Consciousness and Cognition 41 (2016) 93–103

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Evidence for spontaneous level-2 perspective taking in adults

Fruzsina Elekes [a,b,c,*], Máté Varga [d], Ildikó Király [b,e]

[a] Doctoral School of Psychology, Eötvös Loránd University, Izabella str. 46, Budapest 1064, Hungary
[b] Institute of Psychology, Eötvös Loránd University, Izabella str. 46, Budapest 1064, Hungary
[c] Cognitive Development Center, Central European University, Október 6. str. 7, Budapest 1051, Hungary
[d] Department of Cognitive Science, Budapest University of Technology and Economics, Műegyetem rkp. 3., Budapest 1111, Hungary
[e] Institute for Advanced Study, Central European University, Október 6. str. 7, Budapest 1051, Hungary

ARTICLE INFO

Article history:
Received 10 December 2015
Revised 11 February 2016
Accepted 12 February 2016
Available online 18 February 2016

Keywords:
Spontaneous level-2 perspective taking
Aspectuality
Awareness
Social partner
Two systems approach in ToM

ABSTRACT

Social interactions are fostered by humans' propensity to compute their partner's perspective online. However, due to the mindreading system's limited capacity perspective taking (PT) was argued to occur spontaneously only for level-1, but not level-2 perspectives. We propose that level-2 perspectives (containing aspectual information) can also be computed spontaneously if participants have reason to assume that the partner is indeed aware of the objects' aspectual properties. Pairs of adult participants took part in the modified version of Surtees, Butterfill, and Apperly's (2012) number verification paradigm. Participants had prior information on their partner's task, which either called for processing aspectual properties or did not. The partner's inconsistent perspective was found to interfere with RT-s providing evidence for spontaneous level-2 PT. However, such interference only occurred when the partner's task involved processing the perspective dependent object feature, suggesting that PT was sensitive to the other's awareness of the to be represented information.

1. Introduction

Visual perspective taking refers to the ability to mentally map how a certain scene looks from another person's point of view. Being part of the mind reading system, perspective taking (PT) provides the basis for attributing knowledge or beliefs to others and thus lays the foundation for smooth social interactions (Aichhorn, Perner, Kronbichler, Staffen, & Ladurner, 2006; Apperly, 2008; Wimmer, Hogrefe, & Perner, 1988). While its relevance is widely recognized, the features and functioning of PT are strongly debated. In recent literature it has been argued that visual perspective taking is not a unitary capacity either in terms of the computed representation, or regarding the mechanism that leads to that representation (Apperly & Butterfill, 2009; Flavell, Everett, Croft, & Flavell, 1981; Rakoczy, 2012). The opposition to this view claims that there is only one mindreading system that, at times, recruits other cognitive faculties as well, but uses the same concepts regardless (Carruthers, 2015a). Our findings indicate that the division between mindreading systems is not as rigid as proposed by the former approach.

Based on empirical findings in preschoolers, Flavell et al. (1981) proposed that two types of information could be achieved regarding the visual perspective of others. Level-1 PT refers to representing *whether* an agent can see an object, while level-2 PT means representing *how* exactly that object appears to him, that is, under what aspect the agent sees the object. This distinction indicates that there is a qualitative gap between knowing what the other does and does not see, and being able to

* Corresponding author at: Doctoral School of Psychology, Eötvös Loránd University, Izabella str. 46, Budapest 1064, Hungary.
E-mail address: elekes.fr@gmail.com (F. Elekes).

http://dx.doi.org/10.1016/j.concog.2016.02.010
1053-8100/

represent the scene as it is visible to him/her. Underlying this notion it has been demonstrated that the ability to compute someone else's level-1 perspective develops earlier in life than the representationally more complex capacity of level-2 PT (Moll & Meltzoff, 2011; Moll & Tomasello, 2006; Sodian, Thoermer, & Metz, 2007).

Despite the bias often shown in children and adults towards egocentrism (for a review see, Samson & Apperly, 2010), there is evidence that level-1 perspective taking can emerge in a speeded way and without instruction to do so (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010). In the number verification paradigm used by Samson et al. (2010), participants had to verify the amount of dots presented on the walls of a virtual room either from their own perspective, or from the perspective of an avatar located in the room. The amount of dots visible to the avatar either matched or did not match that visible to the participant, potentially creating a conflict with one's own perspective.

The results suggested (Samson et al., 2010) that adults computed the avatar's perspective online despite the fact that it was irrelevant for decision-making. This, in turn, interfered with participants' decisions when the avatar's perspective was inconsistent with their own (altercentric intrusion). Importantly, similar altercentric interference emerged when participants only had to make judgments based on their own perspective throughout the experiment. This rules out the possibility that the high inhibition demands of switching back and forth between perspectives, or the situation that trained participants to place themselves into the other's perspective played a role in the effect. Similar to adults, school-aged children also showed altercentric intrusions in this paradigm (Surtees & Apperly, 2012). Finally, computation of the other's perspective was found to be independent of parallel cognitive load, indicating that the process was indeed cognitively efficient (Qureshi, Apperly, & Samson, 2010).

Unlike level-1 PT, level-2 PT has not been reported to occur in a spontaneous way. Surtees, Butterfill, and Apperly (2012) presented subjects with single numerals that were either symmetric/unambiguous in nature (0,8) or asymmetric/ambiguous (6,9). The numbers were presented either lying on the table between the participant and the avatar (asymmetric stimuli looked different from the two perspectives), or were displayed on the wall (all stimuli looked the same independent of perspectives). Participants had to perform a number verification task from their own or the avatar's perspective. In this case, the avatar's inconsistent perspective did not interfere with egocentric perspective judgments, suggesting that adults did not compute how the scene looked from the avatar's perspective spontaneously.

Before outlining current views on the cause of the above difference, an important distinction has to be drawn between automatic and spontaneous processing. Although there are many different approaches to automaticity, an automatic cognitive process is thought to be independent of both the participant's overt goal, and of any covert goals he might have (Carruthers, 2015a). On the other hand, while spontaneous processes are also independent of overt goals or external prompting, they do depend on participants' covert goals (for example, the general motivation to understand others, Carruthers, 2015a), or on contextual factors (Back & Apperly, 2010).[1] Samson et al.'s (2010) findings were interpreted as evidence for the "relatively automatic" computation of level-1 perspectives (Qureshi et al., 2010; Surtees & Apperly, 2012). For the sake of definitional clarity we will continue to refer to these results as spontaneous PT.

