linear increases depending on an items angular disparity to the participant as well as the other person, by using both disparities as orthogonal predictors in a simple regression model, for each participant and condition separately (and then comparing them against zero). Overall, these revealed very strong evidence for independent contributions of both the angular disparity to the participant, mean β =1.39, $t(60)=19.181$, $p<.001$, $d=2.46$, $BF_{10}=1.189e+24$, and to the other person, mean β =.38, $t(60)=7.209$, $p<.001$, $d=.92$, $BF_{10}=1.000e+07$, showing that recognition times can be described by independent mental rotation functions from one's own and the other person's perspective.

To test how these linear relationships were affected by participants' ability to move freely, beta estimates were entered into a 2 x 2 repeated measures ANOVA with Movement (Free-movement, No-movement) and Viewpoint (Self, Other) as within-subject factors. As expected, this analysis provided decisive evidence for a main effect of Viewpoint, $F(1,60)=129.57$, $p<.001$, $\eta p^2=.68$, $BF_{10}= 2.117e+32$, showing that the angular disparity towards the participants determined recognition times to a stronger extent that angular disparity to the other person. As the main analysis, it provided considerable evidence against any influence of the ability to move freely on the linear relationships. There was neither a main effect of Movement, $F(1,60)=.615$, $p=.436$, $\eta p^2=.010$, $BF_{10}=.154$, nor an interaction of Movement and Perspective, $F(1,60)=.016$, $p=.9$, $\eta p^2=.00$. $BF_{10}=.141$. Thus, there was neither an overall change in how strongly angular disparities to the participant and the other person determined recognition times, nor a specific change in the contribution of either the angular disparity to the participant and the other person. Indeed, in the no-movement trials, both the angular disparity away from upright to the participant, mean β=1.37, $t(60)$=16.01, p<.001, d=2.05, BF_{10}=1.632e+20, and to the actor, mean β=.35, $t(60)$=5.02, p<.001, d=.64, BF_{10}=3507.56, determined recognition times. The same was also true for free-movement trials, where angular disparity to the participant, mean β=1.41, $t(60)$=17.426, p<.001, d=2.23, BF_{10}=9.920e+21, and to that of the other person, mean β=.402, $t(60)$=6.87, p<.001, d=.88, BF_{10}=2.818e+06 provided reliable contributions to RTs.

Relationships to individual differences

A second goal of the study was to test how individual differences in the tendency to judge the items from other's perspective is related to individual differences in Schizotypy, Autistic Traits and Reactivity in social interactions. We therefore correlated each participant's perspective taking score (the difference between perceptual biases when another person was sitting on the left or the right) separately with each of their three questionnaire scores and their age, each time correlating separately against perspective taking scores for all conditions. Other correlations of interest are reported in table S3.1.

We first correlated participants' age against perspective taking scores, replicating our prior finding (Ward et al., 2019) that people take another's perspective more as they increase in age, r=.38, p=.003, BF_{10}=13.52. We then tested whether perspective taking was negatively correlated with participants' self-reported measures of schizotopy as seen in prior work (Langdon et al., 2001; Langdon & Coltheart, 2001). Participants' STQ scores were correlated against perspective taking scores, replicating the negative relationship between schizotopy and visual perspective taking, r=-.26, p=.044, BF_{10}=1.166. Note that here *BF* is under 3, indicating the possibility that this is a spurious effect. We then correlated perspective taking scores against IRI scores (M=68.45; SD=11.34) and AQ scores (M=16.64; SD=6.29) giving a measure of the relationship between perspective taking and social ability. Neither revealed a reliable relationship, r <.08, p >.543, BF_{10}<.191, for all.

Multiple linear regressions analyses were conducted to test whether these questionnaire scores and participants' age were reliable predictors of perspective taking. Using the enter method, all variables were hierarchically entered into the model individually, revealing that when all variables were included, the model reliably predicted perspective taking score, R^2=.24, F(4,53)=4.059, p=.006. With all variables included, beta coefficients confirmed that both age, β=.376, t=3.013, p=.004, and STQ score, β=-.268, t=-2.154, p=.036, provided reliable contributions to the model, whilst AQ score, β=.093, t=.679, p=.5, and IRI score, β=.094, t=.716, p=.477, did not.

The addition of STQ scores increased the predictive power of a model containing only age as a predictor by 6%, F(1,55)=4.210, p=.045, but the individual addition of AQ and IRI scores as predictors did not improve the model, R^2change <.07%, F(1,54/53)<.513, p>.477 for all, further confirming that AQ and IRI scores are unrelated to visual perspective taking.

4.2.3 Discussion

We tested whether people's spontaneous tendency to take another's visual perspective depends on the being able to make active head movements to physically align one's own perspective with that of another. In a version of our recent task (Ward et al., 2019), we asked participants to judge the presentation (mirror-inverted or canonical) of alphanumeric characters on a table, shown at varying orientations. Between trials, an incidentally presented person appeared to either the left or the right of the item. The results replicated, first, the well-established mental rotation effect (e.g. Shepard & Melzer, 1971), with recognition times increasing linearly the more items were rotated away from the participant's own viewing perspective, in line with the idea that the items first have to be mentally rotated back into their canonical (upright) orientation before they can be judged. The results also replicate our finding (Ward et al., 2019; Ward et al., 2020) that participants can rely on the other person's perspective to make these judgments. Participants recognized items oriented away from themselves more quickly when the items would appear upright to the other persons in the scenes. Thus, leftward-oriented items were recognized more quickly when another person was present who saw the letter from the left, and rightward-oriented items were recognized more quickly when another person was viewing them from the right. Moreover, regression analyses showed that recognition times increased linearly not only with the item's angular disparity to the participant, but also to the other person. Together, these data therefore confirm that people spontaneously represent other's visual perspectives in a (quasi-)perceptual manner that can "stand in" for own perceptual input. Once represented in such a manner, the others' view on the scene can drive item recognition and mental rotation processes like one's own perceptual input, facilitating item judgments that would be more difficult from one's own perspective.

The crucial question was whether these spontaneous shifts into the other's perspective depend on people's ability to shift their own body posture into the other's position, either because such movements physically more closely align one's own viewpoint with that of another, or because they trigger "embodied" rotation processes into the other's location in space. Prior research has shown that several cognitive processes, such as emotion recognition (e.g. Neal & Chartrand, 2011) or mathematical reasoning (Cook et al., 2012) are supported by the body movements people make at the same time, with performance decreasing if these movements are restricted or not compatible with the mental operation (Neal & Chartrand, 2011; Parsons, 1994). We did not, however, find this to be the case for visual perspective taking here. When participants' movements were restricted with a chinrest, the shifts into the others' perspective were just as strong as when participants were free to move, ruling out that the mechanisms which enable people to represent how a scene looks to another person rely on the ability to physically shift their head orientation.

These findings show, first, that perspective taking, as measured in our task (Ward et al., 2019), cannot simply be explained as a consequence of participants' physical alignment with the persons on the screen, which could potentially make item recognition easier. Several proposals argue that perception is not a passive process, but a process of "active inference" in which people actively try to sample the information required (e.g. Friston, 2010). Our data strongly rule out that, in our task, simulation of others' perspectives is achieved by bringing one's own perspective into physical alignment with that of the other person in such a manner.

Second, the findings offer some additional insights into the proposal that "embodied" processes play a causal role in visual perspective taking. Several recent studies have revealed that in order to take another's visual perspective, people have to mentally rotate their own body into the location and orientation of the other person, with time taken to judge another's perspective increasing the more this person is rotated away from oneself (Kessler & Thompson, 2010; Kessler & Rutherford, 2010; Kozhevnikov et al., 2006; Surtees et al., 2013). Our data do not argue against such a mental rotation process *per se*. They might suggest, however, that it is not one of motorically simulated bodily rotation, which would be affected by a person's perceived ability to move (Decety, 1995; Moreau, 2012; Parsons, 1994; Wohlschläger & Wohlschläger, 1998). Instead, our findings argue that people that shifts into the other's visual perspective do not draw on processes that track one's (perceived) ability to move their head to actively sample relevant input, but emerge from a more mental transformation into the other's space. The body movements of participants that we sometimes observed in our task are therefore likely to reflect an epiphenomenal "leakage" of these simulated changes in viewpoint, as is typically observed for other forms of imagined action (e.g., Colton et al., 2018; Jacobson, 1930).

It is important to note that whilst we do anecdotally report that participants moved their bodies when they were free to do so, that we did not include a condition in the present study, which actively instructed participants to move. While this leaves open whether actively inducing body movements would help or hinder visual perspective taking, it provides direct evidence that participants either do not make such movements spontaneously (in the free-movement condition) or that the inability to make these movements (in the no-movement condition) does not interfere with simulated changes in viewpoint.

'A limitation of this study is the overall strength of the chosen manipulation to test the embodied nature of VPT-2, specifically the use of a chin rest to restrict deliberate motion. Whilst a chin rest is a useful tool for minimising head movements, and therefore a good measure of how this particular aspect of embodied processing might influence VPT-2 calculations, the standard chin rest used in this study did not fully restrict the potential for a bodily rotation. In other words, whilst participants could not move their heads, they could, in principle, still align their torso with the persons appearing in the scenes. Indeed Kessler and Thompson (2010) directly employed manipulations in their work which have demonstrated that aligned or misaligned postures (participant relative to on-screen avatar), but *not head position* can, respectively, speed up or slow down allocentric judgements. Further work should address this by employing a more active embodiment manipulation, which would seek to directly test whether the findings of Kessler and Thompson (2010) are also apparent in implicit VPT-2, as measured by our task. For example, by manipulating the congruency of the rotation of the participants' own body (so that it is either closer or further away from the specific orientation of the avatar in the scenes) we could test directly whether people's tendency to spontaneously compute others' perspectives is driven by this postural, rather than head, alignment. Moreover, if perspective taking is affected by one's own body (but not head) rotation, it would go some way of revealing that perspective taking occurs through a mental transformation of one's of body (Kessler & Thompson, 2010), rather than through a rotation of the object in question (Shepard & Metzler, 1971).

A second, more exploratory, goal of the study was to determine whether individual differences can explain the stronger (or weaker) tendencies to take another's visual perspective across participants. To this end, we correlated our individual measures of perspective taking with common individual difference measures that have been empirically or conceptually linked to perspective taking and Theory of Mind, such as Schizotypical traits (STQ; Claridge & Brocks, 1984), Autistic traits (AQ; Baron-Cohen et al., 2001), or the ability to coordinate social interactions (IRI; Davis, 1983). Our data replicate the existing weak (negative) link between Schizotypical symptoms and problems with taking others' perspective (Langdon & Coltheart, 2001; Langdon et al., 2001). Measures of social ability and empathy (AQ and IRI), however, did not correlate with perspective taking, and Bayesian analyses provided considerable evidence *against* such a link.

These findings may be surprising in light of the proposed link between visual perspective taking and mentalizing and other coordination processes in social interactions (e.g., Batson et al., 1997; Erle & Topolinski, 2015; Tomasello et al., 2005; Mattan, et al., 2016; Freundlieb et al., 2016) and that those with autism find it more difficult to make perceptual judgments from another's visual perspective (Hamilton et al., 2009; for a review, see Pearson, Ropar & Hamilton, 2013). Of course, our task involved people with autistic traits only, without actual diagnoses of ASD. Moreover, it did not involve actual social interactions, so that any fundamentally "social" mechanisms may not be engaged as effectively. Nevertheless, it is noteworthy that even in prior literature, such relationships are inconsistent, and often found most robustly in children but not adults with ASD, implying that, while perspective taking might be delayed in ASD, it may have mostly caught up when participants reach adulthood, such as in the present sample. In addition, difficulties with representing another's view in ASD are usually observed in tasks in which people have to explicitly take others' perspectives (Pearson et al., 2013), but less so in more implicit tasks such as ours (e.g., Zwickel, et al., 2010), suggesting that those with ASD may have primarily problems in the more cognitively demanding process of intentionally *selecting* one of several possible perspectives (Ramsey, Hansen, Apperly & Samson, 2013; Schwarzkopf, Schilbach, Vogely & Timmermans, 2014; Qureshi, Apperly & Samson, 2010) rather than in perspective-taking per se.

