Similarities and differences in visual and spatial perspective-taking processes

CrossMark

Andrew Surtees [a,*], Ian Apperly [b], Dana Samson [a]

[a] Psychological Sciences Research Institute, Université catholique de Louvain, IPSY - Place, Cardinal Mercier 10 bte L3.05.01,1348 Louvain-la-Neuve, Belgium
[b] School of Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

ARTICLE INFO

ABSTRACT

Processes for perspective-taking can be differentiated on whether or not they require us to mentally rotate ourselves into the position of the other person (Michelon & Zacks, 2006). Until now, only two perspective-taking tasks have been differentiated in this way, showing that judging whether something is to someone's left or right does require mental rotation, but judging if someone can see something or not does not. These tasks differ firstly on whether the *content* of the perspective is visual or spatial and secondly on whether the *type* of the judgement is early-developing (level-1 type) or later-developing (level-2 type). Across two experiments, we tested which of these factors was likely to be most important by using four different perspective-taking tasks which crossed orthogonally the *content* of judgement (visual vs. spatial) and the *type* of judgement (level-1 type vs. level-2 type). We found that the level-2 type judgements, of how something looks to someone else and whether it is to their left or right, required egocentric mental rotation. On the other hand, level-1 type judgements, of whether something was in front of or behind someone and of whether someone could see something or not, did not involve mental rotation. We suggest from this that the initial processing strategies employed for perspective-taking are largely independent of whether judgements are visual or spatial in nature. Furthermore, early developing abilities have features that make mental rotation unnecessary.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

For everyday thought and action we often require an accurate representation of our environment and also the ability to predict how that environment might change. Of particular importance within that environment are other people. The requirement to be able to cooperate and compete with others has helped to shape human cognition (Tomasello, 2008). Successful cooperation and competition are often thought to require accurate perspective-taking (Keysar, Lin, & Barr, 2003), this is our ability to non-ego-centrically represent some aspect of the world as it is to another.

Perspective-taking has been taken to mean a variety of things in different literatures. Commonly it is thought to imply an ability to put ourselves into the shoes of another. It is, however, an open question for cognitive scientists as to whether or not the computational processes we use actually involve mentally positioning ourselves where the other person is (Michelon & Zacks, 2006). For any perspective-taking task, three basic components are involved: a perspective-taker (Self), who judges another's perspective; a target (Other) perspective that can be judged (this is most commonly another person, but can in fact be a directional object or an imagined self-perspective); and an object or circumstance (Object) upon which the perspective is taken. Seen this way, perspective-taking involves two levels of relation (and two levels of representation,

* Corresponding author. Address: Faculté de psychologie et des sciences de l'éducation, Institut de recherche en sciences psychologiques, Université catholique de Louvain, IPSY – Place, Cardinal Mercier 10 bte L3.05.01, 1348 Louvain-la-Neuve, Belgium. Tel.: +32 (0)10 47 9062.
E-mail address: adrsurtees@gmail.com (A. Surtees).

Rakoczy, 2012). Firstly, regarding the relationship between the Self and the Other: How does the Self represent the Other's perspective? Secondly, regarding the relationship between the Other and the Object: What relationship between the Other and the Object is represented? It is worth noting that this, rather minimal, definition does not require that the Other actually perceives something. Researchers in the area of Theory of Mind (ToM, the ability to represent, understand or act sensitively based on the mental states of others) have put a premium on understanding others' minds (Premack & Woodruff, 1979). It is, however, very difficult in practice to assure this, with non-mental short cuts often providing consistently correct responses (Perner & Ruffman, 2005). Adding complexity to Other–Object relations can make mentalistic interpretations more parsimonious, but here we focus on relatively simple relations.

Distinct types of Other–Object relations are involved in visual perspective-taking and spatial perspective-taking. Visual perspective-taking involves a relationship based on if and how an object is seen by the other. Spatial perspective-taking involves a relationship based on the relative spatial locations of the other and the object. The literature on perspective-taking often treats such abilities as separate, but there are clear commonalities between such judgements (Surtees, Noordzij, & Apperly, 2012). They both require the ability to identify the position and orientation of someone else and the ability to understand that their perspective may be different from our own: It may be in front of her, but behind me; I may be able to see it, even if she cannot. Often, but not always, the spatial relation "in front" or "behind" can act as a successful proxy for whether something is seen by someone else, and adults have been found to be particularly sensitive to the fronts of other people (Surtees, Noordzij, et al., 2012) and also to whether objects are seen by them (Samson, Apperly, Braithwaite, Andrews, & Scott (2010)). In this paper, we directly compare visual and spatial perspective-taking with the goal of identifying commonalities and differences in processing strategies.

1.1. Visual Perspective-taking

Successful cooperation with and competition against others can often depend on judging what objects are seen by someone and how they see them. For a chimpanzee the ability to judge if food can be seen by another chimpanzee can mean the difference between an appropriate and an inappropriate behavior in the presence of a dominant other (Brauer, Call, & Tomasello, 2007). Visual perspective-taking has received a large amount of attention from developmental and comparative psychologists (Brauer et al., 2007; Emery & Clayton, 2004; Flavell, Everett, Croft, & Flavell, 1981; Piaget & Inhelder, 1956). This is, at least in part, because it is often considered to form the starting point of the broader ability of ToM (Flavell, 1988).

Visual perspective-taking is not thought to be a unitary ability (Flavell et al., 1981). Flavell and colleagues in the 1970s and 1980s showed a distinct developmental trajectory (Flavell et al., 1981; Masangkay et al., 1974). Specifically, they found that 3-year-old children passed

tasks requiring "level-1" perspective-taking, the ability to judge that someone else might not see an object that you yourself can see (Flavell et al., 1981; Masangkay et al., 1974). This ability has since been shown in explicit responding by the age of 24-months (Moll & Tomasello, 2006) and in implicit monitoring of gaze direction in infancy (Song & Baillargeon, 2008; Sodian, Thoermer & Metz, 2007). It has also been demonstrated by a number of non-human animals, including other primates (Brauer et al., 2007; Santos et al., 2007), more distant mammalian relations such as dogs (Hare, Brown, Williamson, & Tomasello, 2002) and goats (Kaminski, Riedel, Call, & Tomasello, 2005) and even certain species of bird (Emery & Clayton, 2004). This has suggested that level-1 perspective-taking requires only limited cognitive control and may be part of a low-level system for mind-reading (Apperly & Butterfill, 2009). Samson et al. (2010) found evidence consistent with this hypothesis in that adults seemed to automatically calculate the level-1 visual perspectives of others. Further studies have suggested that this automaticity is also present in the responses of 6–11 year old children (Surtees & Apperly, 2012) and is relatively undemanding on general cognitive resources (Qureshi et al., 2010). Interestingly, though, this ability is not sufficient for making explicit responses. When adults make direct responses about other's level-1 perspectives, they suffer an extra cost if they have to ignore their own conflicting perspective (Samson et al., 2010). Such evidence is also consistent with findings from studies of referential communication (Keysar, Barr, Balin, & Brauner, 2000; Keysar et al., 2003) showing that adults require effort to avoid responding as if another person sees all the items in their own view.

There is, of course, more to visual perspective-taking than judging if someone can see a given object or not. Flavell et al. defined level-2 perspective-taking as understanding that "an object simultaneously visible to both the self and the other person may nonetheless give rise to different visual impressions or experiences in the two if their viewing circumstances differ" (Flavell et al., 1981, p. 1). The ability to do so develops later in childhood, at around four years of age (Flavell et al., 1981; Masangkay et al., 1974). Level-2 perspective-taking abilities are yet to be found in infants or non-human animals (Apperly & Butterfill, 2009) and there is no evidence of automatic level-2 perspective-taking (Surtees, Apperly & Samson, submitted for publication; Surtees, Butterfill, & Apperly, 2012).