The difference in spontaneity of computation found between level-1 and level-2 PT might bring us closer to understanding the mechanism behind these abilities. It has been argued that the two skills, level-1 and level-2 PT, differ in terms of their reliance on perspective computation, and relatedly, in the degree to which they are social in nature. According to some, the ability to figure out which objects someone does or does not see (level-1 PT) might not even require reasoning about perspectives at all (Aichhorn et al., 2006; Michelon & Zacks, 2006). Aichhorn and colleagues (2006) argue that differing perceptual experiences have to refer to the same objects or scene in order to qualify as *perspectives* on those objects, while in level-1 decisions the different percepts can be interpreted as resulting from a difference in the objects that are looked at. Without using the concept of seeing, perceptual access to an object can be judged based on the spatial relation between the other person's eyes and the target object (Aichhorn et al., 2006; Michelon & Zacks, 2006). Empirical findings support this proposal. Adults are quicker to make explicit level-1 decisions when the avatar is close to the target object and are slower when the distance is greater, but the speed of computation is not affected by the angular disparity between participant and avatar (Michelon & Zacks, 2006; Surtees, Apperly, & Samson, 2013a). This indicates that the information (see/does not see) is reached through tracing the person's line of sight, which line takes longer to "draw" if the path is longer.

The idea that tracking visual access to certain objects does not involve representing the perspectives of social agents gains further support from a different line of investigation as well. Studies show altercentric interference in Samson et al.'s (2010) number verification task also when the avatar is replaced by a less or non-social, directional stimulus (Nielsen, Slade, Levy, & Holmes, 2015; Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014). Based on this, Santiesteban et al. (2014) argue that the phenomenon referred to as level-1 perspective taking is driven by domain general factors, like attentional cueing, rather than Theory of Mind. The effect in Nielsen et al.'s (2015) study was, however, stronger in the social condition compared to the less and non-social conditions. Furthermore, the effect correlated with self-reported measures of Theory of Mind in the social, but not in the other two conditions. This suggests that processes specific to the social domain also contribute to spontaneous level-1 PT.

As opposed to tracing someone's line of sight, representing appearances from another point of view, that is taking someone's level-2 perspective, presumes understanding that the same object from a different angle may give rise to different percepts. Hence, seemingly contradictory contents regarding the same referent (e.g. the object's perceived identity) may all be

[1] A related term is involuntariness (Bargh, 1989). While automaticity/spontaneity refers to the features of launching a process, involuntariness of computation indicates that a process will necessarily be performed to the end if started, it cannot be down-regulated or controlled even if the perceiver is aware of its' operation.

true, given they belong to different perspectives (Perner, Stummer, Sprung, & Doherty, 2002). Accordingly, when making explicit perspective decisions, adults were influenced by the angular disparity between their own and the avatar's perspective, responding to greater angles more slowly (Surtees, Apperly, & Samson, 2013b, 2013a). This indicates that the mechanism behind explicit level-2 PT resembles mental rotation, where participants mentally rotate themselves in space to align to the other's perspective (Surtees et al., 2013b). This process builds heavily on working memory and is thus cognitively demanding in nature.

The above difference between the characteristics of level-1 and level-2 PT is in line with Apperly and Butterfill's (2009) two systems account, which postulates that the mindreading system, and perspective taking as part of it, consists of two kinds of processes. One process is cognitively demanding and slow but flexible in nature, while the other is quick and efficient at the expense of being inflexible. The latter is called the Minimal ToM and it enables the online guidance of behavior in social contexts (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). However, the quick and efficient system has a content specific limit, being unable to represent representations *as such*. Instead, it uses certain principles to track mental states without actually representing them. Briefly, the online mindreading system can track *relations* between agents and objects (i.e. level-1 perspectives), but it fails when information on object *appearance* (i.e. level-2 perspectives), or propositional mental states have to be computed.

Apperly and Butterfill's (2009; Butterfill & Apperly, 2013) theory has been criticized heavily recently (Carruthers, 2015a, 2015b; Christensen & Michael, 2015). Carruthers (2015a, 2015b) argues that the data can be accounted for by assuming the existence of one mindreading system that can work automatically, spontaneously, or deliberately depending on context. Its functioning is automatic (or at least spontaneous) whenever mental state attribution does not necessitate the use of other cognitive capacities, like working memory or executive functions. Whenever other cognitive faculties have to be recruited (like the working memory requirement of mental rotation in the case of level-2 PT), it is the encoder's explicit goal that determines whether he will allocate cognitive resources into computing the other's mental states. Hence, computation in this case will not be automatic. Importantly, in opposition to Apperly and Butterfill (2009) and Butterfill and Apperly (2013), the model does not assume a qualitative difference between representations attributed automatically and intentionally.

In addition to its sensitivity to the cognitively demanding nature of perspective computation, the mindreading system has a further restriction according to Carruthers. Namely, that due to its limited capacity, the system cannot provide a complete model of the *whole* visual scene. Perspective attributions are only made for those objects that the other person is aware of (Carruthers, 2015a). Awareness of certain objects is indicated by, for example, the other's gaze direction. Thus, the one system model of mindreading predicts that automatic computation of another's perspective will occur whenever that does not require executive functions, *and* appropriate guidance is available for the encoder to judge the other's awareness of the target stimuli. The lack of automatic PT might be caused by either or both of these factors.

For the sake of argument, let us assume that under certain conditions (say, sufficient background knowledge on the object's aspectual nature to enable shortcutting the mental rotation process), level-2 PT can be achieved without exerting a high load on working memory. Under these conditions, the factor that determines if the mindreading system will be engaged automatically is what the encoder knows about the other's awareness of the stimuli. Carruthers (2015a) argues that for level-1 PT the mindreading system selects *which objects* to make attributions about. However, even if the same concept of seeing is used for level-1 and level-2 PT, as hypothesized by Carruthers (2015a, 2015b), level-2 PT involves representing more detailed information (the object's appearance from another aspect) than level-1 PT. We propose that in the case of level-2 PT, the partner's awareness is taken into account on the level of object *features*. Hence, cues indicating that the other is aware of the *presence* of an object might not be sufficient to launch the process.