If these considerations are taken seriously, than our data is more consistent with the view that perspective taking may not have specifically developed – either in ontogeny or phylogeny – to support social interactions, but may build upon a more fundamental process of navigation and action planning (e.g., Ward et al., under review; Kozhevnikov et al., 2006). To effectively act in the world, humans constantly need to be able to derive from which position they may be able to see an object clearly or operate on it effectively. Visual perspective taking may have developed from this basic skill to imagine the world from another location that one could occupy, with other people providing simple landmarks to drive these processes. Several findings seem to support such an account. For example, it has been known for a long while that people's ability to take another's perspective are correlated with navigation skills (Allen, Kirasic, Dobson, Long & Beck, 1996; Hegarty & Waller, 2004; Kozhevnikov et al., 2006) and it has been reported that people are as ready to view the world from another person's perspective as from the perspective of a landmark that supports navigation towards it, such as an empty chair (Gunalp, Moossaian & Hegarty, 2019). In neuroimaging studies of the mentalizing and perspective taking network, the temporo-parietal junction (TPJ) is also a node in navigation skills, and (virtual) lesion of this area can induce out of body experiences (Blanke et al., 2005), effectively moving oneself mentally into other possible locations one could occupy. In our own work with the present task, we have found that perspective taking is not sensitive to "social" features of the other person, such as whether they are looking at the item to be judged or not (Ward et al., under review), but specifically their location in space.

If these links were borne out by future research, it may suggest that at least the spontaneous shifts of perspective measured in our and the above tasks puts rely on fundamental spatial abilities, which put one into another's shoes, but do not necessarily let one see through their eyes. Future studies should include measures of navigational skill in perspective taking tasks. When these abilities are properly accounted for, and parcelled out, it may be possible to uncover the more social components that drive perspective taking. It may then be possible to describe not only how visual perspective taking has developed out of basic skills for spatial navigation, but also how more sophisticated processes for mentalizing and theory of mind build upon these processes, to help us understand other people better and interact with them more effectively.

4.2.4 Conclusions & Future directions

Our results confirm, first, that people represent others' viewpoints in a quasi-perceptual manner, such that other's perspectives can "stand in" for own input and drive subsequent item recognition and mental rotation processes. Second, they show that people can derive others' visual perspectives irrespective of whether they could move their heads to align with the other person, ruling out that this type of physical movement is necessary either to trigger "embodied" perspective taking processes or to physically align one's own perspective with that of the other person. Third, people's ability to perspective take is relatively independent from their autistic traits and their competency in coordinating social interactions, point towards a reliance on fundamental processes of mental travel.

4.2.5 References

Allen, G. L., Kirasic, K. C., Dobson, S. H., Long, R. G., & Beck, S. (1996). Predicting environmental learning from spatial abilities: An indirect route. *Intelligence*, *22*(3), 327-355.

American Psychiatric Association (APA). (2013). *Diagnostic and statistical manual of mental disorders* (5th ed., pp. 5–25). Arlington, VA: American Psychiatric Publishing.

Bach, P., Allami, B. K., Tucker, M., & Ellis, R. (2014). Planning-related motor processes underlie mental practice and imitation learning. *Journal of Experimental Psychology: General,* 143(3), 1277–1294. https://doi.org/10.1037/a0035604

Bach, P., Fenton-Adams, W., & Tipper, S. P. (2014). Can't touch this: the first-person perspective provides privileged access to predictions of sensory action outcomes. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 457.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of autism and developmental disorders*, *31*(1), 5-17.

Batson, C. D., Early, S., and Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imaging how you would feel. *Personality and Social Psychology Bulletin*. 23(7), 751-758.

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of cognitive neuroscience*, *11*(5), 551-559.

Blanke, O., Mohr, C., Michel, C. M., Pascual-Leone, A., Brugger, P., Seeck, M., ... & Thut, G. (2005). Linking out-of-body experience and self processing to mental own-body imagery at the temporoparietal junction. *Journal of Neuroscience*, *25*(3), 550-557.

Bukowski, H., Hietanen, J. K., & Samson, D. (2015). From gaze cueing to perspective taking: Revisiting the claim that we automatically compute where or what other people are looking at. *Visual cognition*, *23*(8), 1020-1042.

Claridge, G., & Broks, P. (1984). Schizotypy and hemisphere function—I: Theoretical considerations and the measurement of schizotypy. *Personality and Individual Differences*, *5*(6), 633-648.

Colton, J., Bach, P., Whalley, B., & Mitchell, C. (2018). Intention insertion: Activating an action's perceptual consequences is sufficient to induce non-willed motor behavior. *Journal of Experimental Psychology: General*, *147*(8), 1256.

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and cognitive processes*, *27*(4), 594-610.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), 113.

De Corte, K., Buysse, A., Verhofstadt, L. L., Roeyers, H., Ponnet, K., & Davis, M. H. (2007). Measuring empathic tendencies: Reliability and validity of the Dutch version of the Interpersonal Reactivity Index. *Psychologica Belgica*, *47*(4), 235-260.

Erle, T. M., & Topolinski, S. (2015). Spatial and empathic perspective-taking correlate on a dispositional level. *Social Cognition*. 33(3), 187-210.

Flavell, J. H., Everett, B. A., Croft, K., and Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Dev Psychol*. 17(1), 99.

Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2016). When do humans spontaneously adopt another's visuospatial perspective?. *Journal of experimental psychology: human perception and performance, 42(3)*, 401

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature Rreviews nNeuroscience*, *11*(2), 127.

Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference?. *PloS one*, *4*(7), e6421.

Gilet, A. L., Mella, N., Studer, J., Grühn, D., & Labouvie-Vief, G. (2013). Assessing dispositional empathy in adults: A French validation of the Interpersonal Reactivity Index (IRI). *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *45*(1), 42.

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science, 1(4)*, 586-596.

Gunalp, P., Moossaian, T., & Hegarty, M. (2019). Spatial perspective taking: Effects of social, directional, and interactive cues. *Memory & cognition*, 1-13.

Hamilton, A. F. D. C., Brindley, R., & Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition*, *113*(1), 37-44.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, *32*(2), 175-191.

Hughes, G., & Waszak, F. (2011). ERP correlates of action effect prediction and visual sensory attenuation in voluntary action. *Neuroimage, 56(3)*, 1632-1640.

Jacobson, E. (1930). Electrical measurements of neuromuscular states during mental activities. I. Imagination of movement involving skeletal muscle. *Am. J. Physiol. 91*, 567–608. doi: 10.1152/ajplegacy.1930.91.2.567

Kampis, D., Parise, E., Csibra, G., & Kovács, Á. M. (2015). Neural signatures for sustaining object representations attributed to others in preverbal human infants. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1819), 20151683.

Kessler, K., & Wang, H. (2012). Spatial perspective taking is an embodied process, but not for everyone in the same way: differences predicted by sex and social skills score. *Spatial Cognition & Computation*, *12*(2-3), 133-158.

Kessler, K., and Rutherford, H. (2010). The two forms of visuo-spatial perspective-taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology,. Psychol. 1*, 1–12.

Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88.

Kozhevnikov, M., Motes, M. A., Rasch, B., & Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *20*(3), 397-417.

Kunde, W., Koch, I., & Hoffmann, J. (2004). Anticipated action effects affect the selection, initiation, and execution of actions. Quarterly Journal of Experimental Psychology: Part A: Human Experimental Psychology, 57(1), 87.

Langdon, R., & Coltheart, M. (2001). Visual perspective-taking and schizotypy: evidence for a simulation-based account of mentalizing in normal adults. *Cognition*, *82*(1), 1-26.

Langdon, R., Coltheart, M., Ward, P. B., & Catts, S. V. (2001). Visual and cognitive perspective-taking impairments in schizophrenia: A failure of allocentric simulation?. *Cognitive Neuropsychiatry*, *6*(4), 241-269.

Mattan, B. D., Rotshtein, P., & Quinn, K. A. (2016). Empathy and visual perspective taking performance. *Cognitive neuroscience*, *7*(1-4), 170-181.

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural networks*, *9*(8), 1265-1279.

Moreau, D. (2013). Constraining movement alters the recruitment of motor processes in mental rotation. *Experimental brain research*, *224*(3), 447-454.

Morrison, I., Tipper, S. P., Fenton-Adams, W. L., & Bach, P. (2013). “Feeling” others' painful actions: the sensorimotor integration of pain and action information. *Human bBrain Mmapping*, *34*(8), 1982-1998.

Neal, D. T., & Chartrand, T. L. (2011). Embodied emotion perception: amplifying and dampening facial feedback modulates emotion perception accuracy. *Social Psychological and Personality Science*, *2*(6), 673-678.

Parr, T., & Friston, K. J. (2017). Working memory, attention, and salience in active inference. *Scientific reports, 7(1)*, 14678.Freundlieb, M., Sebanz, N., & Kovács, Á. M. (2017). Out of your sight, out of my mind: Knowledge about another person’s visual access modulates spontaneous visuospatial perspective-taking. *Journal of experimental psychology: human perception and performance*, *43*(6), 1065.

Parsons, L. M. (1994). Temporal and kinematic properties of motor behavior reflected in mentally simulated action. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(4), 709.

Pearson, A., Ropar, D., & Hamilton, A. F. D. (2013). A review of visual perspective taking in autism spectrum disorder. *Frontiers in Human Neuroscience, 7*, 652.

Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspectives on Psychological Science, 1*, 110–122

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition, 117(2)*, 230-236.

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 Neuroscience, 25(5)*, 670-684.

Roelfsema PR, and de Lange FP (2016). Early visual cortex as a multi-scale cognitive blackboard. *Annual Reviews in Vision Science*. 2,131-151.

Schwarzkopf, S., Schilbach, L., Vogeley, K., & Timmermans, B. (2014). “Making it explicit” makes a difference: Evidence for a dissociation of spontaneous and intentional level 1 perspective taking in high-functioning autism. *Cognition,* 131(3), 345-354.

Samson, D., Apperly, I.A., Braithwaite, J.J., Andrews, B.J., and Bodley Scott, S.E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *J Exp Psychol Hum Percept Perform*. 36(5), 1255.

Shepard, R. N., and Metzler, J. (1971). Mental rotation of three-dimensional objects. Science. 171(3972), 701-703.

Surtees, A., Apperly, I., & Samson, D. (2016). I’ve got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, *150*, 43-52.

Surtees, A., Apperly, I., & Samson, D. (2013). Similarities and differences in visual and spatial perspective-taking processes. *Cognition*, *129*(2), 426-438.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav Brain Sci*. 28, 675–691.

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition, 110(1)*, 124-129.

Vargas, C. D., Olivier, E., Craighero, L., Fadiga, L., Duhamel, J. R., & Sirigu, A. (2004). The Influence of Hand Posture on Corticospinal Excitability during Motor Imagery: A Transcranial Magnetic Stimulation Study. *Cerebral Cortex, 14(11)*, 1200–1206. https://doi.org/10.1093/cercor/bhh080

Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. *Current Biology*, *29*(5), 874-880.

Ward, E., Ganis, G., McDonough, K. L., & Bach, P. (2019, July 19). Perspective taking as virtual navigation? Perceptual simulation of how others see the world reflects their location in space but not their gaze. Retrieved from osf.io/2cxpr

Wohlschläger, A., & Wohlschläger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human perception and performance*, *24*(2), 397.

Zwickel, J., & Müller, H. J. (2010). Observing fearful faces leads to visuo-spatial perspective taking. *Cognition*, *117*(1), 101-105

Zwickel, J., White, S. J., Coniston, D., Senju, A., & Frith, U. (2010). Exploring the building blocks of social cognition: spontaneous agency perception and visual perspective taking in autism. *Social Cognitive and Affective Neuroscience, 6(5),* 564-571.

4.2.6 Supplementary tables

Table S3.1. IRI, AQ, and STQ scores, age, handedness (left, right), gender (male, female), dyslexia diagnosis (yes, no), ASD diagnosis (yes, no). Correlations (*r*) with perspective taking score *p<.05. **p<.01, ***p<.001

Variables	1	2	3	4	5	6	7	8	9
1. Perspective taking score	-								
2. AQ score	.08	-							
3. IRI score	.03	-.36**	-						
4. STQ score	-.26*	.19	.05	-					
5. Age	.38**	.24	-.08	-.06	-				
6. Handedness	.07	.18	-.01	.09	.04	-			
7. Gender	.11	-.05	.22	-.18	-.18	-.07	-		
8. Dyslexia	-.17	.02	-.28*	-.1	-.03	.38**	-.4**	-	
9. ASD	.11	.07	.05	-.09	.12	.57***	.07	-.02	-

[6.] coded as 1=left and 0=right [7.]coded as 1=female, 0=male [8.]coded as 1=yes, 0=no, [9.]coded as 1=yes, 0=no.

5. Chapter Five - General Discussion

Knowing how the world looks to another person is critical to our everyday social interactions. Representing space from another's perspective allows us to probe their epistemic stance, and may therefore allow us to engage in joint actions (Freundlieb et al., 2015), understand their beliefs, desires and goals (Hamilton et al., 2009; Schurz et al., 2015), and empathise with them (Mattan et al., 2016). Several studies have shown that the way we perceive a scene can be influenced by the presence of another person and the way they see it (Freundlieb et al., 2015; Freundlieb et al., 2018; Sampson et al., 2010; Surtees et al., 2016), indicating some form of overlap between what we see with our own eyes, and what we know about how an object or scene appears from another's perspective.