1.2. Spatial perspective-taking

Spatial perspective-taking is the ability to understand where something is located relative to someone else. This kind of perspective-taking has received far less interest from researchers in social perspective-taking. This is, at least in part, likely to be because spatial perspectives need-not necessarily imply mental content: for me to know that something is to your left is in no way dependent on you representing it being as such. In addition, there is not necessarily more than an arbitrary distinction between adopting a spatial perspective of another person and any object with a front (Surtees, Noordzij, et al., 2012). For example a meaningful distinction can be made about the location of a

ball being in front of a chair in the same way as it can about the location of a ball being in front of a person (Surtees, Noordzij, et al., 2012). Related to spatial perspectives are frames of reference, which allow us to encode spatial information relative to ourselves (the relative frame of reference), relative to another person/thing (the intrinsic frame of reference) or to some non-varying element in the environment (the absolute frame of reference, Levinson, 1996, 2003). There is evidence that adults automatically activate multiple frames of reference when making linguistic decisions (Carlson-Radvansky & Irwin, 1993; Carlson-Radvansky & Jiang, 1998). On the other hand, there is evidence of cross-cultural differences in the conditions under which different frames of reference are adopted (Bowerman & Choi, 2003; Levinson, 1996). There is some evidence that spatial perspectives involving people are treated somewhat differently to those solely involving objects, with children and adults showing different frame of reference use for a person and an object (Surtees, Noordzij, et al., 2012). There is also evidence that people spontaneously take the spatial perspectives of others even when it is not necessary (Tversky & Hard, 2009) and evidence of people mentally adopting another's position in tasks involving judging simple actions (Pickering, McLean, & Gambi, 2012).

Like for Visual perspective-taking, there is some evidence that the ability to make judgements about spatial perspectives does not undergo a uniform developmental pattern. Specifically there is evidence of early sensitivity to notions of front and back (Bialystok & Codd, 1987; Cox, 1981; Harris & Strommen, 1972). Children consistently use the words "in front" and "behind" by the age of 3–4 years (Bialystok & Codd, 1987; Cox, 1981; Harris & Strommen, 1972), but are not able to show consistent use of "to the left of" or "to the right of" until much later (Hands 1972; Harris & Strommen, 1972). Whilst it is clear that these distinctions do not map directly onto Flavell and colleague's (Flavell et al., 1981) definitions of level-1 and level-2 visual perspective-taking, we will consider judgements of in front or behind as "level-1 type" and judgements of to the left or to the right as "level-2 type". This is based on the developmental delay between the two abilities and for ease in describing similarities and differences across visual and spatial dimensions. This is, of course, somewhat arbitrary and is aimed at providing clarity of comparison, rather than classifying spatial judgements on the basis of existing definitions of visual perspective-taking.

1.3. Processes for visual and spatial perspective-taking

Within the two separate strands of research there has been some important investigation of the processes involved when we adopt another's perspectives. For level-1
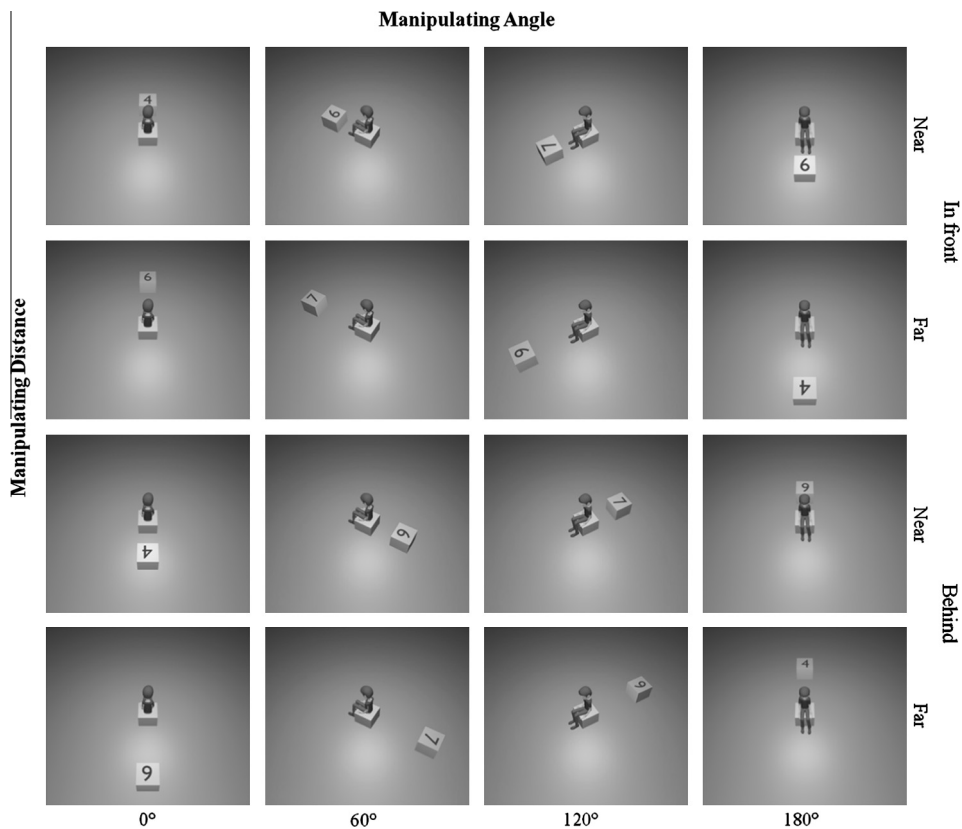

Fig. 1. Subset of stimuli used in all four experiments. Note that for the Spatial II condition, the block with the numeral on it was off-set at a 45° angle from the avatar.

visual perspective-taking, there is evidence that inhibition is needed to select between self and other perspectives (Qureshi et al., 2010). There is similar evidence that inhibition is required to select between different frames of reference (Carlson-Radvansky & Jiang, 1998). As well as looking for whether inhibition is required to avoid our own point of view, researchers have begun to be interested in the strategies involved in different types of perspective-taking (Kessler & Thomson, 2010; Michelon & Zacks, 2006; Zacks, Mires, Tversky, & Hazeltine, 2000). In a seminal study aimed to identify the strategies used in perspective-taking, Michelon and Zacks (2006) had participants complete perspective-taking in a task where they systematically manipulated two factors: the angular disparity between the participant and the target other (Angle herein) and the distance between the other person and the object about which the perspective was taken (Distance herein). By examining two different kinds of perspective-taking task, Michelon and Zacks identified two different performance patterns. When participants were asked to judge if a doll could see a given object or not (a level-1 visual perspective-taking task), their performance varied in relation to Distance, but did not vary linearly in relation to Angle. Michelon and Zacks concluded that participants were tracing a line from the doll's eyes. When participants were asked to judge if an object was to the left or the right of the doll, performance varied linearly in relationship to Angle, but less so for Distance. Michelon and Zacks concluded that, for this kind of perspective-taking, participants were mentally rotating themselves to the doll's position. Interestingly, in addition to this, Kessler and Thomson (2010) showed that the participants' own body orientation affected these left/right judgements suggesting that this process was dependent on the current embodied representation of one's own egocentric position. Taken together, these findings are highly informative about processes for visuo-spatial perspective-taking. Whilst Michelon & Zacks entitled their paper "two kinds of visual perspective-taking", it is clear that judging "to the left of" or "to the right of" is not prima facie a *visual* perspective judgement at all. Left/right judgements can equally well be made in relation to a blindfolded person or a fronted object such as a chair, neither of which has any visual perspective on the scene.