Merely looking in one direction does not obligatorily result in perceiving objects in that trajectory (Mack & Rock, 1998) and more specifically, looking at an object in space does not necessarily lead to consciously perceiving all of its features (Levin & Simons, 1997). For example, one might look at an object that has perspective dependent properties, e.g. form, and attend only to its non-perspective dependent features, e.g. texture or color. Importantly, adults' expectations have been found to be sensitive to the distinction between looking in one direction and perceiving information present there (Teufel & et al., 2009). Based on these considerations, we assume that, as a precondition of spontaneous level-2 perspective taking, the other's awareness of the stimuli is taken into consideration on the level of object *features*, not objects in general. Limiting perspective computation to the information that the partner is aware of is a beneficial strategy, as the partner's behavior will most likely be guided by that subset of information.

Attention is one of the cognitive processes that modulates conscious perception by highlighting certain bits of visual information and filtering out others (Broadbent, 1958); and while it might not be sufficient in itself, attention certainly is necessary for information to reach awareness (Cohen, Cavanagh, Chun, & Nakayama, 2012). Importantly, when reasoning about others, knowledge about the partner's goal can be used to determine what pieces of information he/she pays attention to, which, in turn, is a reasonably good indicator of what the partner is aware of out of all of the information that he/she has perceptual access to.

We hypothesized that if adults had prior information that the partner was not only looking at an object but also had awareness of the aspectual, perspective dependent properties of that object, they would compute the partner's level-2 perspective spontaneously. On the other hand, no PT was expected if participants' had no reason to presume awareness of those features.[2] The modified version of Surtees et al.'s (2012) number verification paradigm was used to test the proposal. Instead of using avatars as "social" partners, pairs of participants took part in the experiment sitting on opposite sides of a short table on which stimuli were displayed. Subjects participated in one of two groups. In the perspective-dependent (PD) group, both participants performed a number verification task. This ensured that the participant knew that the partner was attending to and was consequently aware of the object's form, that determines the number it represents, and potentially looks different from different aspects. In the non-perspective-dependent (NPD) group, however, the participant performing the number verification task knew that the partner's task (an n-back task) did not entail encoding aspectual information about the object.

We predicted interference from the other's inconsistent perspective for mutually attended stimuli in the PD group, as there participants had knowledge that the partner was necessarily aware of the aspectual object property. On the other hand, no interference was expected in the NPD group, due to the lack of evidence of the confederate's awareness of perspective dependent properties. Interference in the current study will be interpreted as an indicator of spontaneous, rather than automatic PT, as it can be argued that a live interaction makes the computation of the partner's perspective relevant without external prompting to do so.

2. Study 1

2.1. Method

2.1.1. Design

A 2 × 2 × 2 (Jointness [individual, joint] × Symmetry [symmetric number, asymmetric number] × Task [PD, NPD]) factorial design was used with Jointness and Symmetry as within subject factors and Task as a between subject variable. We used a within subject individual control instead of the "wall" trials of Surtees et al. (2012), as we placed participants into real 3D space, where that could not be implemented.

2.1.2. Participants

Data was collected from 54 university students (47 females, $M_{age} = 21.85$, $SD = 4.56$). Participant pairs were randomly assigned to groups. Eighteen participants were assigned to the PD group and 36 to the NPD group. As our question was whether the other's inconsistent perspective interfered with number verification we did not analyze data from the *n*-back task. Thus, 18 of the 36 people in the NPD group merely served as pairs for the participants. Participation was rewarded by course credit. The experiment was approved by the ethical committee of our university. All participants signed an informed consent form before starting the experiment.

2.1.3. Materials

Following Surtees et al. (2012) four numbers, two symmetric (0,8) and two asymmetric (6,9), were drawn using Matlab R2013a. The style of the visual stimuli resembled the "digital" numbers (for an illustration see, Table 1). The numbers were displayed in two different widths and two different colors (blue and green) which were varied randomly to increase the overall perceptual variability. The numbers were 11.5 cm in height, and 5.2 or 6.2 cm in width on a 21.5 in. flat monitor.

The audio stimuli were presented in a female voice with neutral and descending intonation. In the case of both symmetric and asymmetric numbers one of the two words contained one syllable while the other contained two syllables in Hungarian, the native language of the participants. The length of the audio stimuli was 760 ms for "eight" (nyolc), 760 ms for "zero" (nulla), 600 ms for "six" (hat), and 760 ms for "nine" (kilenc). The audio and visual stimuli pairings provided four types of events defined by the symmetry of the visual stimuli and correspondence between visual and audio stimuli (for details, see Procedure).

2.1.4. Procedure

Upon arriving to the lab, participants were introduced to each other and were seated next to each other in the reception room to sign the informed consent form and listen to instructions. The two groups, PD and NPD tasks, differed only in the instructions that were given (see, Appendix). In the PD group participants had to decide whether the number heard was the same as the number seen (number verification task). In the NPD group, one of the participants performed the number verification task, while the other had to decide whether the color of the stimulus currently on screen was the same as it was of the previous stimulus (*n*-back task). In contrast to Surtees et al. (2012), in the current experiment participants only had to make decisions based on their own perspective (self trials) to make the expected interference effect a stringent measure of perspective taking, rather than biased by the additional effects of task-switching.

After completing the consent forms, the experimenter escorted participants to the test room and seated them at opposite sides of a short table, facing each other. A 21.5 in. flat monitor was used to provide stimuli for both participants, and was laid

[2] Note, that spontaneous PT is necessarily limited to situations where there is a possibility to compute the other's level-2 perspective without exerting a high load on working memory or executive functions (Carruthers, 2015a). We argue that this is possible if one has pre-existing conceptual knowledge on the object's aspectual properties.

Table 1
Mean RT-s in milliseconds (with Standard Deviations in parentheses) in Study 1 and 2 as a function of Jointness, Symmetry, Correspondence and Task.

