



University  
of Glasgow

Kessler, K., and Rutherford, H. (2010) *The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions.* Frontiers in Psychology, 1. ISSN 1664-1078

<http://eprints.gla.ac.uk/63639/>

Deposited on: 16<sup>th</sup> May 2012



# The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions

Klaus Kessler\* and Hannah Rutherford

Institute of Neuroscience and Psychology, Social Interaction Research Centre, University of Glasgow, Glasgow, UK

**Edited by:**

Anna M. Borghi, University of Bologna, Italy

**Reviewed by:**

Paul D. Siakaluk, University of Northern British Columbia, Canada  
Tad Brunye, US Army NSRDEC and Tufts University, USA

**\*Correspondence:**

Klaus Kessler, Department of Psychology, Room 610, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, UK.  
e-mail: k.kessler@psy.gla.ac.uk

We set out to distinguish level 1 (VPT-1) and level 2 (VPT-2) perspective taking with respect to the embodied nature of the underlying processes as well as to investigate their dependence or independence of response modality (motor vs. verbal). While VPT-1 reflects understanding of what lies within someone else's line of sight, VPT-2 involves mentally adopting someone else's spatial point of view. Perspective taking is a high-level conscious and deliberate mental transformation that is crucially placed at the convergence of perception, mental imagery, communication, and even theory of mind in the case of VPT-2. The differences between VPT-1 and VPT-2 mark a qualitative boundary between humans and apes, with the latter being capable of VPT-1 but not of VPT-2. However, our recent data showed that VPT-2 is best conceptualized as the deliberate simulation or emulation of a movement, thus underpinning its embodied origins. In the work presented here we compared VPT-2 to VPT-1 and found that VPT-1 is not at all, or very differently embodied. In a second experiment we replicated the qualitatively different patterns for VPT-1 and VPT-2 with verbal responses that employed spatial prepositions. We conclude that VPT-1 is the cognitive process that subserves verbal localizations using "in front" and "behind," while VPT-2 subserves "left" and "right" from a perspective other than the egocentric. We further conclude that both processes are grounded and situated, but only VPT-2 is embodied in the form of a deliberate movement simulation that increases in mental effort with distance and incongruent proprioception. The differences in cognitive effort predict differences in the use of the associated prepositions. Our findings, therefore, shed light on the situated, grounded and embodied basis of spatial localizations and on the psychology of their use.

**Keywords:** grounding, embodiment, movement simulation, perspective taking, social cognition, spatial language

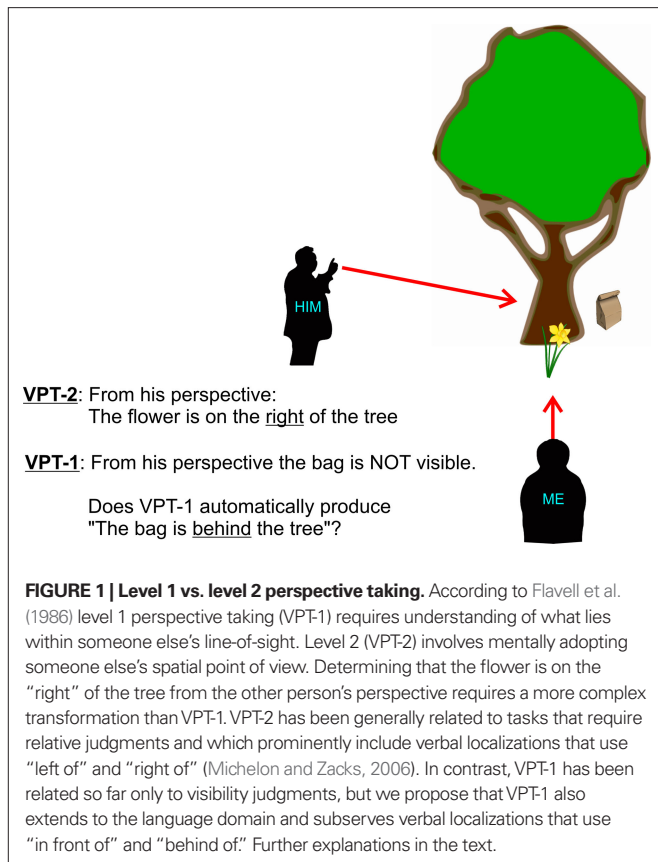
## INTRODUCTION

In this study we set out to investigate the differences between two forms of visuo-spatial perspective taking in terms of embodied processing and regarding the consequences for the situated and grounded use of projective spatial prepositions (i.e., "left," "right," "in front," and "behind").