In summary, visual and spatial perspective-taking literatures have, until recently, employed separate approaches. Two exciting studies have paved the way for our understanding of the computational processes involved in visuo-spatial perspective-taking (Kessler & Thomson, 2010; Michelon & Zacks, 2006), but cut across existing distinctions emanating from developmental psychology. The findings show clearly that one strategy (line of sight tracing) is used for an early developing form of visual perspective-taking; that which developmental psychologists refer to as level-1 visual perspective-taking. They also show clearly that a different strategy (embodied egocentric transformation) is used for a later developing form of spatial perspective-taking. However, these findings point to two very different conclusions. A Content specificity hypothesis would suggest that line of sight tracing might be used for all visual perspective judgements, whilst egocentric mental rotation might be used for all spatial

perspective judgements. In contrast, a Type specificity hypothesis would suggest that line of sight/orientation tracing might be used for early developing abilities (be they visual or spatial) whilst egocentric mental rotation might be responsible for later developing abilities (be they visual or spatial). The current studies sought to test these hypotheses.

In two experiments, we tested the relative contribution of Angle (the angular displacement between the perspective-taker and an avatar whose perspective they judged) and Distance (the linear distance between the avatar and an object) on four types of perspective-taking task. The requirements of these tasks varied on two dimensions. Firstly, whether the task was a visual or a spatial perspective-taking task. Secondly, whether the ability to pass the task develops earlier or later in childhood, considered as level-1 type and level-2 type respectively. For the "Visual I" condition, participants judged if a numeral could be seen (a level-1 perspective-taking task- Flavell et al., 1981). For the Visual II condition, they judged what number the numeral represented (a level-2 visual perspective-taking task- Flavell et al., 1981). For the Spatial I condition, they judged whether the numeral was in front of or behind the avatar. For the Spatial II condition, they judged whether the numeral was to the left of or to the right of the avatar. We measured Response Times as a proxy for how difficult each judgement was and how this depended on Angle and Distance. We further measured error rates to test for trade-offs between speed and accuracy. Using this design, we can made distinct predictions based on our two hypotheses.

1.4. Type specificity hypothesis

According to this hypothesis, level-2 type judgements (visual and spatial) would require egocentric mental rotation while level-1 type judgements (visual and spatial) would not. This hypothesis would predict an interaction between Type of perspective-taking (level-1 type/ level-2 type) and Angular disparity. For level-2 type, there should be a significant and linear relationship between Angle and difficulty, such that a greater angle corresponds to a greater difficulty. For level-1 type, there should be no relationship between Angle and difficulty. The type specificity hypothesis makes no strong claims about an effect of distance.

1.5. Content specificity hypothesis

According to this hypothesis, spatial perspective judgements (level-1 type and level-2 type) would require egocentric mental rotation while visual perspective judgements (level-1 and level-2) would not. This hypothesis would predict an interaction between Content of perspective-taking (Visual/Spatial) and Angular disparity. For spatial perspective-taking, there should be a significant and linear relationship between Angle and difficulty, such that a greater angle corresponds to a greater difficulty. For Visual perspective-taking, there should be no relationship between Angle and difficulty. For the content hypothesis to be correct does not require Distance to affect

perspective-taking differentially across our conditions. However, an extension to the content hypothesis would be to argue that all visual perspective judgements are more sensitive to Distance than all Spatial judgements (as has been shown in a subset of these judgements, Michelon & Zacks, 2006).

2. Experiment 1

2.1. Method

2.1.1. Participants

Participants were 64 undergraduate students (15 male) from the University catholique de Louvain, Belgium. They all participated in the study in exchange of course credit or a small honorarium of 5 Euros. Participants had an average age of 20.63 years (Range 18–24).

2.1.2. Stimuli

Picture stimuli all presented a human avatar seated within a room (see Fig. 1). The stimuli were created using Blender (www.blender.org). Alongside the avatar was a cube and on the cube a single numeral (4, 6, 7 or 9). The position of the avatar and the cube/numeral were varied on two orthogonal dimensions. The angle at which the avatar was facing differed from the angle at which the virtual camera was positioned by one of 4 magnitudes: 0°, 60°, 120°, 180°. For 60° and 120° pictures this included both clockwise and anticlockwise variants. The distance at which the numeral was placed from the avatar took one of 2 magnitudes: Near and Far. The Far condition was placed at a distance that was twice as far within the virtual world as the Near condition. In three of the conditions (Spatial I, Visual I and Visual II), stimuli presented were taken from the same set. In these cases, the numeral appeared either directly in front or directly behind the avatar. For Spatial II, the location of the numeral was displaced 45° from the angle of the avatar, so that meaningful judgments could be made as to whether it was to the left or the right of the avatar.

2.2. Procedure

Participants were randomly divided into one of four groups Visual I, Visual II, Spatial I or Spatial II. Participants were instructed that they were to perform a perspective-taking task in which they were going to make judgments based on picture stimuli. After brief instructions, in all conditions they completed 16 practice trials followed by 256 experimental trials divided into four blocks, all using the same basic procedure (see Fig. 2). Participants saw a written cue before an image. On presentation of the image, participants' task was to respond, by pressing the left or right mouse button, as to whether the cue matched the picture.

For Visual I, this meant responding as to whether the avatar could or could not see the number. For this they were cued by either "vu" (seen) or "pas vu" (not seen). Trials were divided equally on four experimental factors: Whether the picture and the cue matched (Match, Mismatch), whether the avatar saw or did not see the numeral

(Yes, No), the distance between the avatar and the numeral (Near, Far) and the angular displacement from the avatar (0°, 60°, 120°, 180°). Previous studies (Michelon & Zacks, 2006) have found equivalence between clockwise and anti-clockwise rotations, therefore for our 60° and 120° conditions, there were an equal number of trials in which the displacement was clockwise (see Fig. 1) and anticlockwise. For Spatial I, the exact same picture cues were used. The only difference here was that the word cues "devant" (in front) and "derrière" (behind) replaced seen and unseen respectively. In Visual II, for the 128 match trials, these matched those trials in the Visual I and Spatial I conditions where the numeral was seen by the avatar. So again we had an equal number of trials for each angle and distance. Non-match trials were of two-distinct types to ensure two vital elements of the task. In half of non-match trials, the number cue did not match the number on screen; this ensured that participants truly judged how the number looked from the avatar's perspective. In the other half, the number was correct, but was located behind the avatar. This was to ensure that participants had to pay attention to the avatar himself, rather than just the numeral. For Spatial II, match trials were the same as those for Visual II except for the fact that the number was displaced, equally often, to the left or the right of the avatar. Here, in all non-match trials, the number was in front of him, but not to the side indicated by the cue. For all conditions, we also varied evenly the non-experimental factor of the numeral that was presented.

In summary, for Visual I and Spatial I, experimental conditions were identical, other than the probe used and the instructions given. For example in Visual I a participant would respond that the avatar could indeed see the numeral, whilst an equivalent Spatial I trial had the participant respond that the numeral was indeed in front of the avatar. The match trials (those included in the final analysis) differed from this for the Visual II and Spatial II conditions, in that the numeral was always in front of the avatar (for Visual II, this is a pre-requisite for judging how the numeral is seen). Match trials in Visual II and Spatial II only differed on two dimensions: Firstly the cue given either related to a direction or a number and secondly for Spatial II the numeral was displaced to the left or to the right to allow for the appropriate judgement. Conditions thus differed only on aspects that were central to the specific perspective-taking ability itself.