	Study 1							
	0/8 6/9							
	Individual				Joint			
	Symmetric		Asymmetric		Symmetric		Asymmetric	
	Corres	Noncorres	Corres	Noncorres	Corres	Noncorres	Corres	Non corres
PD	685.2 (109.97)	752.91 (129.11)	677.85 (135.9)	692.07 (122.04)	696.69 (116.03)	772.16 (112.09)	730.62 (134.15)	740.48 (102.12)
NPD	704.46 (128.31)	786.13 (152.66)	724.45 (138.38)	720.32 (155.44)	713.56 (160.61)	795.54 (190.39)	729.74 (182.12)	738.26 (174.65)
	Study 2							
	0/8 6/9							
PD	684.77 (121.25)	757.58 (122.87)	694.76 (135.08)	700.99 (129.74)	700.26 (92.28)	785.53 (94.1)	737.34 (132.96)	753.8 (118.41)

Note: The numbers (0/8, 6/9) are examples of the visual stimuli used in different widths and colors. PD refers to the perspective-dependent, NPD refers to the non-perspective-dependent group.

on the table between them. This way, symmetric stimuli (0, 8) looked the same irrespective of perspective, while asymmetric stimuli (6, 9) looked different from the two opposing perspectives. The audio stimulus was displayed through a pair of loudspeakers placed at equal distance from the two participants.

The script that was used for stimulus presentation and response recording was written in PsychoPy 1.81. Each trial started with a fixation cross presented at the center of the screen. The onset of the audio stimulus was 500 ms after the fixation cross appeared. The visual stimulus was presented after 300 ms following the onset of the audio stimulus and remained on the screen as long as the program received an answer (button press) from one or both participants depending on condition. Half of the participants used their right hand to indicate "yes" answers, while half used their left hand. Responses were made on Cedrus type response boxes.

Surtees et al.'s (2012, p. 79) presented the picture right after the audio ended ("participants viewed successive fixation stimuli ... followed by a 1800 ms auditory stimulus ... and then the test picture"). The aim of the shorter stimulus onset asynchrony in the current experiment (the visual image appeared before the audio ended) was to make it less likely that participants could form a mental image of the number presented in the auditory modality before they saw the picture. This mental image could then be matched in visual features with the visually presented number, without actually deciding what number the visual character depicted. We call this the visual matching shortcut. Using this strategy, participants would not engage in the number verification decision that is expected to be influenced by the other's perspective ("What number is depicted on the screen?").

The experiment consisted of three main test phases and an individual practice block containing 16 trials for both members of the pair. The test started with 16 practice trials for participant A, followed by the individual condition for the same participant. After this, participant B took part in the practice phase, followed by the joint condition. Finally, participant B proceeded with the individual condition. This setup ensured that half of the participants had the individual trials first, while the other half the joint trials. Each test phase was divided into two blocks with a 30-s-long rest phase between them. For the individual blocks, when only one of the participants performed the task, the other turned 180°, sitting with her back to the participant.

All three test phases contained a total of 112 trials, half of those depicted symmetrical visual stimuli (0, 8), half depicted asymmetrical visual stimuli (6, 9). Visual stimuli were paired with either corresponding or noncorresponding audio, resulting in the following four event types: symmetric corresponding (0 – "zero" and 8 – "eight"), symmetric noncorresponding (0 – "eight" and 8 – "zero"), asymmetric corresponding (6 – "six", 9 – "nine"), and asymmetric noncorresponding (6 – "nine", 9 – "six").

2.1.5. Data analyses

Only data collected from the number verification task was analyzed that is, both participants' responses in the PD group and one participant's response from each pair in the NPD group. When analyzing RT-s, incorrectly answered trials and outlier data points were removed. An outlier data point was defined as an RT that differed by more than two standard deviations from the mean RT of the given participant. The percentage of correctly answered trials (hit rate) was also calculated and analyzed.

2.2. Results and discussion

2.2.1. Mean reaction times

We expected to find a Jointness × Symmetry interaction for both corresponding and noncorresponding trials in the PD, but not in the NPD group. As a first step, we performed a 2 × 2 × 2 × 2 mixed ANOVA (Jointness [individual, joint] ×

Symmetry [symmetric, asymmetric] × Correspondence [corresponding, noncorresponding] × Task [PD, NPD]) on the mean RT-s (see Table 1, and Fig. 1) with Task as a between subject factor. A Jointness × Symmetry × Task interaction would confirm our predictions and allow us to run separate ANOVA-s for the two experimental groups.

The main effect of Correspondence, $F(1,34) = 19.88$, $p < .0001$, $\eta_p^2 = .369$, and Symmetry, $F(1,34) = 32.71$ $p < .0001$, $\eta_p^2 = .490$, reached significance, with shorter RT-s on corresponding and asymmetric trials. The main effect of Jointness approached significance, $F(1,34) = 3.22$, $p = .082$, $\eta_p^2 = .087$, responses being quicker on individual trials. We found a significant Jointness × Symmetry interaction, $F(1,34) = 7.97$, $p = .008$, $\eta_p^2 = .190$. As expected this was further qualified by a significant three way interaction between Jointness, Symmetry and Task, $F(1,34) = 6.18$, $p = .018$, $\eta_p^2 = .154$. Additionally, the interaction of Symmetry and Correspondence was also significant, $F(1,34) = 16.62$, $p < .0001$, $\eta_p^2 = .328$. The Symmetry × Correspondence interaction was specifically caused by the longer RT-s for symmetric noncorresponding stimuli and importantly it emerged independently of both Task and Jointness. This indicates that whatever caused the Symmetry × Correspondence interaction it did not interfere with the specific comparisons that were targeted.

To further clarify these data and to test if the expected RT pattern had emerged for both the corresponding and the noncorresponding trials 2 × 2 (Jointness [individual, joint] × Symmetry [symmetric, asymmetric]) repeated measures ANOVA-s were performed for these in both groups separately. In line with our predictions, in the PD group the Jointness × Symmetry interaction was significant for the corresponding trials, $F(1,17) = 7.41$, $p = .014$, $\eta_p^2 = .304$, and also approached significance for the noncorresponding trials, $F(1,17) = 3.92$, $p = .064$, $\eta_p^2 = .187$. Additionally, the main effect of Jointness was marginally significant for both the corresponding, $F(1,17) = 3.62$, $p = .074$, $\eta_p^2 = .176$, and the noncorresponding trials, $F(1,17) = 3.82$, $p = .067$, $\eta_p^2 = .183$. The main effect of Symmetry was significant for the noncorresponding trials, $F(1,17) = 12.47$, $p = .003$, $\eta_p^2 = .423$.