This aim of this thesis was to investigate the form that this overlap takes. One proposal that has previously been implied in many accounts of VPT but, prior to this work, never directly tested, is that we represent others' visual perspectives perceptually, in a similar way to our own. This explanation proposes that the representation of others' visual input takes a (quasi)-perceptual form, and that VPT 'paints' an image of the content of others' perspectives onto our own perceptual system. Indeed, several studies have shown reliable interference effects from the perspective of another person (e.g. Surtees et al., 2010; Sampson et al., 2016), where responses to judgements about a stimulus from one's own perspective are slowed down when this judgement would be made differently by someone else (e.g. judging a number as a "6" when it appears as a "9" to a person standing opposite oneself). However, what such studies have not been able to establish is at what level of processing these intrusions occur. Do they indeed reflect perceptual intrusions, or can they be accounted for by processes at the response or conceptual level?

Using a modified mental rotation task (e.g. Shepard & Metzler, 1971), the experiments presented in this thesis provide evidence that VPT-2 takes a (quasi)-perceptual form, and can indeed 'stand in' for our own perceptual input in the form of a perceptual intrusion, painting the content of another's perspective onto our own perception. These studies also showed that spontaneous simulation of others' perspectives happens specifically in the presence of a human, but not an object, even when this person is task-irrelevant (**Chapter 2).** This thesis also explored some of the social cues involved in the deployment of VPT-2 mechanisms. **Experiments 4a** and **4b** demonstrated that the perceptual representation of another's visual perspective reflects their position in space, but is very much insensitive to a person's gaze (**Chapter 3**). Finally, **Experiment 5** tested the causal role of movement in the simulation of others' visual perspectives (**Chapter 4**), revealing that implicit VPT-2 is not impaired by head-movement restriction, but rather head movements that arise when making judgements from another person's perspective are a bodily consequence of the mental transformations of a person's actual to imagined position in space.

Together these findings support a perceptual simulation framework for explaining VPT-2, and argue against the long-standing view that representing how the world looks to another person is a deliberate and cognitively demanding process. Instead, these data point to VPT-2 as being a spontaneous, low-level process which reflects someone's spatial position more generally, allowing us to efficiently derive how (in principle) another person perceives an object of interest from their individual vantage point.

5.1 Summary of results

Experiments **1a** and **1b** of this thesis tested whether people have direct (quasi-)perceptual access to the content of another's perspective. Participants were presented with a series of scenes showing either a table in the corner of a room, or the same scene but with a person inserted either to the left or the right of the table. An alphanumeric character, either mirror-inverted or normal, would then appear on the table either upright to the participant, or at one of seven other possible orientations away from upright. Participants simply judged whether the characters they observed were mirror-inverted or normal.

These, and all subsequent studies in this thesis, showed the classic mental rotation pattern, where participants' response times increased linearly as a function of increasing angular rotation away from upright (e.g. Shepard & Metzler, 1971), showing that it takes longer to make a judgement the more an item has to be mentally rotated into its upright position. Critically, these data show that the angular disparity relative to the other person also reliably produced this response pattern, showing that items were identified faster the more they were oriented towards the other person and slower the more they were oriented away from them. As a consequence, judgements that would normally be difficult from the perspective of the participants were faster if items were oriented towards the inserted persons, compared to when the other person was not present.

It has been argued that mental object rotation is known to operate using the same neural machinery as visual perception (Albers et al., 2013). When we mentally rotate an object of interest, we do so sequentially from the observed orientation to upright via a series of mental manipulations, which simulate the perceptual experience of actually *seeing* the object rotate. The recognition time data obtained in all experiments indicate that in the presence of another person, the entry point of this rotation reflects an integration of both own and others' perspectives, particularly when this other person would have a clearer view of the object of interest. This observation therefore supports the proposal that others' perspectives are derived as a perceptual simulation, which can then drive perceptual decision making in the same way as own input (e.g. Roelfsema & de Lange, 2016).

In Experiment **2** we manipulated between-subjects whether participants viewed another person (human condition), or a lamp in the scene (object condition). In the human condition we replicated the effects observed in Experiments **1a** and **1b**, implying again a perceptual overlap of own and others' perspectives. For the lamp, however, no bias was observed despite the lamp sharing similar directional properties (front/back) as a human agent. Participants did not show the response time benefit for items oriented towards the lamp, and moreover the position of the lamp did not reliably predict the mental rotation response pattern indicating that spontaneous perceptual simulation of others' visual perspectives is therefore a human-specific mechanism.

In Experiment **3** we used the same experimental paradigm as Experiments **1a** and **1b**, but for the simple change that we varied the instructions given to participants. Half of participants were only given instructions relating to the mental rotation task (implicit condition), thereby avoiding drawing attention to the presence of the people who would appear in the scene. The other half of participants were told explicitly to adopt the perspective of the actors when they appeared in the scene (explicit condition). As in earlier experiments, participants in both the implicit and explicit conditions showed a response time bias in the direction of the inserted persons. This bias was of equivalent size to previous studies (Experiments **1a – 2**) in the implicit condition, whilst in the explicit condition this bias was even more pronounced, implying that the bias observed when the actors were viewed passively was a consequence of participants' (involuntary) simulation of the content of the actors' perspectives. Overall, therefore, the preliminary work presented in Chapter Two of this thesis provide reliable support for a perceptual framework of VPT.

Subsequent experiments applied this same experimental paradigm to test which cues drive the spontaneous simulation of another's visual perspective. Experiment **4a** tested whether spontaneous shifts into the visual perspective of another person depend on whether or not the other person is looking at the to-be-judged item on the table. Here, we asked participants to perform the same mirror vs. canonical judgements of alphanumeric characters as before, but this time we manipulated the looking direction of the person on the screen. In some trials, the person looked directly at the item on the table, and in others the person looked away from the table towards the corner of the room.

As in previous experiments, the data replicated the finding that the mere presence of another person induced spontaneous shifts into their visual perspective. We found, again, that participants identified characters more quickly that were facing to the left if there was also another person pictured to the left, and vice versa for items that were facing the right. We also replicated the finding that the orientation of the items relative to both the participant and the inserted persons reliably predicted recognition times, with those oriented away from both the participant and the actor taking longer to judge compared to those appearing upright to the participant/actor. Surprisingly, this effect was not found to be mediated by where the actor was looking. In both table and away facing trials, participants adopted the perspective of the other person when making their responses.

Experiment **4b** replicated these findings. We used the same experimental design as experiment **4a** with the exception of one methodological difference. In this study we varied looking directions between, rather than within-subjects. The results from experiment **4a** can therefore not be explained by carry over (MacFie et al., 1989) or influential companion effects (Poulton, 1982). Indeed, they confirmed again that perceptual simulation of another's perspective is not sensitive to the gaze direction of the other person (i.e. whether they are looking towards or away from a to-be-judged item), but that we rather represent how the world *might* look from their position in space. These data therefore demonstrate that VPT does not necessarily reflect a looking 'through the eyes' of another person in order to represent their perspective, but rather being 'in their shoes' and simulating how the world might (in principle) look to them from their general position in space.

The final experiment (Experiment **5**) in this thesis tested whether the simulation of another's visual perspective is a motorically embodied process, as has been suggested in several studies (e.g. Kessler & Rutherford, 2010; Kessler & Thompson, 2010). Specifically, we tested whether people's spontaneous tendency to take another's visual perspective depends on the ability to make active body movements to physically align one's own perspective with that of another. In experiment **5** we therefore restricted participants' head movement during half of the experimental trials, and in the other half, participants were able to move freely whilst making their judgements. As in all previous experiments, we replicated the known mental rotation pattern (Shepard & Metzler, 1971), finding that response times (judging whether a character is canonical or mirror inverted) became gradually longer with increasing angular disparity away from upright to both the participant *and* the other person present in the scene. We also replicated the finding that judgements about items oriented upright *to the other person* were faster compared to when items were oriented away from them, showing once again that people spontaneously shift into the perspective of the other person, giving rise to faster than usual responses to (otherwise) difficult perceptual decisions. However, we did not find that these shifts were altered when participants were not able to actively shift their heads to better align themselves with the actors in the scenes. This suggests therefore, that the ability to change one's head position does not have a causal influence on our tendency to adopt another's visual perspective, and suggests instead that head movements that occur when taking another's perspective are an incidental bodily consequence of the mental transformation from a person's real to imagined position in space.

A secondary aim of experiment **5** was to test whether the previously reported relationships between VPT ability and measures of autistic traits (e.q. Hamilton et al., 2009), empathy (e.g. Mattan et al., 2016) and schizotypy (e.g. Langdon & Coltheart, 2001) would be observed in our implicit task. Prior work has demonstrated a link between social ability and VPT, where those found to underperform on measures of ToM, in particular false belief reasoning (e.g. Frith, 2001), also underperform in VPT tasks (Hamilton et al., 2009). We therefore asked participants to complete three self-report questionnaires which were then correlated against participants' individual VPT scores (the mean difference in left/right bias when a person is seen on the right of the table compared to when they appear on the left). We replicated the prior findings (Langdon & Coltheart, 2001; Langdon et al., 2001) that individual measures of schizotypic personality traits (STQ; Claridge & Broks, 1984) are negatively linked to

VPT performance. Our data showed a negative correlation between VPT and STQ scores, indicating that individuals with schizophrenic type thinking were less influenced by spontaneous perceptual intrusions of others' perspectives when making their judgements. We note, however, that Bayesian analyses could not fully support such a link, indicating a possibly spurious effect.

Surprisingly we found no relationship between social ability and spontaneous VPT. Neither the IRI (Davis, 1980), which measures empathy, nor the AQ (Baron-Cohen et al., 2001), which measures autistic traits, were found to correlate with VPT. Unsurprisingly, these scores were also not reliable predictors of VPT when entered into a regression model. Indeed, Bayesian analyses of these relationships provided fairly strong evidence *against* a link.

Together, the findings from the experiments presented in this thesis reveal substantial implications for our knowledge about (1) how we represent others' level-two visual perspectives, (2) the spontaneity of this process, (3) the cues which drive VPT-2, and (4) the role our own actions play in VPT. Indeed, whilst our results support a perceptual framework for explaining VPT, they also argue against several assumptions that have been made in prior literature, pertaining largely to the cues involved in the deployment of VPT (e.g. Freundlieb et al., 2016; 2017), and the mechanisms involved (e.g. Kessler & Rutherford, 2010; Kessler & Thompson, 2010; Michelon & Zacks, 2006; Surtees et al., 2013). This may indicate, then, that prior conclusions have been drawn prematurely or perhaps that the conclusions drawn from explicit measures of VPT-2 may not automatically apply to implicit VPT-2 as observed in our, and others' (e.g. Freundlieb et al., 2015; Surtees et al., 2016) tasks. The results of the experiments presented herein are discussed in relation to these assumptions in the following sections.

5.2 Implications

5.2.1 Visual perspective taking as perceptual simulation

The primary aim of this thesis was to test whether level-two VPT takes the form of a perceptual simulation, whereby (quasi)-perceptual intrusions from the perspective of another agent are painted onto our own perceptual system, driving perceptual decision making in the same bottom up way as own perceptual input, in the same way as has been implied by work on imagery in non-social situations (e.g., Csibra, 2008; Kok et al., 2013; Roelfsema & de Lange, 2016; Kosslyn et al., 1997). In recent years, such a view has been *implied* in several accounts of VPT (e.g. Freundlieb et al., 2018; Sampson et al., 2010; Surtees et al., 2016), however to our knowledge, no study has directly tested the perceptual nature of the 'overlap' in perspectives, often termed '*altercentric interference'* (Sampson et al., 2010).

We concluded from the experiments presented in this thesis that VPT takes a (quasi)-perceptual form, essentially 'painting' another's perspective onto our own perception. This conclusion rests on existing data which show that mental rotation of objects recruits low-level visual areas (e.g. Albers et al., 2013), wherein the primary representation of a seen object is sequentially manipulated to appear (in the mind's eye) at different orientations until the current image matches the known canonical representation of the judged object (Shepard & Metzler, 1971). Prior studies have shown an overlap in neural machinery for mental imagery and visual perception (e.g. Ganis et al., 2004; Kosslyn et al., 1997) indicating that visual simulations happen at the perceptual level. Note, however, that it is important to consider limitations in our understanding of mental imagery in general (Pylyshyn, 2002) when considering the completeness of any perceptual overlap that occurs during perspective taking. For example, it has not yet been resolved whether mental images reflect a pictoral or a symbolic representation of a scene or object (e.g. Pylyshyn, 2002).