Flavell et al. (1986) have categorized the ability to understand someone else's visuo-spatial perspective into level 1 and level 2 perspective taking (VPT). While level 1 (VPT-1) reflects understanding of *what* lies within someone else's line-of-sight, i.e., which objects are visible and which occluded ("I know what you can see;" see **Figure 1**), level 2 (VPT-2) involves mentally adopting someone else's spatial point of view and understanding *how* the world is represented from this virtual perspective (i.e., "I see the world through your eyes") as shown in **Figure 1**. As another example please imagine we would like to tell a friend that she has an eyelash on her left cheek, which would require determining "left" and "right" from our friend's perspective – independently from our own point of view. Or think of way descriptions, where an instruction like "in front of the building turn left" assumes that the instructing and the instructed persons are aligned into the same virtual perspective, i.e., that they both either mentally face the entrance from the outside or imagine coming out of the building.

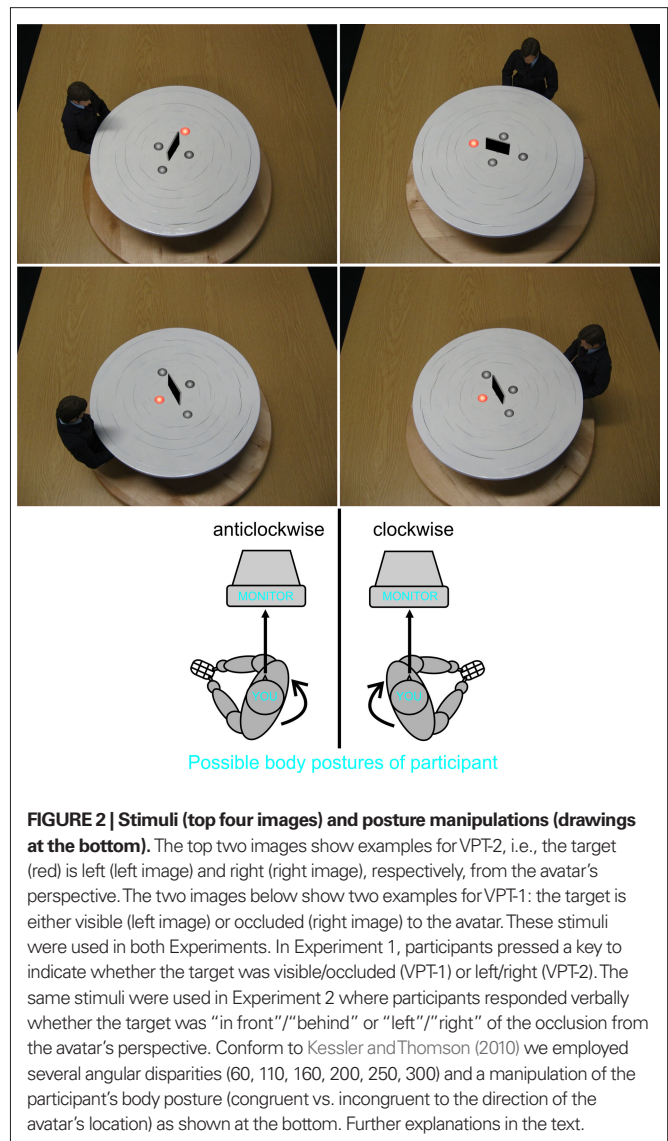
These examples point out the importance of VPT in communication, e.g., for establishing a common reference frame for understanding spatial localizations or more generally for establishing a shared view of the world (Frith and Frith, 2007). VPT-2 is regarded as the more complex process of the two, which is evidenced by a later ontogenetic development, specific difficulties experienced by autistic children, and by phylogenetic differences. VPT-1 develops around the age of 2 years and autistic children do not experience particular difficulties with this task (Leslie and Frith, 1988; Baron-Cohen, 1989). In contrast, VPT-2 develops around 4–5 years (Gzesh and Surber, 1985; Hamilton et al., 2009), but not in children diagnosed with autism spectrum disorder (ASD), and VPT-2 performance is predicted by theory of mind (ToM) score (Hamilton et al., 2009). Primates seem capable of certain forms of VPT-1 but not at all of VPT-2 (Tomasello et al., 2005). The latter conforms to their inability to perform simple ToM tasks (Call and Tomasello, 1999), which pose no problem for 5-year-old (non-autistic) children.