On the other hand, in the NPD group the Jointness × Symmetry interaction did not reach significance either for the corresponding, $F(1,17) = 0.73$, $p = .790$, $\eta_p^2 = .004$, or for the noncorresponding trials, $F(1,17) = 0.36$, $p = .555$, $\eta_p^2 = .021$, suggesting that the other's perspective did not interfere with decision making in this case. Again, the main effect of Symmetry was significant only for the noncorresponding trials, $F(1,17) = 22.37$, $p < .0001$, $\eta_p^2 = .568$. Overall, the data suggest that our findings were consistent over both corresponding and noncorresponding trials.

2.2.2. Hit rate

A 2 × 2 × 2 × 2 mixed ANOVA (Jointness [individual, joint] × Symmetry [symmetric, asymmetric] × Correspondence [corresponding, non-corresponding] × Task [PD, NPD]) was conducted on hit rates (for descriptive statistical data see Table 2).

Jointness had a significant main effect on performance, participants being more successful on individual trials, $F(1,34) = 4.831$, $p = .035$, $\eta_p^2 = .124$. Additionally, we found a significant Jointness × Correspondence interaction, $F(1,34) = 4.506$, $p = .041$, $\eta_p^2 = .117$. Corresponding decisions ("yes" answers) were more error prone in the joint compared to the individual trials than noncorresponding ("no") decisions. A tendency level interaction of Jointness and Task was also found that was caused by a decrease of performance from individual to joint trials in the PD group, but not in the NPD group, $F(1,34) = 3.483$, $p = .071$, $\eta_p^2 = .093$. The Jointness, Symmetry, Task three-way interaction was not significant, $F(1,34) = 0.028$, $p = .868$. These results converge with the RT findings in that Jointness tended to worsen performance only when the partner had to process the perspective dependent stimulus features. This however, did not differentially affect symmetric and asymmetric trials.

3. Study 2

As, to our knowledge, this is the first empirical support for spontaneous level-2 perspective taking in adults it was deemed important to test the robustness and replicability of this effect. Thus, the aim of Study 2 was to replicate the interference effect found in the PD group in Study 1 with a different group of participants and slightly modified visual stimuli.

As a consequence of the previously used digital font it might have been easier to make a decision concerning asymmetric numbers than symmetric numbers, as the former (6,9) differed in two line segments, while the latter (0,8) differed in one line segment (for an illustration see Table 1). To eliminate any possible effects of this difference, in Study 2 we displayed analog numbers where 0 and 8 are also easily distinguishable.

3.1. Methods

3.1.1. Design and participants

A total of 18 people participated in the study in pairs ($M_{age} = 22.78$, $SD = 3.67$, 11 women). All participants took part in the PD group where the within subject individual condition served as a control.

3.1.2. Materials

Only the visual stimuli were modified. Instead of the digital characters, we used analog numbers drawn in Matlab R2013a (for an illustration see Table 1). As in Study 1, the numbers were displayed in two different widths and two different colors, varied randomly. The numbers were 11.5 cm in height, and 5.2 or 6.2 cm in width. The visual images were blue and pink and

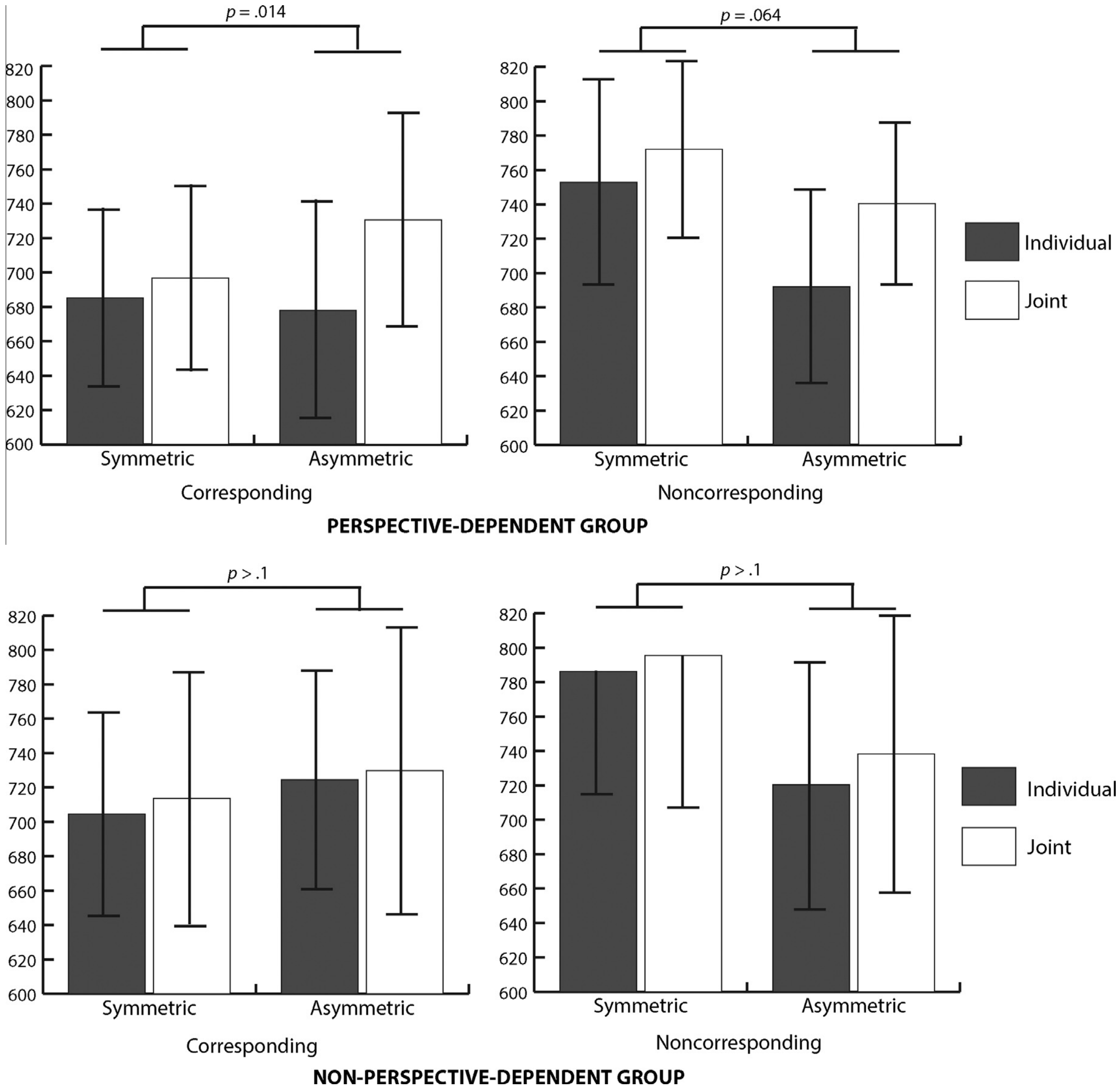

Fig. 1. Mean RT-s as a function of Jointness, Symmetry, Correspondence and Task. Error bars represent 95% Confidence Intervals.