Insofar as the mental rotation of objects is concerned, behaviourally, this mental transformation is demonstrated in the classic response time pattern that has reliably been observed in mental rotation tasks (Shepard & Metzler, 1971). Studies show that the time taken to mentally orient an object to its upright position increases linearly as a function of increasing angular disparity away from upright. Therefore in our own work, participants reliably took longer to judge the presentation of objects (e.g. “R” vs “Я”), the further they were rotated away from upright, and response times were reliably accounted for by the items’ angle of orientation relative to the participant. The finding that this same response time pattern was reliably predicted from the perspectives of the other people in the images provides critical evidence that VPT does indeed take a perceptual form. Rather than mentally rotating items to upright to themselves in our task, participants spontaneously derived how the items appeared to the other person, and used this alternative view as an entry point into the mental rotation. Responses were therefore faster when items were facing these other inserted persons, and slower when this judgement would be harder from the actor’s perspective. These data therefore imply that people are able to access to the content of another’s visual perspective and can rely on it to drive own recognition or mental rotation processes, as implied by perceptual accounts of perspective taking or imagery in general (Roelfsema & de lange, 2016; Kosslyn et al., 1997).

Our findings provide answers to the previous problems with our understanding of how others’ perspectives are represented, and how they drive behaviour in social interactions. By representing them in a (quasi)-perceptual manner, others’ perspectives can drive processes such as perceptual decision making (e.g.

Freundlieb et al., 2018) and action control in the same way as our own input (Kampis et al., 2015; Roelfsema & de Lange, 2016), explaining why we spontaneously shift into others' perspectives when making spatial judgements (e.g. Tversky & Hard, 2009), and also explaining the link between joint action and VPT (Creem-Regehr et al., 2013; Freundlieb et al., 2015). Our data also give insight into the link between VPT and more sophisticated social abilities such as Theory of Mind and false belief understanding (Hamilton et al., 2009; Schurz et al., 2015). By providing fast access to a simulation of others' view of the world in a (quasi)-perceptual manner, others' perspectives may allow us access to their epistemic state, and thereby facilitate reasoning about their beliefs, desires, goals and so on.

At a theoretical level, too, our data confirm the link between VPT and mental imagery, and provide further evidence that others' perspectives take the form of a perceptual simulation. Mental imagery is proposed to serve as a mechanism which allows us to re-live, and mentally manipulate a seen object or event by way of simulation (Moulton & Kosslyn, 2009). The studies in this thesis confirm that others' simulated perspectives can indeed provide the entry point to such mental manipulations, again explaining the link between joint actions and action understanding (Creem-Regehr et al., 2013; Freundlieb et al., 2015) both of which play an important role in facilitating efficient social interactions.

It is tempting to speculate that most, if not all previously reported cases of altercentric intrusions, irrespective if measured through response time lengthening (e.g. Freundlieb et al., 2018; Sampson et al., 2010; Surtees et al., 2016) or through spontaneous shifts into another's spatial perspective (e.g. Freundlieb et al., 2018; Tversky & Hard, 2009

) in the presence of another agent, can be accounted for by a perceptual framework. One example comes from Surtees et al. (2016). In their task, participants were required to judge the magnitude of a number either alone or in the presence of another person standing opposite them. They found that judgements about numbers that would have different *correct* interpretations dependent on perspective (e.g. “6” or “9”) were more difficult in the presence of another agent. The data strongly indicated an overlap in representations of items with perspective dependent interpretations (“6” or “9”), however unlike in our mental rotation task, it is unclear whether these overlapping representations only interfered with judgements, or whether they facilitated them, too.

When there were two conflicting interpretations of the presented number, this perceptual overlap of own and other perspectives appeared to slow down responses, but the numbers which did *not* differ between perspectives (“8” and “5”) were recognised more quickly in the presence of another person compared to when participants responded alone. One explanation for this finding is that ‘easy’ tasks (i.e. judging a number with only one possible interpretation) were sped up simply because another person was present. Zajonc (1965) proposed, and indeed others have demonstrated (e.g. Guerin & Innes, 1982; Rajeki, Ickes, Corcoran & Lenerz, 1977) that task performance generally can be improved in the presence of another person. However, the present work demonstrated that the presence of another agent in our task did not just improve mental rotation performance generally, but specifically facilitated (and interfered with) judgements based on their orientation relative to the actor. It is therefore possible that the acceleration in responses in Surtees et al.’s (2016) task may also be explained by a similar integration of own and others’ perceptual representations of the to-be-judged numbers, where those which match (“5” and “8”) are faster to disentangle than those which do not (“6” and “9”). This overlap could therefore also help to explain the slowing of responses in the presence of another person, where coinciding representations of conflicting input require participants to effortfully ‘select’ the correct perspective before making a response, resulting in relatively longer response durations. It is also possible, however, that the inverse effect of social facilitation, ‘social interference’ may have influenced these response time delays, and further research is needed here.

The same argument could be made for the work of Freundlieb and colleagues (2016; 2018). In one task Freundlieb et al. (2016) demonstrated that participants showed spatial compatibility effects *from the perspective of their co-actor* in a stimulus-response task, where matching spatial mapping of the response button and the presented stimulus is known to accelerate responses, and a mapping miss-match produces the opposite effect (e.g. Fitts et al., 1953). Here, participants responded to horizontally presented stimuli with a left/right button press. These same stimuli appeared laterally from the perspective of their partner, who was positioned 90 degrees to the left or the right of the participant. Therefore, responding to a ‘down’ stimulus with a button positioned on the right should not induce spatial compatibility effects, if participants just represented the stimuli from their own spatial perspective. However, Freundlieb et al. (2016) found that participants were indeed faster to respond to stimuli, which were spatially compatible to response buttons from the perspective of the *other* person. Conversely, responses to stimuli that were spatially incompatible from the perspective of the other person, were slowed down, demonstrating again an integration of own and others’ stimulus representations.

Although such a perceptual account of others’ findings (Freundlieb et al., 2016; Surtees et al., 2016) is tempting, it needs to be established whether they can be accounted for in such a manner. Our own representation of the world is not only derived from perceptual input, but also in terms of a scene's spatial make up (i.e. the distance between oneself and an object, or whether they appear to our left or our right). We also represent our own view of the world on a conceptual level (i.e. what objects are we seeing?) and in terms of the uncertainty we have about our own decisions. As noted above, from the findings themselves it is not clear whether in Surtees et al.'s study it is just uncertainty because one knows that the other person would respond differently. Also in Freundlieb et al.'s studies, it is not clear whether people really *simulate* the perspective of the other person, or whether they merely represent the spatial reference of the other person without representing what they see. Future work needs to resolve these questions.

5.2.2 Level-two VPT as an effortful controlled process

A further implication of our findings is that our data strongly argue against the proposal that VPT-2 is an effortful, controlled cognitive process (Flavell et al., 1981; Surtees et al., 2013). Instead, our findings support the proposal that people sometimes spontaneously derive how others view the world without explicit instructions to do so. Already in prior studies in implicit VPT-2 tasks, spontaneous intrusions from others' perspectives have been shown to influence how people respond to a stimulus, by slowing down responses when own and others' representations do not match (Surtees et al., 2016), or by causing someone to switch to an allocentric spatial reference frame when describing object locations in the presence of another person (Tversky & Hard, 2009). In all cases, these effects occur despite the other agent being largely task irrelevant, and with no explicit instructions for participants to take the perspective of the other person. In contrast, in explicit tasks, explicit judgements from another's perspective take longer with increasing angular disparity between the participant and the other agent (Surtees et al., 2013; Kessler & Rutherford, 2010; Kessler & Thompson, 2010). This suggests that in order to judge, for example, whether a flower appears to the left or the right of a gun from the perspective of another person (e.g. Kessler & Rutherford, 2010), it is necessary to mentally rotate one's own position into that of the other person. Indeed, such findings *do* point towards a controlled and effortful mental transformation as an underlying mechanism for explicit VPT-2 judgements.

One possible explanation is that the task demands of an explicit VPT-2 task, such as making a judgement about the position of two objects from another person's perspective (e.g. Surtees et al., 2013; Kessler & Rutherford, 2010; Kessler & Thompson, 2010) recruit different processes to those employed in tasks, including our own, which measure VPT-2 implicitly (e.g. Freundlieb et al., 2016; Freundlieb et al., 2017; Surtees et al., 2016). Indeed, this explanation is supported by evidence that suggests that individuals with ASD show no performance deficits in implicit measures of VPT but they do struggle in explicit measures (Schwarzkopf et al., 2014; Zwickel et al., 2010). This may therefore indicate that VPT-2 is derived spontaneously, but the selection of an appropriate perspective when making a response is itself effortful. In the case of individuals with ASD, this fits with the research which reliably shows that they are impaired in mentalizing abilities (e.g. Frith, 2001). Failure to disentangle representations of conflicting perspectives could therefore interfere with the processes which allow us to infer others' mental states and lead to failure in such tasks.

This explanation may help to explain data from studies which do not instruct participants to consider a conspecific's point of view, but nonetheless show that participants' responses are reliably influenced by their conflicting perspective of an object or scene (Freundlieb et al., 2015; Freundlieb et al., 2017; Surtees et al., 2016; Tversky & Hard, 2009). In Surtees et al.'s (2016) study for example, it is not clear why a person would engage in a controlled and effortful mental transformation from own to another's position in space without instruction to do so, if doing so would impair their ability to make fast, accurate judgements about a stimulus. However it is possible that the effortful process of selecting one of two representations causes a slowing of responses in this task, rather than the mental travel itself.

Similarly, in Freundlieb et al.'s (2018) word categorisation task, if implicitly deriving how a scene appears to another person is a cognitively effortful process there should be no difference in reaction times when the to-be-judged word appears upright and inverted to the other agent. This is not the case, however. Freundlieb et al. (2018) instructed participants to categorise a series of words that appeared on a table in front of them. These words appeared vertically from the participants' perspective, and dependent upon the seating position of a confederate (either 90degrees to the left or the right of the participant), either upright or inverted from their perspective. The data showed that judgements were faster when the word appeared upright to the confederate, but were relatively slowed down when they were inverted. Only when this perceptual decision would be harder for the other person (because the word appears upside down to them) are there any substantial response delays relative to seeing the word upright to this person, indicating that these judgements were disrupted by a spontaneous perceptual simulation which occurred independently of any effortful mental transformation in space.

In these, and our own tasks, we note that participants had ample time during the course of a trial to mentally reorient themselves into the position of the other person. For example, in our own task, the character of interest appeared in some cases almost 2000ms following the onset of the person in the scene. Similarly, in Surtees et al.'s number recognition task, and Freundlieb et al.'s (2016; 2017; 2018) studies, their partner was present during the course of the experiment, thereby allowing ample time for the rotation process into another's body prior to the onset of any to-be-judged item. We therefore cannot conclude from these studies the extent to which others' perspectives are spontaneously represented, or more specifically whether perceptual simulation of the others' perspective occurs independently of a mental rotation into their body. Future work should aim to manipulate stimulus onset asynchrony as has been done in variants of Sampson et al.'s (2010) avatar task (e.g. Bukowski et al., 2015) in order to investigate this further.

We also note that other studies have identified a mediating effect of the other person's actions when deriving their visual perspective. In many tasks, shifts are only observed when acting with the other person, or seeing them act in an intentional way (e.g. Freundlieb et al., 2016; Furlanetto et al., 2013; Surtees et al., 2016) in contrast to the other person having no involvement in the task, or merely gazing at an object of interest (e.g. Mazzarella et al., 2012; Tversky & Hard, 2009). The spontaneity of these shifts may therefore be driven by a requirement to anticipate the other person's actions and thereby encourage perspective taking, particularly if representing another's perspective helps one predict their upcoming actions. From the present work, it is not clear whether the continued presentation of our actor *in the same place* would induce the same level of perspective taking as observed in the current design. If, for example, observing an action does indeed encourage perspective taking, then we would expect our effects to be reduced when the position of the person on the screen does not change. Again, further work should seek to resolve this.

Despite these remaining questions, the spontaneity of the effects observed across our own, and others' (Freundlieb et al., 2016; Freundlieb et al., 2017; Surtees et al., 2016; Tversky & Hard, 2009) experiments do argue against VPT-2 being a cognitively controlled and effortful process, and point more towards a spontaneous, if not automatic, mechanism that rapidly computes how an object of interest appears from a non-egocentric perspective. Whilst we do not necessarily propose that implicit and explicit VPT-2 operate using different mechanisms altogether, it is instead more likely that when making explicit judgements about how a scene or object appears to another person, it is necessary to 'select' between own and others' representations, leading to slower responses. This may also explain why in explicit tasks we observe a reliable relationship between angular disparity and response times in studies which propose VPT-2 is calculated by way of a mental rotation into the position of another person (e.g. Kessler & Rutherford, 2010), but why in other implicit tasks (e.g. Freundlieb et al., 2016; Tversky & Hard., 2009), people shift into another person's spatial reference frame without considerable response time costs.