3. Results

Only responses in which the cue and the picture matched were analysed, i.e. those trials in which the participant pressed a button to indicate "yes". Three participants who failed to perform above chance were excluded; these were all in the condition Spatial II and all had a pattern of performance consistent with ignoring the cue and clicking the mouse button directly related to the side on which the numeral appeared. Outliers were excluded from the analysis of Response Times on the basis of being more than 2.5 standard deviations away from the
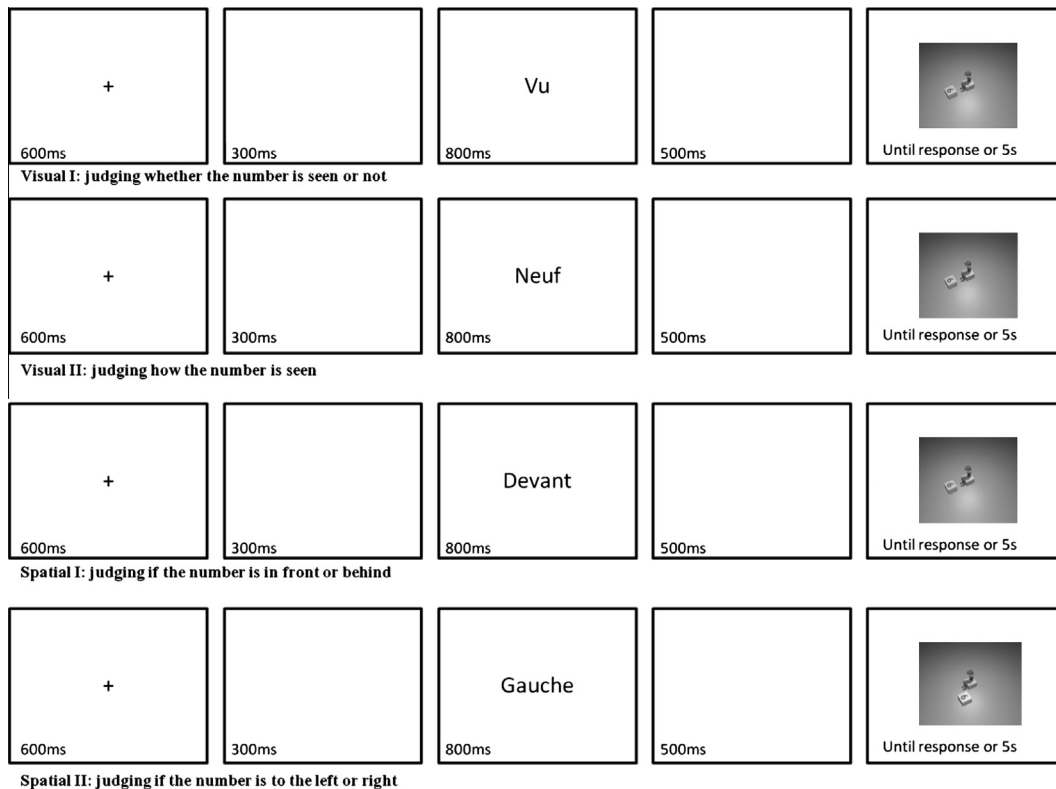
Fig. 2. Procedures for experiments.

mean response time (2.81–3.65% across the four conditions), as were incorrect responses.

The *type specificity hypothesis* predicts an interaction between Type and Angle, representing a significant linear effect of Angle, but only for Level-2 type perspective-taking. The *content specificity hypothesis* predicts an interaction between Content and Angle, representing a significant linear effect of Angle, but only for Spatial perspective-taking.

A $4 \times 2 \times 2 \times 2$[1] ANOVA on response times with Angle (0°, 60°, 120°, 180°) and Distance (Near, Far) as within subjects factors and Content (Visual, Spatial) and Type (Level-1 type, Level-2 type) as between subjects factors revealed a main effect of Distance, $F(1,60) = 7.21$, $p = .009$, $\eta p^2 = .107$, with shorter avatar-object distances processed more quickly. There was also a main effect of Angle, $F(3,180) = 22.00$, $p < .001$, $\eta p^2 = .268$, which represented a linear trend, $F(1,60) = 53.87$, $p < .001$, $\eta p^2 = .473$, such that larger angles were responded to more slowly (See Fig. 3). The between subjects factors, Content and Type also affected response times. Visual judgements were processed more quickly than Spatial judgements, $F(1,60) = 8.96$, $p = .004$, $\eta p^2 = .130$. Level-1 type judgements were processed more quickly than Level-2 type, $F(1,60) = 7.70$, $p = .007$, $\eta p^2 = .110$.

Central to our hypotheses were the interactions between Angle and Content and Angle and Type. In favour of the *type-specificity hypothesis*, there was a significant

interaction between Angle and Type, $F(3,180) = 17.93$, $p < .001$, $\eta p^2 = .230$. For Level-1 type, there was no main effect of Angle, $F(3,90) = .340$, $p = .797$, $\eta p^2 = .034$. For Level-2 type, there was a significant, $F(3,90) = 17.73$, $p < .001$, $\eta p^2 = .647$, and linear, $F(1,60) = 61.10$, $p < .001$, $\eta p^2 = .671$, effect of Angle. Contrary to the *Content specificity hypothesis*, the interaction between Angle and Content, $F(3,180) = 1.19$, $p = .314$, $\eta p^2 = .019$, was not significant. Effects of Angle were independent of whether the judgement was visual or spatial.

There was a significant Content by Type interaction, $F(1,60) = 38.04$, $p < .001$, $\eta p^2 = .390$, resulting from participants performing more quickly for the Spatial variant of Level-1 type perspective-taking, $t(15) = 4.13$, $p < .001$, and conversely performing better for the Visual variant of Level-2 type perspective-taking, $t(15) = 4.96$, $p < .001$. No further interactions were significant, $Fs \leqslant 1.12$, $ps \geqslant .295$.

Error rates across conditions were generally low and did not contradict the findings from response times. Full details of the analysis of errors are included in Appendix A.

The omnibus analysis of Response Times clearly suggests that judging how someone sees something and judging whether something is to someone's left or right require a rotation to align the position of the participant and the avatar, for response times increased with angular disparity. On the other hand, judging if someone sees something and judging whether it is in front of them do not require rotation, for the response times do not increase with angular disparity. The effect of Distance was somewhat more

---

difficult to interpret. There were no interactions between Distance and Type and Distance and Content. This suggests either that line of sight is calculated in all conditions, or that Distance affected speed and accuracy for a reason other than the recruitment of a line of sight process (see discussion).

Alongside our main investigations of processing strategies employed, we also identified gross differences between the four conditions. For level-1 type perspective-taking, performance was better in the spatial condition, for level-2 type perspective-taking, performance was better in the visual condition. That level-1 visual perspective-taking was found to be particularly difficult was surprising, as authors have suggested the basic calculation of such perspectives (though not explicit judgements about them) to be automatic (Samson et al., 2010).

4. Experiment 2

In Experiment 1, participants responded as to whether a preceding cue correctly described the picture presented. This has the advantage of meaning that participants basic responses, "yes" or "no", were matched across the four conditions. Furthermore, it meant that only a single response type, associated with "yes" was actually analysed for each participant. On the other hand this also has two disadvantages. Firstly, it presents a pre-conceived idea of the stimulus that is to follow, meaning participants may try and match the picture to this pre-conceived representation. This leads to the possible interpretation that rotational effects on level-2 visual perspective-taking may be the result of a discrepancy between the expected and actual picture. Secondly, we considered it possible that level-1 visual perspective-taking had been rendered artificially difficult by having to process double negations, such as deciding that the stimulus is not "not seen" by the avatar. To address this, in Experiment 2, participants pressed one of two keys to identify the avatar's perspective. Also in Experiment 2 we varied type of perspective-taking within participant. This provides a stronger test of the independence of the processing features of the two types. Firstly it investigates whether the very same participants would show differing performance patterns for different types. Secondly it tests whether our effects are robust to the context of previously having completed the other type of perspective-taking.