Table 2
Mean percentage of correct responses (with Standard Deviation in parentheses) in Study 1 and 2 as a Function of Jointness, Symmetry, Correspondence and Task.

	Study 1							
	Individual				Joint			
	Symmetric		Asymmetric		Symmetric		Asymmetric	
	Corres	Noncorres	Corres	Noncorres	Corres	Noncorres	Corres	Noncorres
PD	95.23 (4.38)	92.4 (5.35)	93.73 (5.35)	94.35 (5.6)	89.55 (7.61)	91.91 (7.71)	89.84 (9.46)	88.94 (4.35)
NPD	92.95 (4.46)	90.36 (7.57)	94.78 (3.89)	92.5 (5.65)	92.87 (5.58)	91.74 (5.54)	92.01 (7.64)	92.71 (6.44)
	Study 2							
PD	94.49 (4.95)	88.54 (8.31)	93.98 (5.94)	93.57 (5.53)	93.83 (6.8)	94.06 (4.25)	91.47 (8.37)	92.29 (6.93)

participants had to press green and red buttons to indicate their answers (corresponding and noncorresponding respectively).

3.1.3. Procedure

The procedure was the same as in Study 1.

3.2. Results and discussion

3.2.1. Mean reaction times

A 2 × 2 × 2 ANOVA (Jointness [individual, joint] × Symmetry [symmetric, asymmetric] × Correspondence [corresponding, noncorresponding]) was run on the mean RT-s (for descriptive statistical data see Table 1). This revealed the main effect of Correspondense, corresponding decisions ("yes" answers) made more quickly, $F(1,17) = 14.513$, $p = .001$, $\eta_p^2 = .461$. The Symmetry × Correspondence interaction was found to be significant, $F(1,17) = 11.933$, $p = .003$, $\eta_p^2 = .412$. Just as in Study 1, this did not interact with Jointness. Crucially, we could replicate the Jointness × Symmetry interaction, implying that the other's perspective interfered with decision making more for the asymmetric than for the symmetric numbers, $F(1,17) = 9.922$, $p = .006$, $\eta_p^2 = .369$. Additionally, the main effect of Jointness approached significance, RT-s being marginally shorter in individual trials, $F(1,17) = 3.959$, $p = .063$, $\eta_p^2 = .189$.

Finally, we tested whether the same pattern emerged for both the corresponding and the noncorresponding trials, and found the expected interaction in both cases, corresponding: $F(1,17) = 4.417$, $p = .051$, $\eta_p^2 = .206$, noncorresponding: $F(1,17) = 5.345$, $p = .034$, $\eta_p^2 = .239$). Additionally, on the noncorresponding trials the main effect of both Jointness, $F(1,17) = 5.348$, $p = .034$, $\eta_p^2 = .239$, and Symmetry, $F(1,17) = 18.918$, $p < .0001$, $\eta_p^2 = .527$, was significant.

3.2.2. Hit rate

A Jointness [individual, joint] × Symmetry [symmetric, asymmetric] × Correspondence [corresponding, noncorresponding] ANOVA was performed on the hit rates (Table 2). In line with our expectations the analysis revealed a significant interaction between Jointness and Symmetry, $F(1,17) = 4.976$, $p = .039$, =.226. For asymmetric trials participants were more accurate in the individual than in the joint blocks, whereas the reverse was true for symmetric trials. Hence, hit rate provides converging evidence for the spontaneous computation of the partner's level-2 perspective. Additionally, the Jointness × Correspondense interaction was also found to be significant, $F(1,17) = 4.558$, $p = .048$, $\eta_p^2 = .211$, participants were better at making noncorresponding ("no") responses in the joint blocks, and corresponding ("yes") responses in the individual blocks. As the Jointness × Correspondence × Symmetry three-way interaction was significant on a tendency level, $F(1,17) = 3.809$, $p = .068$, $\eta_p^2 = .183$, we performed separate 2 × 2 Jointness [individual, joint] × Symmetry [symmetric, asymmetric] ANOVA-s for corresponding and noncorresponding trials. The Jointness × Symmetry interaction, signaling PT, was found to be specific to the noncorresponding trials, $F(1,17) = 11.432$, $p = .004$, $\eta_p^2 = .402$. In sum, indication of spontaneous level-2 PT was found in participants' accuracy as well, however, this was found to be more limited than the effect on RT-s.

4. General discussion

It has been consensual both in the empirical research and in various theoretical approaches that level-2 PT is a cognitively demanding, thus deliberate process. The findings presented here draw a more nuanced picture, suggesting that in certain contexts level-2 PT does emerge spontaneously. The interference effect we found in reaction times indicated that whenever participants took part in a task that required processing the perspective dependent object features a representation was formed about how the object looked from the partner's perspective. The representation emerged independently of participants' overt goal to make self perspective based decisions, and formed quickly enough to interfere with those decisions. This happened despite the fact that throughout the experiment participants only had to make judgments based on their own perspective. That is, they did not have to switch back and forth between tasks or perspectives, and thus, were not trained by the experimental situation to take the other's perspective as quickly as they could. The interference effect found in RT-s gained partial support from the hit rate measure as well.

Importantly, our findings showed that the fact that a fellow human being was actively making decisions regarding the same object from a different perspective was not sufficient for the interference effect to occur. Perspective interference only emerged when participants had evidence that their partner was processing the perspective dependent object feature as well. Recently, Carruthers (2015a) argued that, due to its limited capacity, the mindreading system cannot possibly give a complete model of the whole environment. Rather, it limits its operation to objects that are looked at by the confederate, that is, to objects that the agent is *aware of*. Our results indicate that cues more subtle than direction of gaze are used as a selection criterion. By varying the task of the partner, we manipulated which *stimulus feature* the other was aware of. Our findings suggest that spontaneous level-2 PT is limited to situations when someone from the other perspective is actually attending the perspective dependent features.