5.2.3 VPT: Seeing through another's eyes or wearing their shoes?

Few studies have directly tested whether a person's gaze direction (i.e. whether they are looking directly at an object of interest) influences our tendency to adopt their visual perspective, and results from these studies have been inconclusive. For example, Mazzarella et al. (2012) tested whether gaze at an object or action towards it more strongly influenced participants' tendency to shift into their spatial perspective when judging object locations. They found that, whilst observing a person reach for an object caused participants to judge the object location from the other person's perspective, simply seeing the other person gaze towards the object did not induce such a shift. In contrast, Furlanetto et al. (2013) showed that viewing someone gaze at an object did numerically increase shifts into the other's perspective, but also showed that the removal of gaze cues altogether did not affect whether participants responded from their own, or the other person's perspective.

The results from Experiments **4a** and **4b** in this thesis indicate that the gaze-direction of the other person does not influence people's tendency to adopt their visual perspective during perceptual decision making. Instead, our data suggest that spontaneous computations of others' visual perspectives reflect their position in space more generally, rather than their gaze. This is a surprising finding given the (folk-psychological) assumptions that VPT allows us to 'see' through another person's eyes (e.g. Furlanetto et al., 2016), and also given the well- established link between eye gaze and social perception more generally (e.g. Frischen et al., 2007; Gardner et al., 2018; Becchio et al., 2008). Our data suggest that such social gaze effects, which are strongly linked to mental state attribution, are separate to the perceptual shifts which underlie VPT-2, and therefore suggest a separate mechanism for mentalising and perceptual simulation of the content of others' visual perspectives.

Spontaneous VPT-2 calculations appear to operate using a mechanism which computes more generally how an item would, in principle, look from another's position in space, even when this person is not looking at the object of interest. This is consistent with the finding that during explicit VPT-2 tasks (e.g. Kessler &

Rutherford, 2010; Surtees et al., 2013), calculation of other's perspectives is reliant on a bodily mental rotation into the position of another person. Rather than linking perspective taking directly to Theory of Mind processes, our data link VPT to a more general, perhaps evolutionary, spatial-navigational ability, indicating that the primary function of this ability is to enable navigation of, and interaction with, our environment. Indeed, support for such a view has been reported in a number of studies which show a correlational relationship between VPT and navigation ability (Kozhevnikov et al., 2006) and also other studies that have found that the presence of an object which signals a position in space which one could occupy (e.g. a chair) can induce spatial shifts into this viewing perspective (e.g. Gunalp et al., 2019). Moreover, the TPJ has been linked to both VPT (Zacks et al., 1999), imagined navigation (Committeri et al., 2015), and to out of body experiences (Blanke et al., 2005), indicating that, though also implicated in Theory of Mind and mentalizing (Perner & Aichhorn, 2008), the primary function of VPT may be to compute what and how we can interact with in our environment.

Our ability to take another's perspective may therefore emerge from an evolutionary ancient mechanism, which emerged to help us to interact safely with our environment and plan actions (e.g. evading capture) from different positions in space. More sophisticated abilities for Theory of Mind and mentalizing may then capitalise on the emerging insights from others' perspectives, allowing us to identify epistemic differences between our own, and others' knowledge about the same scene, giving rise to the common neuronal network for Theory of Mind, VPT, self-other distinction, and navigation. By this view, our ability to perspective take serves both epistemic and action functions, perhaps explaining the link to false belief reasoning (Schurz et al., 2013), and also to joint action (Creem-Regehr et al., 2013).

5.2.4 Action precipitates VPT

Embodied accounts of VPT propose that our own actions, or posture, can interfere with our judgements when responding explicitly from another person's perspective. Work by Kessler and colleagues (2010; 2010) has, for example, demonstrated that when participants align their body position with that of another person, judgements from this other perspective are relatively easy, and vice versa if the postures of both individuals are misaligned these judgements take longer. This finding indicates a possible causal role of movement, or posture, in VPT. In some earlier experiments we noticed that some participants would make subtle postural adjustments during the course of the experiment. We therefore tested whether this movement had a causal influence on the observed perceptual shifts in our task. We tested this directly by restricting participants' movement when they were making judgements, and predicted that if movement was causally linked to the perceptual shifts observed in our earlier experiments, we would see a reduction in these shifts when participants were unable to physically simulate the position of the actor in the scenes.

Prior work investigating embodied mechanisms such as emotion recognition (Neal & Chartrand, 2011) have demonstrated that a person's ability to move, or motorically emulate, key gestures or expressions can interfere with our ability to identify others' emotional expressions. For example, people who are unable to move their own faces have been shown to perform worse than usual in emotion recognition tasks (Neal & Chartrand, 2011). Similarly, estimates of distances between locations are larger when participants are weighed down with a heavy load (Proffitt, 2006), indicating that our cognitive processes are influenced by our real bodily experiences, and critically our perceived ability to move (Decety, 1995; Moreau, 2012; Parsons, 1994; Wohlschläger & Wohlschläger, 1998).

While embodied accounts or VPT endorse a motorically embodied rotation into the position or another person when making allocentric judgements (e.g. Kessler & Rutherford, 2010), we did not find evidence in support of this view. Although we have shown in earlier experiments that a person's position in space, rather than where they are looking, is a critical cue to deploy VPT mechanisms (**Experiments 4a** and **4b**), we did not find that restricting participants' ability to move interfered with these effects (**Experiment 5**). We therefore do not argue against an embodied account of VPT *per se*, rather against a motoric basis for the mental transformation from a person's real to imagined position in space. Our findings suggest that the shifts we observed across these experiments do not draw on processes that track one's perceived ability to move their bodies, but emerge from mental transformation. We conclude, therefore, that the movements sometimes observed during our task are perhaps epiphenomenal leakage of simulated changes in viewpoint, as have been observed in other imagined actions (e.g. Colton et al., 2018; Jacobson, 1930).

It is important to now consider how these shifts might be affected by instructing participants to move their bodies, rather than to allow them to simply have the option to move. As yet, it is not clear whether actively inducing body movements would interfere with, or facilitate VPT, therefore future studies should consider this. Doing so would help to disentangle the processes which underlie explicit and implicit computation of the content of others' visual perspectives. We know from prior work, for example, that intentional postural alignment with a target 'other' does indeed give rise to faster judgements about object locations from the allocentric perspective, however it is not clear whether a similar manipulation would increase spontaneous shifts into the other person's perspective in our task. Such findings would provide large steps not only in supporting this mental rotation view of perspective taking, but would also provide further support of the link between perspective taking and navigation through space, as discussed above.

5.2.5 Individual differences in VPT

A final exploratory goal of this thesis was to examine whether individual differences in social functioning and schizotopy could explain the variance in tendency to take another's perspective in our task. We therefore entered participants' scores from a series of questionnaires (AQ; Baron-Cohen et al., 2001; IRI; Davis, 1983; STQ; Claridge & Brocks, 1984), which have been conceptually, or empirically linked to visual perspective taking in prior work (Hamilton et al., 2009; Langdon et al., 2001; Schurz et al., 2013; Schwarzkopf et al., 2014; Zwickel et al., 2011) into a multiple regression model to identify reliable predictors of VPT. Surprisingly, despite its assumed link to mentalizing and false belief reasoning (Schurz et al., 2013), VPT was not found to correlate with measures of social ability and empathy (AQ and IRI) in our task, though measures of schizotopy (STQ) did provide some contribution in explaining variance in VPT across participants, replicating prior work (e.g. Langdon et al., 2001; Langdon & Coltheart, 2001).

Though initially surprising, these findings support the underlying theme of this work, which proposes that visual perspective taking is not an inherently social mechanisms, but one that builds upon evolutionary ancient mechanism which may allow us to interact with our environment from multiple viewpoints. Already, we note above that there may be distinct differences between implicit perceptual intrusions from another's visual perspective, and the effortful act of selecting one of several representations of the same scene. Indeed, studies have shown that individuals with ASD, whom are characteristically impaired in social functioning, are not impaired in implicit measure of perspective taking (e.g. Schwarzkopf et al., 2014; Zwickel et al., 2011), but struggle with explicit tasks (Hamilton et al., 2009). We therefore do not propose that there is no link between social ability, Theory of Mind, mentalizing etc. and VPT, instead we speculate that these more sophisticated social abilities capitalise on a pre-existing mechanism which simply allows us to navigate and interact with our environment (e.g. Kozhevnikov et al., 2006). This mechanisms is then effortfully selected to solve social tasks.

5.3 Chapter Six - Future directions and remaining questions

This thesis provides by no means an exhaustive explanation of the mechanisms by which we derive the content of others' visual perspectives. We have, however, provided considerable evidence in support of a perceptual simulation account. Moreover, the evidence presented in this thesis support our speculation that representing others' visual perspectives draws on an ancient evolutionary mechanism which developed first to support navigation and interaction with the environment, and later developed to support more sophisticated mentalizing abilities. There are, of course, many remaining questions, some of which are outlined below either in terms of planned studies **(5.3.1 – 5.3.2)**, or theoretical questions **(4.3.4)**.

5.3.1 How complete is the perceptual simulation of others' input?

Since the publication of **Experiments 1a - 3** (Ward et al., 2019), there has been some backlash in the literature against the viability of our perceptual simulation account (e.g. Cole and Millet, 2019). In light of this, there are a number of actions that need to be taken to clarify the extent to which we truly simulate others' perceptual input.

One argument against a perceptual simulation account brought forward by Cole and Millet (2019) is that without utilizing a condition where the inserted actors cannot 'see' the item on the table, we cannot conclude that our effects demonstrate any representation of the content of the other's visual perspective. Whilst the **Experiments 4a** and **4b** in this thesis do indeed employ such a manipulation, it is important to consider possible limitations of this design, and moreover more fully explain why a turned away head (as in our own experiments) does not reduce VPT, but why in other studies (e.g. Freundlieb et al., 2017; Freundlieb et al., 2018) shifts into others' perspectives become much less likely when this other person is wearing occluding glasses. One possibility is that the occluding glasses prevent gaze at the relevant items in principle, while a simply averted gaze does not. After all, a turned away head can, in principle, turn to look at – and judge – an item at any point, whilst a person wearing a blindfold could not. One future study would then be to integrate Freundlieb and colleagues' alternative approach with our existing experiment, so that in some trials a blindfold is applied to the actors' head. This would test whether visual access, rather than gaze direction, is a critical cue in the deployment of spontaneous VPT-2 mechanisms. If our effects remain when the other person cannot see the item due to the blindfold, this would lend support to our proposal that perceptual simulations of another's perspective reflect a more general mechanism which allows us to represent how, in principle, an object of interest appears from an alternative vantage point. If, on the other hand, we saw a reduction in perceptual bias, this might indicate that VPT is driven by the perceived potential for another person to view, or act upon an object.

A further argument, again from Cole and Millet (2019), is that it is not possible that another's perspective could be perceptually represented *in full* when taking their perspective. It is not clear from the present work, for instance, whether the same effects would be observed with the use of different stimuli. Alphanumeric characters are commonly used in mental rotation studies and are particularly useful because they are easily recognised. However, often people already have experience in mentally transforming them (e.g. when reading a word that has been presented vertically or viewing mirrored characters in the rear view mirror of a car), and people could therefore simply rely on these memories of rotated letters rather than fully computing the others' perspective of the item. It would therefore be useful to replace these characters with other objects in future versions of this study for which such prior knowledge is unavailable. Nevertheless, a critical characteristic of these replacement stimuli would be to ensure that, like in our original study, the stimuli appear to be 'different' from the participant and the inserted person's perspective, allowing us to determine whether participants really represent the content of another's perspective, or whether the presence of another person merely triggers a shift into an alternative position from which they might recall a previously encoded version of the item at this specific orientation.

One possibility would be to replace our alphanumeric characters with faces (e.g. Thompson, 1980). Faces are a unique stimulus because we learn faces during ontogeny from an upright perspective, to the extent that face perception appears to be specialized for recognizing and distinguishing faces in this upright orientation only (e.g. the face inversion effect, (Diamond & Carey, 1986). The classic 'Thatcher Illusion' (Thompson, 1980; figure 4.0), which is similarly specific to faces in upright orientations could provide an excellent opportunity to further our understanding of the way we represent others' perspectives. In tasks employing the thatcher illusion (e.g. Edmonds & Lewis, 2007; Thompson, 1980; Lewis, 2001), participants are typically shown a series of faces, some of which are distorted (eyes and mouth inverted) and some of which are un-edited, and are asked to state which faces are normal, and which are grotesque. The classic finding is that people find it more difficult to distinguish between normal and grotesque faces when they are rotated by 180 degrees because it is necessary to mentally rotate the images back to upright before making a judgement. As such, reaction times follow a similar linear pattern to those seen in classic mental rotation tasks, where decisions take longer as a function of increasing angular disparity away from upright (e.g. Lewis, 2001).