However, primates (Tomasello et al., 1998; Brauer et al., 2005) and other species (Scheumann and Call, 2004; Pack and Herman, 2006) have been reported to physically align their perspective with humans. Apes even deliberately change their position to be able to look around obstacles and share what a human experimenter can see (Tomasello et al., 1998; Brauer et al., 2005). This reflects



the basic understanding that a physical (apes) or mental effort (humans, i.e., VPT-2) is sometimes necessary in order to understand someone else's view of the world (Frith and Frith, 2007), which led us to hypothesize that VPT-2 might have originated from deliberate physical alignment of perspectives exhibited by apes (Kessler and Thomson, 2010).

We reasoned that if this was the case then VPT-2 would still be an "embodied" process in the sense that it relies on the posture and action repertoire of the body. Our recent findings (Kessler and Thomson, 2010) have indeed confirmed that although VPT-2 is a high-level cognitive process it is not a purely abstract transformation of a reference frame or coordinate system as had been the established view within linguistics/computational linguistics (Retz-Schmidt, 1988). Instead, we found substantial evidence that VPT-2 is relying on action-related and proprioceptive representations of the body. Specifically, we altered the body posture of the participants before each trial (cf. **Figure 2**) so that their body was either congruent or incongruent (in some experiments also neutral) with the direction of VPT-2. This simple manipulation had a dramatic effect on reaction times, where a congruent posture speeded up processing while an incongruent slowed it down. Based on the pattern of results across four experiments, we concluded that the embodiment of VPT-2 is best conceptualized as the deliberate emulation or simulation of a body rotation, supporting the notion of endogenous sensorimotor embodiment (Kessler and Thomson, 2010). This conforms to Wilson's sixth and most powerful meaning of embodied cognition: "6. Off-line cognition is body based. Even when decoupled from the environment, the activity of the



mind is grounded in mechanisms that evolved for interaction with the environment – that is, mechanisms of sensory processing and motor control" (Wilson, 2002, p. 626).

In the experiments presented here we compared VPT-1 and VPT-2 in terms of their embodiment. While we expected to replicate the evidence for movement simulation/emulation subserving VPT-2 we did not know whether a similar process would also subserve VPT-1. In fact, the onto- and phylogenetic differences we have mentioned above suggested that the underlying processes could differ quite substantially and Michelon and Zacks (2006) provided conclusive evidence for a qualitative difference between VPT-1 and VPT-2: Congruent to our previous findings (Kessler and Thomson, 2010) and to results reported by others (Huttenlocher and Presson, 1973; Levine et al., 1982) VPT-2 showed an increase in reaction times with increasing angular disparity, while reaction times in the VPT-1 task remained flat across angles.

Michelon and Zacks (2006) concluded that VPT-1 is based on imagining the other person's line-of-sight which determines the relevant inter-object spatial relations, while VPT-2 requires some sort of mental rotation. A very similar distinction was suggested by

Kessler (2000) based on a connectionist network model for processing spatial prepositions: While a mental self-rotation (i.e., VPT-2) is necessary for understanding “left” and “right” from a different perspective, “in front” and “behind” from the same perspective are solved based on between-object relations that “compute” the line-of-sight. For the latter imagine someone telling you “the bag is behind the tree.” For determining “behind” it is only necessary to draw a line between that person and the tree and the bag would be on the side of the tree that is occluded from the person’s view (cf. **Figure 1**). Our expectation therefore was that VPT-1 subserves “in front” and “behind” judgments and that VPT-1 does not rely on the simulation of a body movement. Hence, VPT-1 would either not be “embodied” at all or very differently “embodied” than VPT-2.