There is a related possibility that could have contributed to the effect and which can also be derived from the one system model of Carruthers (2015a). Namely, for spontaneous processes to be launched no external prompting or instruction is needed, however, the agent has to have the covert goal to perform that computation. Varying the partner's task also meant varying whether it was the same as one's own. Pursuing the same task as the other could have induced the implicit goal to understand and predict the other's actions to a greater extent than having different tasks. The above two explanations can both be accommodated with the one system model of ToM and can work in an additive way. Future studies will have to disentangle the role that these two factors played.

Evidence for spontaneous level-2 perspective taking is unique in the literature, and thus, the current studies have important implications for the existing models on Theory of Mind. Namely, the representation that caused interference necessarily contained more detailed information about the partner's perspective than merely tracking whether he had perceptual access to the object. Perspective interference could have resulted from representing either that whatever the partner saw looked *different* from what the participant himself could see, or from representing *how exactly* the object appeared to the partner.

Note, however, that according to the two systems account, the minimal ToM, that is responsible for online computation, cannot handle appearance related information (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). In its original form, the one system account of Carruthers (2015a, 2015b) predicts no spontaneous level-2 perspective taking either, albeit for a different reason. The one system model claims that any perspective content can be formed spontaneously given that its' computation does not require cognitively effortful processes. As explicit level-2 perspective judgments had been found to rely on a demanding, mental rotation like mechanism (Surtees et al., 2013a; Surtees et al., 2013b), level-2 PT was thought to be incompatible with spontaneity. Importantly, this model does not exclude the possibility of spontaneous level-2 PT *per se*, it excludes the possibility of a cognitively demanding computation performed spontaneously.

Two questions arise. First, do the results of the current experiments indicate that information that necessitates working memory (in the form of mental rotation) can be computed by the mindreading system spontaneously? Second, was the interfering representation actually *attributed* to the partner? Our answer to both questions is no, not necessarily.

Regarding the first question, we argue that it is possible to reach perspective information in the given paradigm without the cognitively demanding process of mental rotation. The number verification task builds on adults' relatively automatized capacity to recognize numerals. Participants most likely had associations between the concepts "six" and "nine" as they had had prior experience through education that these characters could be rotated into each other. Hence, the information, that the character appeared to depict a different number from the other's point of view, could have been acquired through accessing pre-existing conceptual knowledge, instead of using demanding mental rotation on each trial. There can, of course, be circumstances under which participants rely on mental rotation in this task (Surtees et al., 2013a). However, there is also a possibility to bypass working memory, which shortcut may have allowed participants to quickly map the scene from the other's perspective. The results presented here provide evidence that information on object appearance from a different perspective can be reached spontaneously and online. This finding can be accommodated with the one system account (Carruthers, 2015a), but contradicts the prediction of the two systems account (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013).

Now, turning to the issue of attribution, while it is possible, the interference effect we found does not prove that the representation that interfered was necessarily attributed to the confederate as a mental state in the form of a proposition or otherwise. Merely reaching the conflicting information (the *content* of the partner's perspective) through the conceptual shortcut would have been sufficient to encumber decision-making. There was of course a need to identify which information (e.g. six/nine) belonged to the egocentric perspective in order to make a correct decision. However, it is possible that the content of the partner's perspective was accessed without acknowledging it to be a perspective that belonged to a *specific person*. The phenomenon that a mental state is computed without being bound to a certain agent has already been documented in ToM research (Kampis, Somogyi, Itakura, & Király, 2013). What is more, for the conflicting content to interfere, the information did not even have to be acknowledged as constituting *a perspective at all*. Accessing the object's alternative identity without recognizing that it referred to the mutually viewed object under another aspect could just as well hinder decision-making. According to our findings, what did depend on prior information about the social context was whether the content of the partner's perspective was activated in the first place: it was only accessed if the partner was assumed to be *aware* of the perspective dependent stimulus feature.

It is widely assumed that the primary function of Theory of Mind is to explain and predict behavior in order to adjust one's own behavior (either in a cooperative or a competitive way) to others'. The phenomenon we tapped into clearly improves one's chances to coordinate behavior with others. The findings indicate that given certain contextual cues (e.g. prior information that the other is also attending the perspective dependent features), adults are predisposed to activate all information they have that enables them to map what others might see about the surrounding environment. In a live social interaction the mere activation of a pre-existing piece of knowledge on perspectives that are "occupied" by someone might create a readiness for actions guided by that information. This, in turn, provides the opportunity to respond quickly and adequately to those actions. However, the findings are neutral regarding whether the interference emerged as a consequence of a meta-representation formed on the partner's perceptual experience.

By providing the first evidence for spontaneous level-2 perspective interference the results of the current experiments support those claiming that adult mindreading has to be viewed in a more dynamic way than the two systems account does, devoting attention to the role of contextual factors or cues in launching Theory of Mind processes (German & Cohen, 2012). However, the results also raise numerous questions to follow upon. For instance, instead of using avatars as social partners, we employed a live perspective-taking situation, as it is commonly implemented in studies on joint action (Sebanz, Bekkering, & Knoblich, 2006) or task sharing (Sebanz, Knoblich, & Prinz, 2003). Although, positive findings with avatars certainly show the robustness of level-1 perspective taking, some phenomena might call for more than a schematic sign of a social partner to occur. Future studies will have to clarify whether the presence of a human agent was indeed necessary for level-2 PT to happen.

Finally, the results presented here provide evidence for the spontaneity of computation, as interference emerged independently of the participants' overt goal. The interfering content was reached online, quickly enough to hinder self-perspective based decisions. The effect did, however, depend on what the participant knew about the partner's task. This is in line with German and Cohen's (2012) argumentation that the sub-processes of ToM might only be engaged when specific combinations of stimuli (involving contextual cues as well) apply. According to some, any effect that relies on such preconditions fails to qualify as automatic (Back & Apperly, 2010). Additional research is needed to determine whether the effect resists external incentives to inhibit the distracting non-relevant perspective, providing evidence for its involuntariness, and whether interference also occurs under cognitive load implying the efficiency of the process.

In summary, the results suggest that the mindreading system indeed works in a dynamic way, making use of off-line cues about the confederate's task, and prior knowledge regarding the aspectual nature of objects. Our findings indicate that the dichotomy of automatic ToM processes, tracking only relational information, and flexible mindreading processes, computing mental states, might be less clear-cut than previously thought.