4.1. Method

4.1.1. Participants
Participants were 32 undergraduate students (12 male) from the University catholique de Louvain, Belgium. They all participated in the study in exchange for a small honorarium of 8 Euros. Participants had an average age of 21.70 years (Range 18–25).

4.1.2. Stimuli and procedure
Stimuli used were a subset of those used in Experiment 1, those including the numbers 6 and 9, but not the numbers 4 and 7. Participants were pseudo-randomly divided into two groups Visual or Spatial: For experiment 2, we ensured that male and female participants were equally divided across groups (as Kessler & Wang, 2012, previously found differences between genders in processing style). Participants were instructed that they were to perform a perspective-taking task in which they were going to make judgments based on picture stimuli. Participants completed level-1 type or level-2 type perspective-taking in a counterbalanced order. After brief instructions, for each level, they completed 16 practice trials, followed by 128 experimental trials. On each trial, they saw a fixation cross followed by a picture. For each picture, one of two responses was possible: Level-1 Visual (seen, unseen),Level-2 Visual (six, nine), Level-1 type Spatial (in front, behind), Level-2 type Spatial (left of, right of). Note that, unlike for Experiment 1, here we were able to analyse all trials, rather than just "yes" responses. Participants responded by pressing the up or down arrow key to each trial. The key used for each response was counterbalanced across participants.

5. Results

A 4x2x2x2 ANOVA on Response Times was completed with Angle (0°, 60°, 120°, 180°), Distance (Near, Far) and Type (Level-1 type, Level-2 type) as within subjects factors and Content (Visual, Spatial) as a between subjects factor. There was a main effect of Angle, $F(3,90) = 57.02$, $p < .001$, $\eta p^2 = .655$, this represented a linear trend, $F(1,30) = 99.69$, $p < .001$, $\eta p^2 = .769$, such that larger angles were responded to more slowly. As in Experiment 1, Level-2 type judgements were computed more slowly, $F(1,30) = 59.32$, $p < .001$, $\eta p^2 = .664$, but here there was only a trend for an effect of Distance, $F(1,30) = 3.02$, $p = .093$, $\eta p^2 = .093$. As in Experiment 1, visual judgements were completed more quickly than spatial judgements, $F(1,30) = 57.02$, $p < .001$, $\eta p^2 = .655$.

Central to our hypotheses were the interactions between Angle and Content and Angle and Type (See Fig. 4). In favour of the type-specificity hypothesis, there was a significant interaction between Angle and Type, $F(3,90) = 47.90$, $p < .001$, $\eta p^2 = .615$. For Level-1 type, there was a main effect of Angle $F(3,90) = 6.22$, $p = .001$, $\eta p^2 = .172$, but this was non-linear, $F(1,30) = 2.16$, $p = .151$, $\eta p^2 = .067$, with participants performing best at 60° and 120°. For Level-2 type, there was a significant $F(3,90) = 60.35$, $p < .001$, $\eta p^2 = .668$, and linear, $F(1,30) = 114.07$, $p < .001$, $\eta p^2 = .792$, effect of Angle. Contrary to the Content specificity hypothesis, the interaction between Content and Angle was not significant, $F(3,90) = 2.15$, $p = .121$, $\eta p^2 = .067$. Effects of Angle were independent of whether the judgement was visual or spatial.

Again, as in Experiment 1, there was a significant Content by Type interaction, $F(1,30) = 21.87$, $p < .001$, $\eta p^2 = .422$, resulting from participants performing more quickly for the Spatial variant of Level-1 type perspective-taking, $t(15) = 2.35$, $p = .027$, and conversely performing better for the Visual variant of Level-2 type
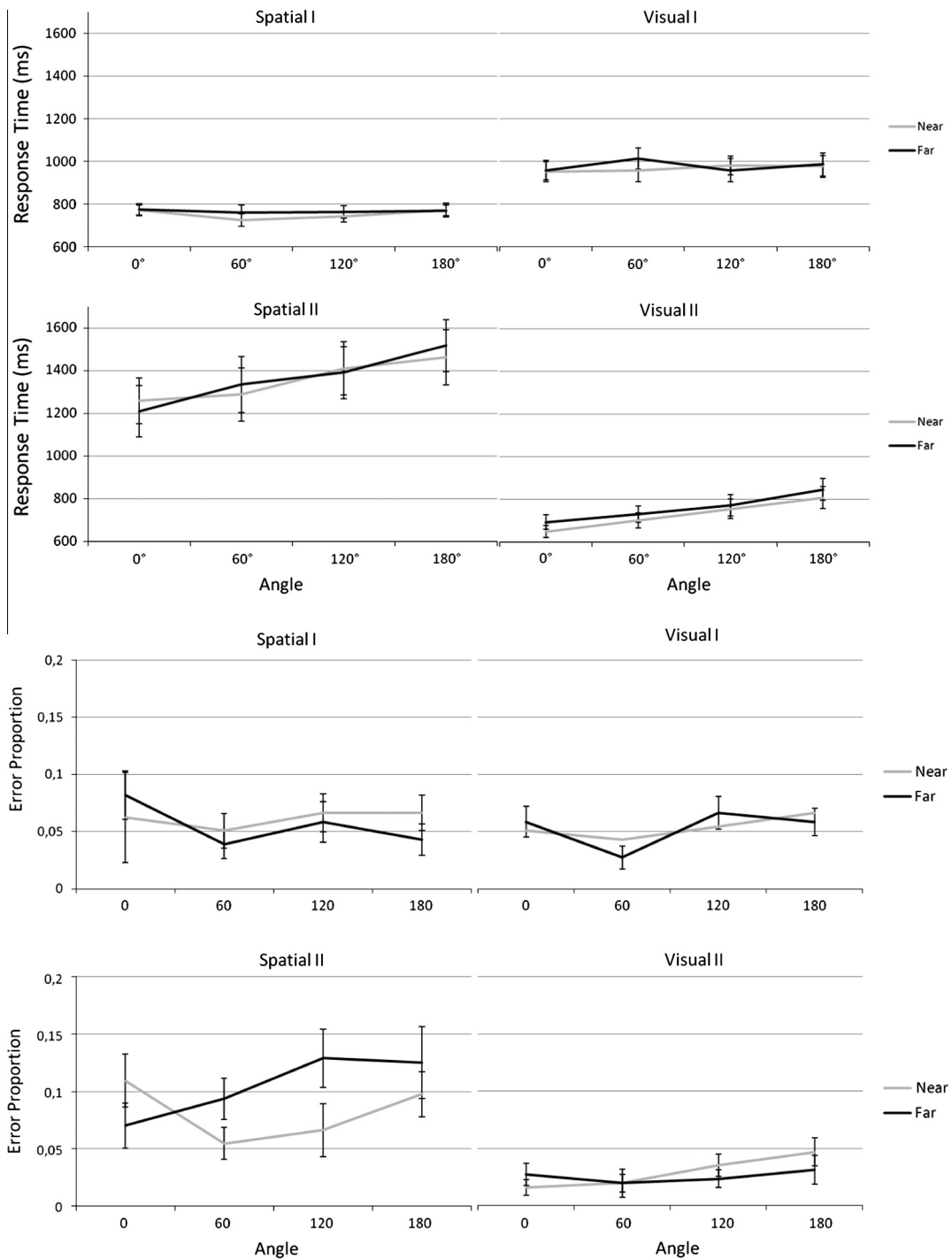
Fig. 3. Response times and error proportions for each of the four conditions of Experiment 1. Error bars represent the standard error of the mean.

perspective-taking, $t(15) = 2.90$, $p = .009$. No other interactions were significant, Fs $\leqslant 1.53$, ps $\geqslant .215$.