Figure 4.0. **Example of 'Thatcherised' face stimuli used by Thompson (1980)**

A planned future study will therefore utilize these stimuli to test the extent to which we represent others' perspectives during perceptual decision making. As the central manipulation, and therefore the best indicator of the 'strength' of the Thatcher illusion, is to fully invert the face stimuli, the faces in this future study will only be presented at 0 or 180 degrees. Participants will be shown grotesque and unedited faces, and using a simple button press will be asked to indicate whether the observed faces are ‘normal’ or ‘grotesque’. In half of the trials, participants will perform the task standing adjacent to another person (same perspective condition), and in the other half, participants will be positioned opposite one another (different perspective condition). If we spontaneously derive how the faces appear to the other person, then judgements about grotesque faces which appear upright to the other person (and therefore inverted to ourselves) will be faster compared to when participants share the same view of the item. Similarly, judgements about upright faces will be slower when this same face appears inverted to the other person. If participants are therefore less ‘fooled’ by the illusion when they are able to recruit another’s perspective during this task, this would provide confirmatory evidence that others’ perspectives take the form of a perceptual simulation, and would strengthen the argument for spontaneous VPT-2. If, on the other hand, people are unaffected by the presence of another individual (with an inconsistent view of the object of interest), this will shed light on the limits of such simulations. For example, it is possible that it is necessary that we have extensive prior experience with the stimuli (as with alphanumeric characters) if we are to accurately represent, and integrate, others’ perspectives when making judgements. The above proposal would help to solve this problem, and would indeed contribute to the generation of a more comprehensive theoretical model of a perceptual simulation framework of visual perspective taking.

5.3.2 How is VPT embodied?

Do we assume ownership of the other person’s body when taking their perspective or do they represent a position in space that we could occupy?

We have shown across these studies that a person's position in space, rather than their gaze (**Experiments 4a and 4b**) or one's own ability to move (**Experiment 5**) is central in deriving how the world appears from this alternative perspective. This has led to speculation that taking another's perspective may have evolved as a general mechanism which allows us to navigate and interact with our environment rather than one which has developed specifically to support complex Theory of Mind processes (Hamilton et al., 2009). Indeed, other studies point towards this view. For example, some studies have demonstrated a link between navigational skill and VPT (e.g. Kozhevnikov et al., 2006), and others have indicated that occupiable spaces (e.g. an empty chair) can induce spatial shifts akin to perspective taking (Gunalp et al., 2019). However, it remains to be tested whether this process is 'social' rather than 'spatial', and specifically whether, when taking another's perspective, we assume ownership of this other person's body (thereby allowing us to simulate their world view) or whether they simply represent a more general position in space that we could – in principle – occupy. We therefore propose a series of studies below, which will help to disentangle this problem.

We will first test whether the recognition-time benefits observed in our prior work are exclusively observed in the presence of another person, or whether the same pattern emerges when the person is substituted for a chair**.** Our first study will test directly whether the mechanism driving these effects is socially embodied (rotating into another person's body) or spatially embodied (rotating into a space that could easily be occupied by oneself). We will replicate our previous method, but for one critical manipulation. Between subjects, we will vary whether in half of the trials, another person, or a chair (representing a position in space that is inviting the participant to occupy it) appears either to the left, or the right of the participant. If

VPT is driven by a socially embodied rotation into another's position, judgements about items oriented towards this person, but not the chair, will be faster than when this item is oriented away from them. If, on the other hand, the inserted person signifies a space which could otherwise be occupied by oneself, then we will see the same response time benefits in the presence of the chair as we do this other person. A further experiment will build upon this by making the same comparisons between a person, a usable chair, and one which is not safe to sit in (e.g. covered with barbed wire). If VPT is driven by the embodiment of another person, we will only see RT benefits for the inserted person. In contrast, if VPT is a spatially embodied, navigational process, by which people mentally rotate themselves into a space that they *could* occupy, we will see RT benefits for the person and the 'safe' chair, but not the 'dangerous' chair.

To assess the social nature of VPT, we will also test whether those who easily assume ownership over others' bodies are more strongly influenced by their perspective during perceptual decision making. We will test whether participants' susceptibility to the Rubber Hand Illusion (Botvinick & Cohen, 1998) predicts VPT as indexed in our task. Participants will observe a rubber hand being stroked, whilst their own concealed hand will also be stroked. Participants will be asked to rate how much they felt that the rubber hand was their own. We will then replicate the first experiment using only trials with a person inserted, and trials featuring just the table. If VPT is facilitated by the embodiment of another person, then people who are more susceptible to such body transfer illusions will show more perspective taking than those who are not so fooled by the RHI. A striking further manipulation would be to induce an outer body illusion but stroking the participant and the other person in the scene at the same time. If perspective taking is a means of inducing body ownership, or reflects ownership of the other's body, then such a manipulation could increase perspective shifts into the other person.

5.3.3 What role do others' actions play in VPT?

Several studies have shown that people are more sensitive to others' visual perspectives when this other person is performing an action (e.g. Surtees et al., 2016; Tversky & Hard, 2009; Freundlieb at al., 2016), however it is not clear whether others' actions serve as a cue to perspective take, or whether the perceived intentionality of an action (compared simply with movement) gives rise to spontaneous perspective taking. Several studies have tested this (e.g. Freundlieb et al., 2016; Surtees et al., 2016; Tversky & Hard, 2009) but the results overall have been inconclusive. During Freundlieb et al.'s (2016) spatial compatibility task, for example, they found that participants would only shift into a confederate's perspective if they perceived them to be an intentionally acting agent. Similarly, Tversky and Hard (2009) found that the mere presence of another person was enough to cause participants to describe the spatial relationships of objects from the other person's perspective. Importantly, when these other people appeared to reach from one of the objects, this shift into the other person's spatial reference frame was even more pronounced, suggesting that a person's action intentions towards an object might play a causal role in VPT. These findings align with studies which link VPT to Theory of Mind process (e.g. Furlanetto et al., 2016; Hamilton et al., 2009), as VPT might play a role in helping the participant understand the actions they observe (e.g., Jeannerod, 2007).

Other studies did not find this relationship, however. Surtees et al. (2016) found in their number recognition task that the aspect of the stimulus that participants were sensitive to (the magnitude) did not have to match the aspect to which their partner was attending (e.g. colour). Regardless of the task being performed by their partner, participants' responses were consistently influenced by the other person's perspective when making their judgements. This suggests a more primitive explanation, whereby the content of others' perspectives is derived spontaneously in response to *any* movement from one's partner. Similarly, in Freundlieb et al.'s (2016) task, the movement, rather the intentional actions, of the confederate may have caused a similar process of regular spatial updating. One possible explanation may then be that it is not action *per se* that drives these effects, more so it is the relative salience of movement which causes participants to continuously update their spatial reference frame, and thereby rotate themselves into the other perspective.

In contrast, in the studies presented in this thesis, the individuals in the images did not reach for the items, nor did they 'move' in any way (with the exception of **Experiments 1a** and **1b**, which was shown in subsequent studies to have no impact on the effects observed). Despite this lack of action, and their irrelevance to the task, the mere presence of the actors in our studies reliably induced VPT across all experiments. We note, however, that whilst there was no action, the random distribution of the trials in our experiments, varying the appearance of a person on either the left or the right if the screen, may have caused participants to spontaneously derive *in each trial* how the scene appeared from the others' perspective, in contrast to these other experiments in which the actor location was usually blocked.

To test whether VPT is triggered by a process of spatial updating (e.g. Wolbers, Hegarty, Buchel & Loomis, 2008), a future version of our mental rotation experiment may therefore adapt the trial distribution such that the actors appear in the same place across blocks of trials. This would eliminate any movement, and would clarify whether the unpredictable position of the actor played a causal role in participants' tendency to take the other's perspective in the present work. Further experiments could introduce an explicit action manipulation. For example, rather than using a computer based task to test spontaneous VPT, in future versions of this task we will project our mirrored and cannonical alphanumeric characters onto a table, and will ask two participants to take part at the same time. By counterbalancing participants' positions relative to one another (left or right) throughout the experiment, we will replicate the existing experiment design in a 'live' setting, thereby allowing us to directly test whether others' actions induce spontaneous shifts into the other's visual perspective. Whilst Surtees et al. (2016) have demonstrated that participants need not be engaged in the same task to induce perspective taking, Tversky & Hard (2009) and Freundlieb et al. (2016) found evidence indicating that the perceived intentionality of observed actions might induce spontaneous shifts into another's spatial frame of reference. Our task could play a unique role in disentangling these possibilities.

5.3.4 VPT-2 is a perceptual simulation, but how do we get there?

The experiments outlined in this thesis provide substantial evidence for a perceptual simulation account of visual perspective taking, but *how* this simulation is achieved remains unanswered. For example, we cannot say from the data presented herein whether we really do have ‘immediate’ access others’ visual perspectives, or indeed whether *any* transformation is necessary to derive what others can see. Although overall the experiments in this study suggest that people do rapidly and spontaneously compute others’ perspectives, there are limitations to the current design if we are to attempt to conclusively argue that VPT-2 is derived in a ‘direct’ or ‘immediate’ way. For example, in all the experiments described throughout this thesis, participants had a minimum of 1500ms to align their perspectives with those of the people in the scenes (i.e. between appearance of the person and the letter to be judged). Although the evidence for VPT-2 as a spontaneous, unintentional process is clear throughout this thesis, the speed at which these computations are achieved remains therefore in question, with an upper limit of 1500 ms. A simple solution to this problem would be to manipulate the stimulus onset asynchrony (SOA) between the onset of the person in the scene, and the to-be-judged item. This could be varied between trials at SOAs of, for example, 0ms, 300ms, 600ms and 1000ms. If others’ perspectives can be accessed immediately, then the shifts into the other person’s perspective would not differ between SOAs, and the RT pattern relative to the onset of the item would not vary between conditions. If, however, access to others’ simulated perspectives is not ‘immediate’ or ‘direct’ but relies on a (costly) transformation of space, then a different pattern would be more likely, where evidence of VPT might only be apparent at longer SOAs

In the introduction, we briefly introduced the distinction between self-rotation (Kessler & Thompson, 2010) and object-centred rotations (Kovhevnikov et al., 2006), where the former describes how we mentally transport our bodies into a different position to sample a scene from an alternative viewpoint, and the latter describes that we anchor ourselves in space, and mentally rotate a scene or object so we can ‘observe’ it from a desired angle. From the data presented in this thesis, we cannot conclude which method of mental transformation is employed in solving our task. It is important to note, however, that regardless of whether the relevant mental transformation is anchored in one's own perspective (object-based transformation) or if VPT-2 is achieved by rotating oneself into the position of another person (self-rotation), the end result would not differ. The simulated perceptual input that drives perceptual decision making in our task would be the same. Therefore, for the purpose of the present work the two processes are viewed interchangeably and the distinction is mentioned here simply with the goal of clarification.

The mechanisms preceding the perceptual simulation of others' perspectives, specifically the presence of a rotation transformation could be tested by manipulating the angular disparity between the participant and the person appearing on the screen. Prior work has shown that response times for explicit allocentric judgements increase linearly as a function of the rotational distance between the 'self' and 'other' perspectives (e.g. Michelon & Zacks, 2005; Kessler and colleagues, 2010; 2010). If the same rotational processes are involved in implicit VPT-2 computations, then this same linear relationship should be present in our task. Here, a future study could simply increase or decrease the angular disparity between the participant and the person appearing in the scene and measure whether shifts in the others' perspective become apparent earlier in time than for people sitting further apart. This would provide further insights into the mechanisms that underpin implicit VPT-2. Further studies could instruct participants explicitly to take the alternative perspective, either by performing a mental self-rotation, or by mentally rotating the scene towards themselves. This, again, might help to disentangle which, if any, rotational transformation occurs during VPT-2.

To conclude, we can speculate that people have immediate (quasi)perceptual access to others' visual perspectives, and that this is achieved by performing a rapid, mental transformation in space, however further work needs to be done to address the mechanics of the transformation that enables such rapid computations of others' visual perspectives.