Acknowledgments

We thank Ágnes Kovács for helpful comments on the manuscript and Caoimhe O'Reilly for providing language check. This work was supported by the OTKA under Grant number K-109352 and K-116779.

Appendix A. Appendix

A.1. Instructions

A.1.1. First part of instructions (both groups)

"You are going to participate in this study together, both of you will have to make decisions on simple audio and visual stimuli. There will be parts when you work alone, during this time the other person will sit with his/her back to the participant who is working, and there will be a part when you work in parallel, during this time you will sit facing each other. Both of you will have a response box to indicate your answers. One after the other you will hear numbers from a loudspeaker. During this, you will see differently colored number characters on the screen, one for each number heard from the loudspeaker."

A.1.2. Second part of instructions (perspective-dependent group)

"You will not have to deal with the color of the numbers. Your task is to decide whether the number you hear was the same as what you see. If the two numbers are the same, press the yellow button on the box, if they are different, press the red button. It is important that you answered correctly, but quickly."

A.1.3. Second part of instructions (non-perspective-dependent group)

"You (name of participant A) will not have to deal with the color of the numbers. Your task is to decide whether the number you hear was the same as what you see. If the two numbers are the same, press the yellow button on the box, if they are different, press the red button. You (name of participant B) will not have to deal with the numbers you hear, or the meaning of the numbers you see. Your task is to decide, whether the color of the character was the same as the color of the previous character. If the two colors are the same, press the yellow button on the box, if they are different, press the red button. It is important that you answered correctly, but quickly."

References

Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *NeuroImage, 30*, 1059–1068. http://dx.doi.org/10.1016/j.neuroimage.2005.10.026.

Apperly, I. A. (2008). Beyond simulation-theory and theory–theory: Why social cognitive neuroscience should use its own concepts to study "theory of mind". *Cognition, 107*, 266–283. http://dx.doi.org/10.1016/j.cognition.2007.07.019.

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

Back, E., & Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition, 115*(2010), 54–70. http://dx.doi.org/10.1016/j.cognition.2009.11.008.

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3–51). New York: Guilford Press.

Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon.

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

Carruthers, P. (2015a). Mindreading in adults: Evaluating two-systems views. *Synthese*. http://dx.doi.org/10.1007/s11229-015-0792-3.
Carruthers, P. (2015b). Two systems for mindreading? *Review of Philosophy and Psychology*. http://dx.doi.org/10.1007/s13164-015-0259-y.
Christensen, W., & Michael, J. (2015). From two systems to a multi-systems architecture for mindreading. *New Ideas in Psychology*, *40*(PartA), 48–64. http://dx.doi.org/10.1016/j.newideapsych.2015.01.003.
Cohen, M. A., Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in Cognitive Sciences, 16*, 411–417. http://dx.doi.org/10.1016/j.tics.2012.06.013.
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*, 99–103. http://dx.doi.org/10.1037//0012-1649.17.1.99.
German, T. C., & Cohen, A. S. (2012). A cue-based approach to 'theory of mind': Re-examining the notion of automaticity. *British Journal of Developmental Psychology, 30*, 45–58. http://dx.doi.org/10.1111/j.2044-835X.2011.02055.x.
Kampis, D., Somogyi, E., Itakura, S., & Király, I. (2013). Do infants bind mental states to agents? *Cognition, 129*, 232–240. http://dx.doi.org/10.1016/j.cognition.2013.07.004.
Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review, 4*, 501–506. http://dx.doi.org/10.3758/BF03214339.
Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.
Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception and Psychophysics, 68*, 327–337. http://dx.doi.org/10.3758/BF03193680.
Moll, H., & Meltzoff, A. N. (2011). How does it look? Level 2 perspective-taking at 36 months of age. *Child Development, 82*, 661–673. http://dx.doi.org/10.1111/j.1467-8624.2010.01571.x.
Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology, 24*, 603–613. http://dx.doi.org/10.1348/026151005X55370.
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*, 1931–1951. http://dx.doi.org/10.1080/17470218.2015.1023206.
Perner, J., Stummer, S., Sprung, M., & Doherty, M. (2002). Theory of mind finds its Piagetian *perspective*: Why alternative naming comes with understanding belief. *Cognitive Development, 17*, 1451–1472. http://dx.doi.org/10.1016/S0885-2014(02)00127-2.
Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition, 117*, 230–236. http://dx.doi.org/10.1016/j.cognition.2010.08.003.
Rakoczy, H. (2012). Do infants have a theory of mind? *British Journal of Developmental Psychology, 30*, 59–74. http://dx.doi.org/10.1111/j.2044-835X.2011.02061.x.
Samson, D., & Apperly, I. A. (2010). There is more to mind reading than having theory of mind concepts: New directions in theory of mind research. *Infant and Child Development, 19*, 443–454. http://dx.doi.org/10.1002/icd.678.
Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1255–1266. http://dx.doi.org/10.1037/a0018729.
Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance, 40*, 929–937. http://dx.doi.org/10.1037/a0035175.
Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences, 10*, 70–76. http://dx.doi.org/10.1177/0956797612439070.
Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition, 88*, B11–B21. http://dx.doi.org/10.1016/S0010-0277(03)00043-X.
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. http://dx.doi.org/10.1111/j.1467-7687.2007.00580.x.
Surtees, A. D. R., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development, 83*, 452–460. http://dx.doi.org/10.1111/j.1467-8624.2011.01730.x.
Surtees, A. D. R., Apperly, I. A., & Samson, D. (2013a). Similarities and differences in visual and spatial perspective-taking processes. *Cognition, 129*, 426–438. http://dx.doi.org/10.1016/j.cognition.2013.06.008.
Surtees, A. D. R., Apperly, I. A., & Samson, D. (2013b). The use of embodied self-rotation for visual and spatial perspective-taking. *Frontiers in Human Neuroscience, 7*, 1–12. http://dx.doi.org/10.3389/fnhum.2013.00698.
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. http://dx.doi.org/10.1111/j.2044-835X.2011.02063.x.
Teufel, C. et al (2009). Social cognition modulates the sensory coding of observed gaze direction. *Current Biology, 19*, 1274–1277. http://dx.doi.org/10.1016/j.cub.2009.05.069.
Wimmer, G., Hogrefe, J., & Perner, J. (1988). Children's understanding of informational access as source of knowledge. *Child Development, 59*, 386–396.