Error rates across conditions were generally low and did not contradict the findings from response times. Full details of the analysis of errors are included in Appendix A.

Results matched the central findings of Experiment 1. We found the influence of Angle to be centrally linked to

Type of perspective-taking. There was a significant interaction between Type and Angle, demonstrating that level-2 type judgements were linearly affected by the angular disparity between the self and other, whilst level-1 type judgements were not. This suggests that the results were not the effect of matching pictures to pre-defined cues. Similarly, we again found participants to be slower in the
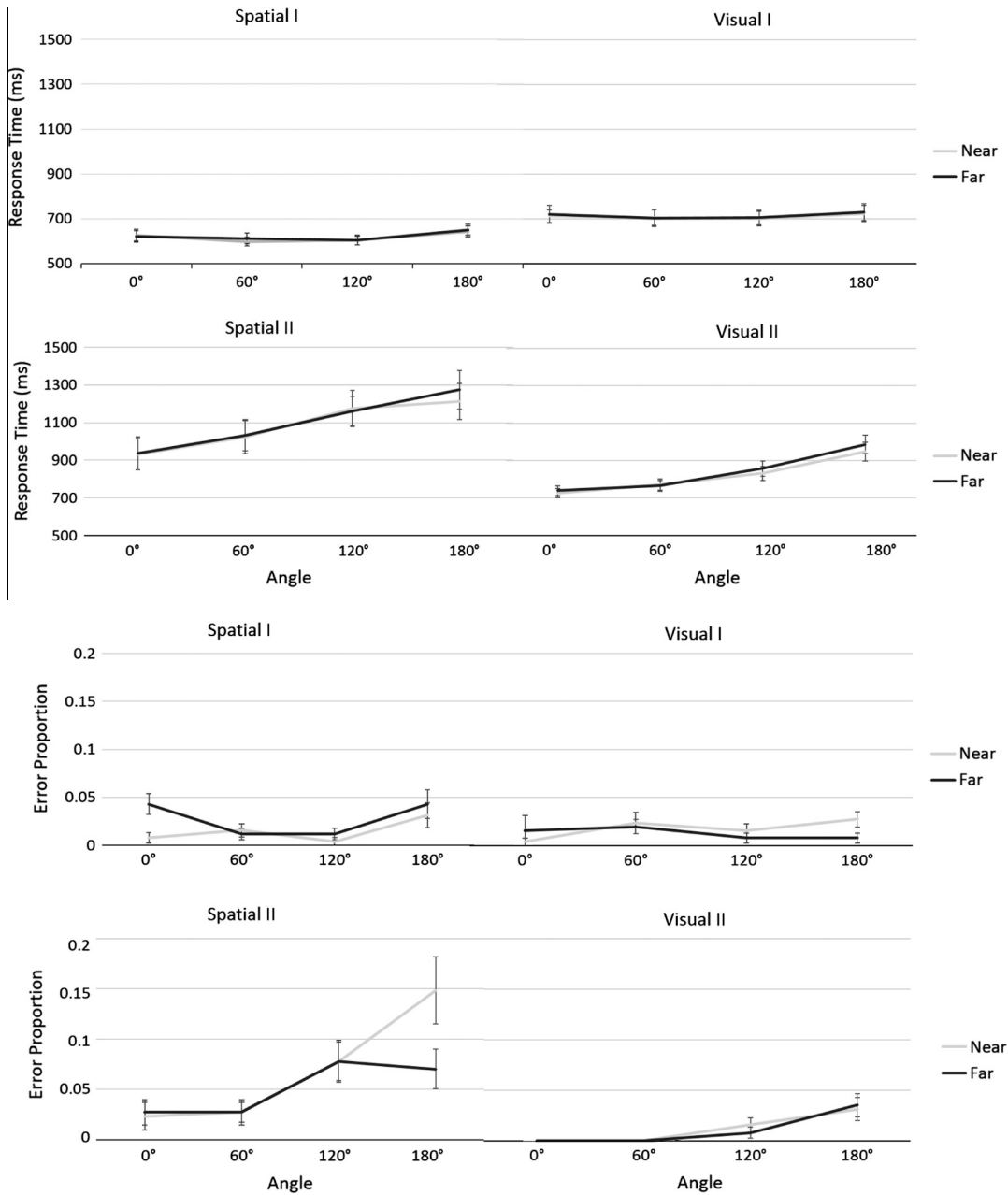
Fig. 4. Response times and error proportions for each of the four conditions of Experiment 2. Error bars represent the standard error of the mean.

visual variant of level-1 type perspective-taking than the spatial variant. In Experiment 2, there was no evidence for level-2 visual perspective-taking being easier than level-1 visual perspectives-taking. This suggests that that surprising finding from Experiment 1 was likely to be the result of the specific experimental set-up. Experiment 2 also ruled out that any effects were driven by gender.

6. Discussion

In two experiments, adult participants made judgements about the perspective of an avatar in one of four conditions. Analysis of their response times and accuracy revealed two distinct processing strategies. One of these strategies involved mentally rotating their own position to align it with that of the avatar. When judging what numeral the avatar could see and when judging whether it was to his left or to his right, the difficulty of the task increased as the angular disparity between the participant and the avatar increased. This was not the case for the other two kinds of perspective-taking we tested. When participants judged if the avatar could see the numeral and when they judged if it was in front of or behind him, angle did not impact on judgements. These findings fit

most closely with our *type specificity hypothesis*, in that it was the type of perspective-taking (level-1 type/level-2 type) which was shown to relate to perspective-taking strategy, rather than whether the judgement was of a visual, or a spatial perspective.

6.1. Visual perspective-taking

For over 50 years, researchers in developmental psychology have considered children's ability to understand the visual perceptions of others to be a landmark ability (Flavell et al., 1981; Piaget & Inhelder, 1956; Masangkay et al., 1974). Here we provide further evidence as to the computational processes involved in mature visual perspective-taking. We support the findings of Michelon and Zacks (2006) and Kessler and Thomson (2010) that judging whether or not someone sees a given object does not require us to mentally rotate to their position. It was not systematically more difficult to judge if someone sees an object if their view on the world is displaced from ours by a greater angle. That is not to say that one cannot, in principle, judge level-1 visual perspectives in this way, but rather this is not what participants did spontaneously on the current tasks. Here, critically, we undertook an equivalent investigation of the processes involved when we calculate *how* another person sees an object. For these level-2 judgements, we provide the first evidence that when we calculate how someone sees something we may use a process of mental rotation. Further research is needed to be certain that this rotational process involves mentally transforming *our own* egocentric position, but previous research suggesting this to be the case for left/right judgements (Kessler & Rutherford, 2010; Kessler & Thomson, 2010) would, at the very least, suggest this as a good working hypothesis.

6.2. Spatial perspective-taking

As for visual perspective-taking, we found evidence that spatial perspective judgements can be made by one of two computational processes. Different processes were employed, depending on the axes about which the judgement was made. Again, in our directly comparable condition, we support the findings of Michelon and Zacks (2006), Kessler and Thomson (2010), and Kozhevnikov and Hegarty (2001). Specifically, when we have to judge whether something is located to someone's left or right, we find evidence that this requires an egocentric mental rotation. On the other hand, when we make what seem to be similar judgements, of whether something is in front or behind someone, we do not do so. One possible explanation of this is that we *only* mentally rotate ourselves to the target's position when discriminating the relevant direction is difficult. In the case of judging whether something is in front or behind someone this discrimination should be relatively easy as we have many cues, most notably someone's face, to distinguish their front. In the case of judging whether something is to the left or the right of someone this is often more difficult as these cues do not exist in the same way because the human body is typically symmetrical on the left–right dimension. In these difficult discrimination

tasks, we adopt the strategy of rotating our own position and using our knowledge of our own body schema.