Accordingly, we also expected VPT-1 and VPT-2 to rely on different neural substrates, although we could not directly test this hypothesis in our behavioral experiments. VPT-2 could either be a form of action simulation that involves action control areas in the posterior frontal cortex together with body schema representations in the parietal lobe, or VPT-2 could be a form of action emulation, where the perceptive and proprioceptive outcomes of the transformation are generated without the need for a full movement simulation that instantiates all the intermediate steps “to get there.” The distinction between emulation and simulation is rather gradual in this context (and “simulation” will be used throughout the document), but in an extreme scenario emulation might not involve action control areas at all, while essentially relying on transformations within body schema and other proprioceptive areas. The body schema would be involved in any case, which is indeed supported by a growing number of findings where the temporo-parietal junction was identified as an essential substrate for VPT-2 (Zacks and Michelon, 2005; Arzy et al., 2006; Keehner et al., 2006).

In contrast, if VPT-1 solely relies on understanding spatial relations between a person and at least two objects, then the primary substrate of VPT-1 should be the dorsal between-object system in parietal cortex (e.g., Goodale and Milner, 1992; Ungerleider and Haxby, 1994). To our knowledge no neuroimaging data are yet available for VPT-1, hence, the behavioral results presented here will provide a first hint for whether motor simulation/emulation networks are a likely or unlikely neural substrate for VPT-1.

Kessler (2000) argued based on his connectionist model that qualitatively different processes are employed for the two dimensions of projective spatial prepositions (“in front”/“behind” vs. “left”/“right”), and explicitly related these processes to the different neural substrates mentioned above (i.e., to the between-object system and to the motor simulation/emulation systems, respectively). In psycholinguistics there have been suggestions to relate the use of “in front” and “behind” to the line-of-sight (Grabowski, 1999; Grabowski and Miller, 2000; Kessler, 2000), yet, to our knowledge no explicit link has been established so far to VPT-1 as the underlying cognitive mechanism. In the model by Kessler (2000) all spatial dimensioning starts with the extraction of a so-called “anchor”-direction, which is basically the line-of-sight of that perspective and which automatically produces the “in front” pole in relation to the relatum (the reference object, i.e., the tree in **Figure 1**). “Behind” requires an additional processing step for determining the opposite direction to “in front.” In agreement with Grabowski (1999) this predicts faster production times for “in front” compared to

“behind,” which has been shown to be the case (Herrmann et al., 1987; Bryant et al., 1992). Furthermore, if the assumption is correct that VPT-1 is the underlying process, then response times for “visible” judgments should be faster than for “occluded.”

Our first two steps to investigate the proposed link between the two horizontal dimensions of projective prepositions and the two levels of VPT were as follows. Firstly, we aimed to show that VPT-1 and VPT-2 are indeed differently embodied cognitive processes. Secondly, we aimed to replicate our effects using a verbal response (i.e., participants saying “left,” “right,” “in front,” or “behind”) in order to test our assumption that “left”/“right” is subserved by VPT-2 whereas “in front”/“behind” by VPT-1. So far we had exclusively employed spatially mapped key presses for VPT-2 (Kessler and Thomson, 2010). We expected that the embodiment effect observed in these experiments would persist as we believed it to be a defining characteristic of VPT-2 that did not depend on the response modality. However, a result to the contrary would be important, forcing us to adjust our theoretical considerations. It would mean that only in the case of a spatially mapped motor response a body rotation is fully simulated (cf. Kessler and Thomson, 2010), while in the case of a verbal response VPT-2 could rely on the transformation of a more abstract “disembodied” (e.g., geometric) representation of the egocentric perspective (e.g., Mahon and Caramazza, 2008). Qualitative differences in spatial representation updating after a physical or an imagined self-rotation have indeed been reported for motor (pointing) vs. verbal (“left”/“right”) responses (de Vega and Rodrigo, 2001). In our case our primary interest was whether the effect of the participant’s body posture on processing time depends on response modality or not.