5.4 Summary and concluding remarks

While it has been inferred in several past accounts of visual perspective taking, the experiments presented in this thesis provide the first known evidence that the content of others' visual perspectives is represented similarly to our own, and takes the form of a perceptual simulation that can drive perceptual decision making in the same way as our own input (Roelfsema & de Lange, 2016). This work has therefore substantially advanced our knowledge about the overlapping representations of own and others perspectives and, using a perceptual framework, helps to explain why others' perspectives can interfere with our own, even when we are not instructed to consider them (Freundlieb et al., 2016; Sampson et al., 2010; Surtees et al., 2016; Tversky & Hard, 2009).

The data presented in the latter part of this thesis have advanced our understanding of the mechanisms involved in deriving how the world looks to others. In contrast to some folk-psychological accounts (e.g. Furlanetto et al., 2016), which imply that we 'see through others' eyes' when taking their perspective, we found here that the mere presence of another person is sufficient to induce spontaneous simulation of their perspective. Moreover, we found that measures of social ability and empathy are not reliably linked to implicit VPT-2 as measured in our task, raising further questions about its evolutionary basis.

The experiments presented in this thesis have provided a strong entry point from which new knowledge and research can now be developed, in particular in relation to the social vs. spatial nature of VPT. Based on our findings we conclude that (1) deriving what another person can see occurs spontaneously in the presence of another person, (2) VPT takes the form of a perceptual simulation, and (3) speculate that VPT emerged primarily to support navigation of, and interaction with the environment. Future studies should now focus on disentangling the spatial from the social determinants of VPT deployment, and investigate the possibility that VPT emerged to support navigation and action planning.

5.5 References

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., and de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. Curr. Biol. 23(15), 1427-1431.

Allen, G. L., Kirasic, K. C., Dobson, S. H., Long, R. G., & Beck, S. (1996). Predicting environmental learning from spatial abilities: An indirect route. *Intelligence*, *22*(3), 327-355.

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. (2018). Raincloud plots: a multi-platform tool for robust data visualization. *PeerJ Preprints*, *6*, e27137v1.

Ambrosini, E., Costantini, M., & Sinigaglia, C. (2011). Grasping with the eyes. *Journal of neurophysiology*, *106*(3), 1437-1442.

American Psychiatric Association (APA). (2013). *Diagnostic and statistical manual of mental disorders* (5th ed., pp. 5–25). Arlington, VA: American Psychiatric Publishing.

Amorim, M. A., Isableu, B., & Jarraya, M. (2006). Embodied spatial transformations:" Body analogy" for the mental rotation of objects. *Journal of Experimental Psychology: General*, *135*(3), 327.

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. Psychological Science, 28(11), 1547-1562.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states?. *Psychological review*, *116*(4), 953.

Astington, J. W., & Jenkins, J. M. (1995). Theory of mind development and social understanding. *Cognition & Emotion*, *9*(2-3), 151-165.

Bach, P., Allami, B. K., Tucker, M., & Ellis, R. (2014). Planning-related motor processes underlie mental practice and imitation learning. *Journal of Experimental Psychology: General,* 143(3), 1277–1294. https://doi.org/10.1037/a0035604

Bach, P., Fenton-Adams, W., and Tipper, S.P. (2014). Can't touch this: the first-person perspective provides privileged access to predictions of sensory action outcomes. J Exp Psychol Hum Percept Perform. 40(2), 457.

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*(4), 508.

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

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of autism and developmental disorders*, *31*(1), 5-17.

Batson, C. D., Early, S., and Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imaging how you would feel. *Personality and Social Psychology Bulletin*. 23(7), 751-758.

Bayliss, A. P., Paul, M. A., Cannon, P. R., & Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic bulletin & review*, 13(6), 1061-1066.

Becchio, C., Bertone, C., & Castiello, U. (2008). How the gaze of others influences object processing. *Trends in Cognitive Sciences*, *12*(7), 254-258.

Becchio, C., Del Giudice, M., Dal Monte, O., Latini-Corazzini, L., and Pia, L. (2011). In your place: neuropsychological evidence for altercentric remapping in embodied perspective taking. Social cognitive and affective neuroscience, 8(2), 165-170.

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of cognitive neuroscience*, *11*(5), 551-559.

Blanke, O., Mohr, C., Michel, C. M., Pascual-Leone, A., Brugger, P., Seeck, M., ... & Thut, G. (2005). Linking out-of-body experience and self processing to mental own-body imagery at the temporoparietal junction. *Journal of Neuroscience*, *25*(3), 550-557.

Boccia, M., Sulpizio, V., Palermo, L., Piccardi, L., Guariglia, C., & Galati, G. (2017). I can see where you would be: patterns of fMRI activity reveal imagined landmarks. *Neuroimage*, *144*, 174-182.

Borke, H. (1975). Piaget's mountains revisited: Changes in the egocentric landscape. *Developmental Psychology*, *11*(2), 240.

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

Bruner, J. S., & Goodman, C. C. (1947). Value and need as organizing factors in perception. *The journal of abnormal and social psychology*, *42*(1), 33.

Bukowski, H., Hietanen, J. K., & Samson, D. (2015). From gaze cueing to perspective taking: Revisiting the claim that we automatically compute where or what other people are looking at. *Visual cognition*, *23*(8), 1020-1042.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology*, *76*(6), 893.

Christophel, T. B., Cichy, R. M., Hebart, M. N., and Haynes, J. D. (2015). Parietal and early visual cortices encode working memory content across mental transformations. NeuroImage. 106, 198-206.

Claridge, G., & Broks, P. (1984). Schizotypy and hemisphere function—I: Theoretical considerations and the measurement of schizotypy. *Personality and Individual Differences*, *5*(6), 633-648.

Cole, G. G., & Millett, A. C. (2019). The closing of the theory of mind: A critique of perspective-taking. *Psychonomic bulletin & review*, 1-16.

Cole, G. G., Atkinson, M., Le, A. T., and Smith, D. T. (2016). Do humans spontaneously take the perspective of others?. Acta Psychol (Amst). 164, 165-168.

Colton, J., Bach, P., Whalley, B., & Mitchell, C. (2018). Intention insertion: Activating an action's perceptual consequences is sufficient to induce non-willed motor behavior. *Journal of Experimental Psychology: General*, *147*(8), 1256.

Committeri, G., Piccardi, L., Galati, G., & Guariglia, C. (2015). Where did you "left" Piazza del Popolo? At your "right" temporo-parietal junction. Cortex, 73, 106-111.

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and cognitive processes*, *27*(4), 594-610.

Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology*, *7*(1), 20-43.

Craver-Lemley, C., & Reeves, A. (1992). How visual imagery interferes with vision. *Psychological Review*, *99*(4), 633.

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.

Csibra, G. (2008). Action mirroring and action understanding: An alternative account. *Sensorymotor foundations of higher cognition. Attention and performance XXII*, 435-459.

David, N., Bewernick, B. H., Cohen, M. X., Newen, A., Lux, S., Fink, G. R., ... & Vogeley, K. (2006). Neural representations of self versus other: visual-spatial perspective taking and agency in a virtual ball-tossing game. *Journal of Cognitive Neuroscience*, *18*(6), 898-910.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), 113.

De Corte, K., Buysse, A., Verhofstadt, L. L., Roeyers, H., Ponnet, K., & Davis, M. H. (2007). Measuring empathic tendencies: Reliability and validity of the Dutch version of the Interpersonal Reactivity Index. *Psychologica Belgica*, *47*(4), 235-260.

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and cognitive neuroscience reviews*, *3*(2), 71-100.

Decety, J., Chen, C., Harenski, C., & Kiehl, K. A. (2013). An fMRI study of affective perspective taking in individuals with psychopathy: imagining another in pain does not evoke empathy. *Frontiers in human neuroscience*, *7*, 489.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General*, *115*(2), 107.

Drayton, L. A., Santos, L. R., & Baskin-Sommers, A. (2018). Psychopaths fail to automatically take the perspective of others. *Proceedings of the National Academy of Sciences*, *115*(13), 3302-3307.

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.

Eack, S. M., Wojtalik, J. A., Keshavan, M. S., & Minshew, N. J. (2017). Social-cognitive brain function and connectivity during visual perspective-taking in autism and schizophrenia. *Schizophrenia research*, *183*, 102-109.

Edmonds, A. J., & Lewis, M. B. (2007). The effect of rotation on configural encoding in a face-matching task. *Perception*, *36*(3), 446-460.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. Behavior Research Methods, Instruments, & Computers, 28(1), 1-11.

Erle, T. M., & Topolinski, S. (2015). Spatial and empathic perspective-taking correlate on a dispositional level. *Social Cognition*. 33(3), 187-210.

Fitts, P. M., & Seeger, C. M. (1953). SR compatibility: spatial characteristics of stimulus and response codes. *Journal of experimental psychology*, *46*(3), 199.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology*, *17*(1), 99.

Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2016). When do humans spontaneously adopt another’s visuospatial perspective? *Journal of experimental psychology: human perception and performance*, *42*(3), 401.

Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2018). Reading Your Mind While You Are Reading—Evidence for Spontaneous Visuospatial Perspective Taking During a Semantic Categorization Task. *Psychological science*, *29*(4), 614-622.

Freundlieb, M., Sebanz, N., & Kovács, Á. M. (2017). Out of your sight, out of my mind: Knowledge about another person's visual access modulates spontaneous visuospatial perspective-taking. *Journal of experimental psychology: human perception and performance*, *43*(6), 1065.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic bulletin & review*, *5*(3), 490-495.

Friesen, C. K., Moore, C., & Kingstone, A. (2005). Does gaze direction really trigger a reflexive shift of spatial attention?. *Brain and cognition*, *57*(1), 66-69.

Frischen, A., Loach, D., & Tipper, S. P. (2009). Seeing the world through another person's eyes: Simulating selective attention via action observation. *Cognition*, *111*(2), 212-218.

Frischen, Alexandra, Andrew P. Bayliss, and Steven P. Tipper. (2007) Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin* 133.4 (2007): 694.

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature Rreviews nNeuroscience*, *11*(2), 127.

Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference?. *PloS one*, *4*(7), e6421.

Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology*, *17*(16), R724-R732.

Frith, C., & Frith, U. (2005). Theory of mind. *Current biology*, *15*(17), R644-R645.

Frith, U. (2001). Mind blindness and the brain in autism. *Neuron*, *32*(6), 969-979.

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*(2), 158.

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.

Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Cognitive Brain Research*, *20*(2), 226-241.

Gardner, M. R., Hull, Z., Taylor, D., & Edmonds, C. J. (2018). 'Spontaneous' visual perspective-taking mediated by attention orienting that is voluntary and not reflexive. *Quarterly Journal of Experimental Psychology*, *71*(4), 1020-1029.

Gardner, M., Bileviciute, A., & Edmonds, C. (2018). Implicit mentalising during level-1 visual perspective-taking indicated by dissociation with attention orienting. *Vision*, *2*(1), 3.

Gilet, A. L., Mella, N., Studer, J., Grühn, D., & Labouvie-Vief, G. (2013). Assessing dispositional empathy in adults: A French validation of the Interpersonal Reactivity Index (IRI). *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *45*(1), 42.

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science, 1(4)*, 586-596.

Gronholm, P. C., Flynn, M., Edmonds, C. J., and Gardner, M. R. (2012). Empathic and non-empathic routes to visuospatial perspective-taking. Conscious Cogn. 21(1), 494-500.

Guerin, B., & Innes, J. M. (1982). Social facilitation and social monitoring: A new look at Zajonc's mere presence hypothesis. *British Journal of Social Psychology*, *21*(1), 7-18.

Gunalp, P., Moossaian, T., & Hegarty, M. (2019). Spatial perspective taking: Effects of social, directional, and interactive cues. *Memory & cognition*, 1-13.

Hamilton, A. F. D. C. (2016). Gazing at me: the importance of social meaning in understanding direct-gaze cues. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150080.

Hamilton, A. F., Brindley, R., and Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition* 113, 37–44.

Happé, F. G. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child development*, *66*(3), 843-855.

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*(4), 771-785.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, *32*(2), 175-191.

Hudson, M., McDonough, K. L., Edwards, R., & Bach, P. (2018). Perceptual teleology: expectations of action efficiency bias social perception. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1884), 20180638.

Hughes, G., & Waszak, F. (2011). ERP correlates of action effect prediction and visual sensory attenuation in voluntary action. *Neuroimage, 56(3)*, 1632-1640.

Hughes, M. (1975). Egocentrism in preschool children.

Jacobson, E. (1930). Electrical measurements of neuromuscular states during mental activities. I. Imagination of movement involving skeletal muscle. *Am. J. Physiol. 91*, 567–608. doi: 10.1152/ajplegacy.1930.91.2.567

Janzen, G., Schade, M., Katz, S., & Herrmann, T. (2001). Strategies for detour finding in a virtual maze: The role of the visual perspective. *Journal of Environmental Psychology*, *21*(2), 149-163.