6.3. Comparing Visual and Spatial perspective-taking processes

Visual and spatial perspective-taking have often been considered quite separately. The main aim of the present work was to examine commonalities and differences between the two. Our findings from both visual and spatial perspective-taking run contrary to the *content hypothesis.* For neither visual nor spatial content was there a consistent process used across level-1 and level-2 types of judgement. Instead, consistent with the *type hypothesis*, we found similarity of processing across visual and spatial judgements. Judgements of whether an object was in front or behind of someone or whether the person saw the object or not did not vary systematically with the angle between the perspective-taker and their target. This suggests that for these judgements people do not mentally rotate themselves to the position of their target. This leads to the obvious question of whether these processes are in fact the same. There are reasons to be cautious in concluding this. Firstly, from our data alone, there was evidence that the visual perspective judgements were more difficult than the spatial perspective judgements. Experiment 2 ruled out the possibility that this was merely the result of having to process cues involving a double negation. It is still not clear however, why level-1 visual perspective judgements were found to be more difficult. Most likely is that it is the result of cues to bodily orientation being stronger than cues to eye direction, in our stimuli at least. Further investigation may look to vary these factors independently. Alternatively, whilst Experiment 2 did not include cues, it still retained the concepts seen/unseen which may require more processing than in front and behind. A second reason to be cautious is that the commonality between these visual and spatial judgements may be the result of participants paying attention to similar cues across the two tasks, specifically cues to the avatar's eyes and his front. In our experiment, these are directly yoked to one another. This is not, however, always the case in the real world. The direct association between an object being in front of someone and visible by them can be broken by a person rotating their head or closing their eyes or by a second object being located between a person and the target object.

We also showed similarities in processing between level-2 visual perspective judgements and judgements of whether an object was to the left or the right of someone. In both cases, participants performed in a way that was consistent with mentally rotating to the position of the avatar. This part of the process for the two operations seems to be equivalent. On the other hand, again, there was a difference in absolute difficulty, with spatial judgements being more difficult than visual judgements. This is consistent with the idea that egocentric transformation is necessary, but not sufficient for these kinds of perspective-taking. That this kind of perspective-taking involves an egocentric transformation is consistent with the idea that perspective-taking is a simulative process (Goldman,

2006). If this is the case, perspective-taking should not only be related to the difficulty of assuming the other's position, but also to the difficulty of completing the task for oneself. This suggests that judging whether an object is to the left or to the right of someone may be more difficult than judging how something looks to someone because judging if something is to one's own left or right is more difficult than judging how something looks from one's own position.

While three conditions used identical stimuli, the stimuli for the Spatial II condition differed. We could not use the same stimuli from the other conditions as in these the object was neither to the left nor the right of the avatar. The only way to fully equate stimuli across conditions would have involved the avatar not looking directly at the object, or looking contrary to his body angle. It was felt that such a circumstance might cause confusion, particularly in judging the front of someone whose legs face one direction, but looks in another. It seems highly unlikely that unique features of the Spatial II stimuli were responsible for the strategy of rotation, for rotation was also evidenced in the Visual II condition. As already noted, however, level-1-type problems do not preclude a mental rotation strategy, and level-2-type problems do not necessitate the strategy. Not only may the spatial requirements of many level-2 tasks be solved with geometric transforms other than rotation, but in principle the level-2 distinction admits of differences of perspective between people with the very same spatial point of view, as when a person with normal vision sees an object as "in focus" whereas someone with impaired vision might see it as "blurry". What the present findings do indicate is that on highly similar tasks it was level-1 or level-2 *type* that seemed to dictate the basic features of the strategy, rather than visual or spatial *content*.

Whilst our most informative effects were related to angular disparity, we also found an effect of distance between the other person and the object in Experiment 1 (and a trend in Experiment 2). This effect was independent of whether the content of the judgement was visual or spatial and of the type of judgement that had to be made. One suggestion is that some form of line of sight/body is required for all these judgements (Kessler & Thomson, 2010; Michelon & Zacks, 2006). A second suggestion is that this difference in difficulty is really more due to the distance across which we must sustain our visual attention to connect the Other and the Object (Carlson & Van Deman, 2008; Michelon & Zacks, 2006). Our results do not distinguish between these hypotheses.

6.4. The difficulty of defining perspective-taking

Earlier, we defined perspective-taking in terms of a target other's relationship with an object. Showing similarities across different contents of representation, spatial and visual, is informative about commonalities in processing strategies. What should be apparent, both in our minimal definition and through our results, is that these processes are not specific to representing the views of other people. One might predict similar findings for judging if something were "in front of" a car for instance. This does not entail that there are no distinctive processes involved in attributing a representational perspective to another person, and of course, visual perspective-taking must take into account factors such as occlusion through a blindfold or by a barrier, whilst this is not necessarily the case for spatial perspective-taking. However, it is also apparent that many processes for perspective-taking or Theory of Mind may be achieved with only a minimal, cognitive architecture that does not involve ascribing mental states as such (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). Similarly, recent, broader definitions of the notion of perspective suggest a commonality across mental and non-mental contents for tasks that require some form of representation or reference (Perner, Mauer & Hildebrand, 2011). Thus, by highlighting important commonalities between visual and spatial perspective-taking, the current findings suggest that the project of identifying processes that are actually distinctive for one or other type is an important direction for futurework.

7. Conclusion

Judgments about where things are located in relation to someone and how they see things share many superficial similarities. Here we tested the hypothesis that they will also show similarities in the computational processes they engage. There are two routes to both visual and spatial perspectives, only one of which requires an egocentric body transformation. When asked about the appearance of an object to someone else or whether the object is located to their left or their right, our evidence suggests that we mentally rotate ourselves to their position before making a judgement. When asked about whether an object is in front of someone or whether they can see an object, we do not do this. What seems more likely is that we directly identify the relationship between the person's front or eyes and the object. Furthermore, it is possible that visual perspective-taking relies on basic spatial perspective-taking processes as a foundation, but sometimes requires more complex processing. When and how this processing occurs should be a focus for further investigation.

Acknowledgements

Appendix A. Analyses of error rates in Experiments 1 and 2

A.1. Experiment 1

A $4 \times 2 \times 2 \times 2$ ANOVA with Angle ($0°$, $60°$, $120°$, $180°$) and Distance (Near, Far) as within subjects factors and Content (Visual, Spatial) and Type (Level-1 type, Level-2 type) as between subjects factors was completed on Error proportion (detailed in Table A). Rates of accuracy across all conditions were high. We only investigated hypothesis driven effects on accuracy to avoid the risk of making

Table A

Analyses of accuracy in Experiments 1 and 2. All statistics are Greenhouse–Geisser corrected to guard against violations of sphericity.