To this end we conducted two experiments. The first experiment employed a motor response to indicate the target’s location, conform to Michelon and Zacks’ as well as Kessler and Thomson’s procedures. This ensures comparability to previously reported results and allows for an optimal first comparison between VPT-1 and VPT-2 with respect to embodiment. The second experiment, however, employed a voice key to measure response time and to record verbal localizations of the target by means of the prepositions of interest. This would reveal whether our previous findings (Kessler and Thomson, 2010) regarding the embodiment of VPT-2 as well as our potential new findings from Experiment 1 would actually generalize to the language domain.

## EXPERIMENT 1

The goal of this experiment was to investigate the differences between VPT-1 and VPT-2. Firstly, we aimed at replicating the difference reported by Michelon and Zacks (2006) regarding the dependence on angular disparity: While reaction times (RTs) strongly increased with angle for VPT-2, RTs for VPT-1 remained constant across angles. This predicted an interaction between task (VPT-1 vs. VPT-2) and angular disparity (60°, 110°, 160°). Secondly, we wanted to replicate our previous results (Kessler and Thomson, 2010) where we found a strong body posture effect for VPT-2 (congruent posture faster than incongruent) and compare this pattern of results to the one obtained for VPT-1 with the identical posture manipulation. We expected that VPT-1 would not be embodied in the same way as VPT-2. Hence, we also expected an interaction between task and body posture (congruent vs. incongruent).

## MATERIALS AND METHODS

### Participants

All procedures were in concordance with the declaration of Helsinki and approved by the local ethics committee. Participants were volunteers, right-handed, had normal or corrected-to-normal vision, were naive with respect to the purpose of the study, and received payment or course credit for participation. Sixteen females and 8 males took part in Experiment 1. Mean age was 23.1 years. One female participant had to be excluded due to excessive response times in the VPT-2 task (more than 3 standard deviations away from the sample mean). According to her self-report she had difficulties in general to determine left and right, even from her own (egocentric) perspective.

### Stimuli and design

We employed a VPT-2 task congruent to the one originally employed by Kessler and Thomson (2010) revealing the embodied nature of VPT-2, and we added a VPT-1 task. The stimuli are shown in **Figure 2**, where an avatar was seated at one of six possible angular disparities (60°, 110°, 160° clockwise and anticlockwise) around a table. Pictures were taken from a vertical angle of 65°. Stimuli were colored bitmaps with a resolution of 1024 by 768 pixels corresponding to the graphic card settings during the experiment. Viewing distance was 100 cm and a chin rest was employed to ensure constancy.

In the center of the stimulus-table four gray hemispheres (potential targets) were arranged around an occluder (**Figure 2**). On each trial one of the hemispheres turned red indicating its status as the target. Hence, from the avatar's perspective the target on a given trial could be left, right (VPT-2), visible, or occluded (VPT-1). In this first experiment participants either pressed a key to indicate the target's left/right location from the avatar's point of view, or they pressed a key from another set of two to indicate whether the target was visible or occluded to the avatar. Due to the different pairs of response keys for the two tasks (VPT-1 vs. VPT-2), tasks were blocked into miniblocks of 24 trials each, to allow for optimal response preparation within each block (index fingers above the appropriate pair of keys).

Conform to Kessler and Thomson (2010) we also varied the body posture of the participants randomly across trials (**Figure 2**). The body in relation to the head/gaze direction could be turned clockwise or anticlockwise, hence, being either congruent or incongruent with respect to the direction of the avatar's sitting position (at either clockwise or anticlockwise angular disparities of 60°, 110°, 160°). Participants moved the response device (Targus® wireless number keypad) together with their body, while their head remained on the chin rest gazing ahead. Markings on the table indicated exactly where to place the numberpad each way to ensure a constant angle of  $\pm 60^\circ$  (clockwise/anticlockwise) between body and gaze direction across trials. We administered a total of 336 trials with 28 trials in each cell of the  $2 \times 3 \times 2$  design consisting of the factors "task" (VPT-1 or VPT-2), "angular disparity" (60°, 110°, or 160°; collapsed across clockwise and anticlockwise disparities), and "body posture" (congruent or incongruent to the clockwise or anticlockwise direction of the avatar's location). In separate analyses for VPT-1 and VPT-2 we also included the factor "response" (visible vs. occluded and left vs. right, respectively). Our dependent variable

was response time (RT) of correct responses only and we employed individual medians for each condition and participant to reduce distortions by outlier RTs. Error rates are reported in **Table 1**.