JASP Team (2018). JASP (Version 0.10.0)[Computer software].

Jeannerod, M. (2007). Being oneself. Journal of Physiology-Paris, 101(4-6), 161-168.

Jeffreys, H. (1961). Theory of Probability, 3rd Edn Oxford: Oxford University Press.[Google Scholar].

Jolicoeur, P., & Cavanagh, P. (1992). Mental rotation, physical rotation, and surface media. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(2), 371.

Kampis, D., Parise, E., Csibra, G., & Kovács, Á. M. (2015). Neural signatures for sustaining object representations attributed to others in preverbal human infants. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1819), 20151683.

Kaufman, J., Csibra, G., & Johnson, M. H. (2005). Oscillatory activity in the infant brain reflects object maintenance. *Proceedings of the National Academy of Sciences*, *102*(42), 15271-15274.

Kessler, K., & Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology*, *1*, 213.

Kessler, K., & Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition*, *114*(1), 72-88.

Kessler, K., & Wang, H. (2012). Spatial perspective taking is an embodied process, but not for everyone in the same way: differences predicted by sex and social skills score. *Spatial Cognition & Computation*, *12*(2-3), 133-158.

Kok, P., Brouwer, G. J., van Gerven, M. A., & de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *Journal of Neuroscience*, *33*(41), 16275-16284.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: evidence from studies of image scanning. *Journal of experimental psychology: Human perception and performance*, *4*(1), 47.

Kosslyn, S. M., Thompson, W. L., & Alpert, N. M. (1997). Neural systems shared by visual imagery and visual perception: A positron emission tomography study. *Neuroimage*, *6*(4), 320-334.

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

Kozhevnikov, M., Motes, M. A., Rasch, B., & Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *20*(3), 397-417.

Kunde, W., Koch, I., & Hoffmann, J. (2004). Anticipated action effects affect the selection, initiation, and execution of actions. Quarterly Journal of Experimental Psychology: Part A: Human Experimental Psychology, 57(1), 87.

Lakens, D., & Albers, C. (2017). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. Journal of Experimental Social Psychology, 74, 187-195.

Lamm, C., Bukowski, H., & Silani, G. (2016). From shared to distinct self–other representations in empathy: evidence from neurotypical function and socio-cognitive disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150083.

Lang, P. J. (1995). The emotion probe: studies of motivation and attention. *American psychologist*, *50*(5), 372.

Langdon, R., & Coltheart, M. (2001). Visual perspective-taking and schizotypy: evidence for a simulation-based account of mentalizing in normal adults. *Cognition*, *82*(1), 1-26.

Langdon, R., Coltheart, M., Ward, P. B., & Catts, S. V. (2001). Visual and cognitive perspective-taking impairments in schizophrenia: A failure of allocentric simulation?. *Cognitive Neuropsychiatry*, *6*(4), 241-269.

Leekam, S., BaronCohen, S., Perrett, D., Milders, M., and Brown, S. (1997). Eye-direction detection: a dissociation between geometric and joint attention skills in autism. *Br. J. Dev. Psychol*. 15, 77–95. doi: 10.1111/j.2044-835X.1997.tb00726.x

Lempers, J. D., Flavell, E. R., & Flavell, J. H. (1977). The development in very young children of tacit knowledge concerning visual perception. *Genetic Psychology Monographs*.

Lewis, M. B. (2001). The Lady's not for turning: Rotation of the Thatcher illusion. *Perception*, *30*(6), 769-774.

MacFie, H. J., Bratchell, N., Greenhoff, K., & Vallis, L. V. (1989). Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of sensory studies*, *4*(2), 129-148.

Mann, D. L., Farrow, D., Shuttleworth, R., Hopwood, M., & MacMahon, C. (2009). The influence of viewing perspective on decision-making and visual search behaviour in an invasive sport. *International Journal of Sport Psychology*, *40*(4), 546-564.

Mattan, B. D., Rotshtein, P., & Quinn, K. A. (2016). Empathy and visual perspective-taking performance. *Cognitive neuroscience*, *7*(1-4), 170-181.

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*(12), 2447-2460.

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural networks*, *9*(8), 1265-1279.

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & psychophysics*, *68*(2), 327-337.

Moors, A., & De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychological bulletin*, *132*(2), 297.

Moreau, D. (2013). Constraining movement alters the recruitment of motor processes in mental rotation. *Experimental brain research*, *224*(3), 447-454.

Morgan, E. J., Freeth, M., & Smith, D. T. (2018). Mental State Attributions Mediate the Gaze Cueing Effect. Vision, 2(1), 11.

Morrison, I., Tipper, S. P., Fenton-Adams, W. L., & Bach, P. (2013). “Feeling” others' painful actions: the sensorimotor integration of pain and action information. *Human bBrain Mmapping*, *34*(8), 1982-1998.

Mosconi, M. W., Mack, P. B., McCarthy, G., & Pelphrey, K. A. (2005). Taking an “intentional stance” on eye-gaze shifts: a functional neuroimaging study of social perception in children. *Neuroimage*, *27*(1), 247-252.

Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: mental imagery as mental emulation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1273-1280.

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current directions in psychological science*, *16*(5), 269-274.

Muto, H., Matsushita, S., & Morikawa, K. (2018). Spatial perspective taking mediated by whole-body motor simulation. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(3), 337.

Neal, D. T., & Chartrand, T. L. (2011). Embodied emotion perception: amplifying and dampening facial feedback modulates emotion perception accuracy. *Social Psychological and Personality Science*, *2*(6), 673-678.

Nielsen, M. K., Slade, L., Levy, J. P., & Holmes, A. (2015). Inclined to see it your way: Do altercentric intrusion effects in visual perspective taking reflect an intrinsically social process?. *The Quarterly Journal of Experimental Psychology*, *68*(10), 1931-1951.

Parr, T., & Friston, K. J. (2017). Working memory, attention, and salience in active inference. *Scientific reports, 7(1)*, 14678.

Parsons, L. M. (1994). Temporal and kinematic properties of motor behavior reflected in mentally simulated action. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(4), 709.

Pearson, A., Ropar, D., & Hamilton, A. F. D. (2013). A review of visual perspective taking in autism spectrum disorder. *Frontiers in human neuroscience*, *7*, 652.

Perner, J., & Aichhorn, M. (2008). Theory of mind, language and the temporoparietal junction mystery. *Trends in cognitive sciences*, *12*(4), 123-126.

Piaget, J. (1929). The child's concept of the world. *Londres, Routldge & Kegan Paul*.

Piaget, J. (1951). Egocentric thought and sociocentric thought. *Sociological studies*, 270-286.

Piaget, J., & Inhelder, B. (1956). *The child's concept of space*. Routledge & Paul.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, *32*(1), 3-25.

Poulin-Dubois, D., Sodian, B., Metz, U., Tilden, J., & Schoeppner, B. (2007). Out of sight is not out of mind: Developmental changes in infants' understanding of visual perception during the second year. *Journal of Cognition and Development*, *8*(4), 401-425.

Poulton, E. C. (1982). Influential companions: Effects of one strategy on another in the within-subjects designs of cognitive psychology. *Psychological Bulletin*, *91*(3), 673.

Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspectives on Psychological Science, 1*, 110–122

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and brain sciences*, *25*(2), 157-182.

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition, 117(2)*, 230-236.

Rajecki, D. W., Ickes, W., Corcoran, C., & Lenerz, K. (1977). Social facilitation of human performance: Mere presence effects. *The Journal of Social Psychology*, *102*(2), 297-310.

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 neuroscience*, *25*(5), 670-684.

Roelfsema, P. R., & de Lange, F. P. (2016). Early visual cortex as a multiscale cognitive blackboard. *Annual review of vision science*, *2*, 131-151.

Samson, D., Apperly, I.A., Braithwaite, J.J., Andrews, B.J., and Bodley Scott, S.E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *J Exp Psychol Hum Percept Perform*. 36(5), 1255.

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing?. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 929.

Schurz, M., Aichhorn, M., Martin, A., & Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. *Frontiers in human neuroscience*, 7, 712.

Schurz, M., Kronbichler, M., Weissengruber, S., Surtees, A., Samson, D., & Perner, J. (2015). Clarifying the role of theory of mind areas during visual perspective taking: Issues of spontaneity and domain-specificity. *NeuroImage*, *117*, 386-396.

Schwarzkopf, S., Schilbach, L., Vogeley, K., & Timmermans, B. (2014). “Making it explicit” makes a difference: Evidence for a dissociation of spontaneous and intentional level 1 perspective taking in high-functioning autism. *Cognition*, *131*(3), 345-354.

Shepard, R. N., and Metzler, J. (1971). Mental rotation of three-dimensional objects. Science. 171(3972), 701-703.

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.

Surtees, A. D. R., Apperly, I. A., & Samson, D. (2013). The use of embodied self-rotation for visual and spatial perspective-taking. Frontiers in Human Neuroscience. 7, 698

Surtees, A., Apperly, I., & Samson, D. (2013). Similarities and differences in visual and spatial perspective-taking processes. *Cognition*, *129*(2), 426-438.

Surtees, A., Apperly, I., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, *150*, 43-52.

Surtees, A., Samson, D., & Apperly, I. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, *148*, 97-105.

Thompson, P. (1980). Margaret Thatcher: a new illusion. *Perception*.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav Brain Sci*. 28, 675–691.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97-136.

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, *110*(1), 124-129.

Vargas, C. D., Olivier, E., Craighero, L., Fadiga, L., Duhamel, J. R., & Sirigu, A. (2004). The Influence of Hand Posture on Corticospinal Excitability during Motor Imagery: A Transcranial Magnetic Stimulation Study. *Cerebral Cortex, 14(11)*, 1200–1206. https://doi.org/10.1093/cercor/bhh080

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... & Meerhoff, F. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic bulletin & review*, *25*(1), 58-76.

Wang, Y., & Hamilton, A. F. D. C. (2012). Social top-down response modulation (STORM): a model of the control of mimicry in social interaction. *Frontiers in human neuroscience*, 6, 153.

Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. *Current Biology*, *29*(5), 874-880.

Ward, E., Ganis, G., McDonough, K. L., & Bach, P. (2019, July 19). Perspective taking as virtual navigation? Perceptual simulation of how others see the world reflects their location in space but not their gaze. Retrieved from osf.io/2cxpr

Warreyn, P., Roeyers, H., Oelbrandt, T., & De Groote, I. (2005). What are you looking at? Joint attention and visual perspective taking in young children with autism spectrum disorder. *Journal of Developmental and Physical Disabilities*, *17*(1), 55-73.

Watson, A. C., Nixon, C. L., Wilson, A., & Capage, L. (1999). Social interaction skills and theory of mind in young children. *Developmental psychology*, *35*(2), 386.

Wertheimer, M. (1912). Experimentelle studien uber das sehen von bewegung. *Zeitschrift fur Psychologie*, *61*.

Williams, J. H., Waiter, G. D., Perra, O., Perrett, D. I., & Whiten, A. (2005). An fMRI study of joint attention experience. *Neuroimage*, *25*(1), 133-140.

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.

Wohlschläger, A., & Wohlschläger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human perception and performance*, *24*(2), 397.

Wolbers, T., Hegarty, M., Büchel, C., & Loomis, J. M. (2008). Spatial updating: how the brain keeps track of changing object locations during observer motion. *Nature neuroscience*, *11*(10), 1223.

Xu, B., Tanaka, J., & Mineault, K. (2012). The head turn cueing effect is sustained at longer SOA's in the presence of an object distractor. *Journal of Vision*, *12*(9), 396-396.

Yirmiya, N., Sigman, M., & Zacks, D. (1994). Perceptual perspective-taking and seriation abilities in high-functioning children with autism. *Development and psychopathology*, *6*(2), 263-272.

Zacks, J., Rypma, B., Gabrieli, J. D. E., Tversky, B., & Glover, G. H. (1999). Imagined transformations of bodies: an fMRI investigation. *Neuropsychologia*, *37*(9), 1029-1040.

Zajonc, R. B. (1965). Social facilitation. *Science*, *149*(3681), 269-274.

Zwickel, J. (2009). Agency attribution and visuospatial perspective taking. *Psychonomic Bulletin & Review*, *16*(6), 1089-1093.

Zwickel, J., & Müller, H. J. (2010). Observing fearful faces leads to visuo-spatial perspective taking. *Cognition*, *117*(1), 101-105.

Zwickel, J., White, S. J., Coniston, D., Senju, A., & Frith, U. (2010). Exploring the building blocks of social cognition: spontaneous agency perception and visual perspective taking in autism. *Social Cognitive and Affective Neuroscience*, *6*(5), 564-571.