	dF	F	p	ηp^2
EXPERIMENT 1				
Angle	3	2.97	.044	.047
Linear	3	1.60	.211	.078
Distance	1	.454	.503	.008
Content	1	11.33	.001	.159
Type	1	.172	.680	.003
Angle × Type	3	.733	.509	.012
Angle × Content	3	.183	.875	.003
Angle × Distance	3	.733	.509	.012
Distance × Type	1	2.13	.15	.034
Distance × Content	1	1.53	.222	.025
Content × Type	1	8.18	.006	.120
Angle × Content × Type	3	.477	.682	.007
Angle × Content × Distance	3	1.22	.304	.02
Angle × Distance × Type	3	2.054	.114	.033
Distance × Content × Type	1	3.23	.078	.051
Angle × Content × Type × Distance	3	3.12	.032	.049
Level-1 type				
Angle	3	1.68	.196	.053
Linear	3	.023	.881	.001
Level-2 type				
Angle	3	2.01	.128	.063
Linear	3	2.61	.116	.080
EXPERIMENT 2				
Angle	3	9.21	.005	.235
Linear	3	18.70	<.001	.384
Distance	1	.638	.431	.021
Content	1	17.41	<.001	.367
Type	1	9.21	.005	.235
Angle × Type	3	10.69	<.001	.263
Angle × Content	3	2.78	.077	.085
Angle × Distance	3	4.71	.005	.136
Distance × Type	1	3.55	.069	.106
Distance × Content	1	0	1.00	0
Content × Type	1	13.75	.001	.314
Angle × Content × Type	3	2.85	.054	.087
Angle × Content × Distance	3	2.06	.114	.064
Angle × Distance × Type	3	1.58	.213	.050
Distance × Content × Type	1	5.87	.022	.164
Angle × Content × Type × Distance	3	3.878	.023	.114
Level-1 type				
Angle	3	2.11	.122	.066
Linear	3	.685	.415	.022
Level-2 type				
Angle	3	14.67	<.001	.328
Linear	3	24.97	<.001	.454

t-tests to assess separately the effect of content at each level			
	df	t	p
EXPERIMENT 1			
Level-1 Type			
Content	15	.911	.725
Level-2 type			
Content	15	4.45	<.001	Visual < Spatial
EXPERIMENT 2			
Level-1 type			
Content	15	1.18	.246
Level-2 type			
Content	15	4.31	<.001	Visual < Spatial

Type I errors (Barr, 2008), this allows for the investigation of any trade-offs between speed and accuracy without relying on post hoc interpretations. The main effect of Angle was somewhat hard to interpret, with data showing a quadratic, rather than linear distribution. Neither the Angle x Type interaction, nor the Angle x Content interaction reached significance, suggesting that the findings of the Response time analysis were not the

result of differing trade-offs between speed and accuracy across conditions.

A.2. Experiment 2

A $4 \times 2 \times 2 \times 2$ ANOVA was completed with Angle ($0°$, $60°$, $120°$, $180°$), Distance (Near, Far) and Type (Level-1 type, Level-2 type) as within subjects factors and Content (Visual, Spatial) as a between subjects factor, the results are presented in Table A. Rates of accuracy across all conditions were high. In support of our analysis of response times, there was a linear effect of Angle and an interaction between Angle and Type detailing a linear effect for Level-2 type, but not for Level-1 type. Again, the Angle by Content interaction was not significant.

References

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

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, Special Issue: Emerging Data Analysis., 59*, 457–474.

Bialystok, E., & Codd, J. (1987). Children's interpretations of ambiguous spatial descriptions. *British Journal of Developmental Psychology, 5*, 205–211.

Bowerman, M., & Choi, S. (2003). Space under construction: Language-specific spatial categorization in first language acquisition. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind*. Cambridge MA: MIT Press, pp. 387–428.

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

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind and Language, 28*.

Carlson, L. A., & Van Deman, S. R. (2008). Inhibition within a reference frame during the interpretation of spatial language. *Cognition, 106*, 384–407.

Carlson-Radvansky, L. A., & Irwin, D. E. (1993). Frames of reference in vision and language – Where is above. *Cognition, 46*, 223–244.

Carlson-Radvansky, L. A., & Jiang, Y. H. (1998). Inhibition accompanies reference-frame selection. *Psychological Science, 9*, 386–391.

Cox, M. V. (1981). Interpretation of the spatial prepositions in front of and behind. *International Journal of Behavioral Development, 4*, 359–368.

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

Flavell, J. H. (1988). The development of children's knowledge about the mind: From cognitive connections to mental representations. In J. Astington, P. Harris, & D. Olson (Eds.), *Developing theories of mind* (pp. 244–267). New York: Cambridge University Press.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young childrens knowledge about visual-perception – further evidence for the level 1-level 2 distinction. *Developmental Psychology, 17*, 99–103.

Goldman, A. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. New York: Oxford University Press.

Hands, L. J. (1972). Discrimination of left and right and the development of the logic of relations. *Merrill-Palmer Quarterly, 18*, 307–320.

Hare, B., Brown, M., Williamson, C., & Tomasello, M. (2002). The domestication of social cognition in dogs. *Science, 298*, 634–1636.

Harris, L. J., & Strommen, E. A. (1972). Role of front-back features in childrens front, back, and beside placements of objects. *Merrill-Palmer Quarterly of Behavior and Development, 18*, 259–271.

Kaminski, J., Riedel, J., Call, J., & Tomasello, M. (2005). Domestic goats (Capra hircus) follow gaze direction and use some social cues in an object choice task. *Animal Behaviour, 69*, 11–18.

Kessler, K., & Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology (Special Issue on: Embodied and Grounded Cognition), 1*(213), 1–12.

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

Kessler, K., & Wang, H. (2012). Spatial perspective taking is an embodied process, but not for everyone in the same way: Differences for gender and social skills score. *Spatial Cognition and Computation. Special Issue on: "Unusual Bodies, Uncommon Behaviors: Embodied Cognition and Individual Differences in Spatial Tasks, 12*, 133–158.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. *Psychological Sciences, 11*, 32–38.

Keysar, B., Lin, S. H., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25–41.

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition, 29*, 745–756.

Levinson, S. C. (1996). Language and space. *Annual Review of Anthropology, 25*, 353–382.

Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity*. Cambtridge: Cambridge University Press.

Masangkay, Z. S., Mccluske, K. A., Mcintyre, C. W., Simsknig, J., Vaughn, B. E., & Flavell, J. H. (1974). Early development of inferences about visual percepts of others. *Child Development, 45*, 357–366.

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

Moll, H., & Tomasello, M. (2006). Level I perspective-taking at 24 months of age. *British Journal of Developmental Psychology, 24*, 603–613.

Perner, J., Mauer, M. C., & Hildenbrand, M. (2011). Identity: Key to Children's Understanding of Belief. *Science, 333*, 474–477.

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science, 308*, 214–216.

Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London: Routledge & Kegan Paul.

Pickering, M. J., McLean, J. F., & Gambi, C. (2012). Do addressees adopt the perspective of the speaker? *Acta Psychologica, 141*, 261–269.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind. *Behavioral and Brain Sciences, 1*, 515–526.

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

Rakoczy, H. (2012). Do infants have a theory of mind? *British Journal of Developmental Psychology, 30*, 59–74.

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

Santos, L. R., Flombaum, J. I., & Phillips, W. (2007). The evolution of human mindreading: How non-human primates can inform social cognitive neuroscience. In S. Platek, J. P. Keenan, & T. K. Shackelford (Eds.), *Evolutionary cognitive neuroscience* (pp. 433–456). Cambridge, MA: MIT Press.

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*, 199–204.

Song, H. J., & Baillargeon, R. (2008). Infants' Reasoning About Others' False Perceptions. *Developmental Psychology, 44*, 1789–1795.

Surtees, A. D. R., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development, 83*, 452–460.

Surtees, A. D. R., Apperly, I. A. & Samson, D. (in revision). Automatic perspective-taking calculates whether something is seen, but not how it is seen.

Surtees, A. D. R., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology, 30*, 75–86.

Surtees, A. D. R., Noordzij, M. L., & Apperly, I. A. (2012). How to lose yourself in space: Children and adults' use of frames of reference and perspectives. *Developmental Psychology, 48*(185), 191.

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

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

Zacks, J., Mires, J., Tversky, B., & Hazeltine, E. (2000). Mental spatial transformations of objects and perspective. *Spatial Cognition and Computation, 2*, 315–332.