### Procedure

E-Prime® 2.0 was used for experimental control. The two tasks (VPT-1 vs. VPT-2) were presented in alternating miniblocks of 24 trials each (14 miniblocks in total; initial task balanced across participants). Every block started with an instruction about the given task. For VPT-2 participants were instructed to press the "4" key (colored yellow) with their left index for "left" and the "6" key (colored blue) with their right index for "right" on the wireless numberpad. For VPT-1 participants were instructed to press the "0" key (colored green) with their left index to indicate "visible" and the "." key (colored brown) with their right index to indicate "occluded." We did not choose a vertical key alignment in the latter case in order to avoid interference or congruence effects between target location and key location at 160° angular disparity. For instance if one would choose the top key for "occluded" responses and the bottom key for "visible" responses, then the key alignment and the target locations could mismatch at 160° (the "visible" target is above the "occluded" target, cf. **Figure 2**). In essence our response key mapping corresponded to the mapping employed by Michelon and Zacks (2006).

Every trial started with a picture displaying the posture instruction (cf. **Figure 2B**). When participants had assumed the correct posture they pressed both response buttons to proceed. A fixation cross was then shown for 500 ms and was automatically replaced by the experimental stimulus. Participants were instructed to respond as quickly and as accurately as possible. Twelve practice trials were administered in form of a miniblock of 6 VPT-2 trials and a miniblock of 6 VPT-1 trials.

## RESULTS

In our first analysis we compared the two tasks (VPT-1 vs. VPT-2) together with the factors "body posture" (congruent vs. incongruent) and "angular disparity" (60°, 110°, 160°). We also conducted two separate analyses for each task, where we included the two possible responses as an additional factor: "left" vs. "right" for VPT-2 and "visible" vs. "occluded" for VPT-1. This allowed us to test for asymmetries within each task. For instance all participants were right-handed; it could therefore be that "right"-responses that

**Table 1 | Mean error rates per condition for both experiments (a value of 1.00 would mean that one mistake was committed on average in a particular condition).**

	60		110		160		Sum
	Cong	Incong	Cong	Incong	Cong	Incong	
<b>EXPERIMENT 1</b>							
VPT-1	0.87	0.70	0.91	0.57	0.78	0.39	4.22
VPT-2	0.70	0.57	0.26	0.83	0.83	1.13	4.30
<b>EXPERIMENT 2</b>							
VPT-1	0.00	0.13	0.04	0.00	0.04	0.00	0.22
VPT-2	0.00	0.17	0.30	0.26	0.39	0.35	1.48

required a keypress with the right index were faster than “left”-responses with the left index. With respect to VPT-1 we expected faster responses to visible targets than to occluded ones –due to fewer inter-object-relations required for processing. That is, determining visibility requires only the direct relation between the avatar’s line-of-sight and the target, while determining occlusion also requires the occluder to be processed, involving three objects and their inter-relations (see Introduction).

Prior to each of these three multifactorial analyses we conducted Mauchly’s sphericity tests and whenever the sphericity assumption was violated ( $p < 0.05$ ) we conducted multivariate analyses of variance (MANOVA) conform to recommendations in the literature, as this test does not assume sphericity (Davidson, 1972; O’Brien and Kaiser, 1985; Vasey and Thayer, 1987). Error Rates are shown in **Table 1** and were low and inconsistent with a general speed-accuracy-trade-off.

### Combined analysis for VPT-1 and VPT-2

Sphericity was violated, so a  $2 \times 3 \times 2$  MANOVA was employed with “task,” “angular disparity,” and “body posture” as factors. The analysis revealed significant main effects of all three factors ( $p < 0.01$ ), but most importantly the expected interactions between task and angular disparity ( $F(2,21) = 26.1, p < 0.00001, \eta_p^2 = 0.658$ ) as well as between task and body posture ( $F(1,22) = 30.7, p < 0.00001, \eta_p^2 = 0.583$ ) reached significance and strong effect sizes (compare **Figure 3A**). These interaction effects will be analyzed further in relation to each task separately in the next two sub-sections (Analysis for VPT-2 and Analysis for VPT-1).

### Analysis for VPT-2

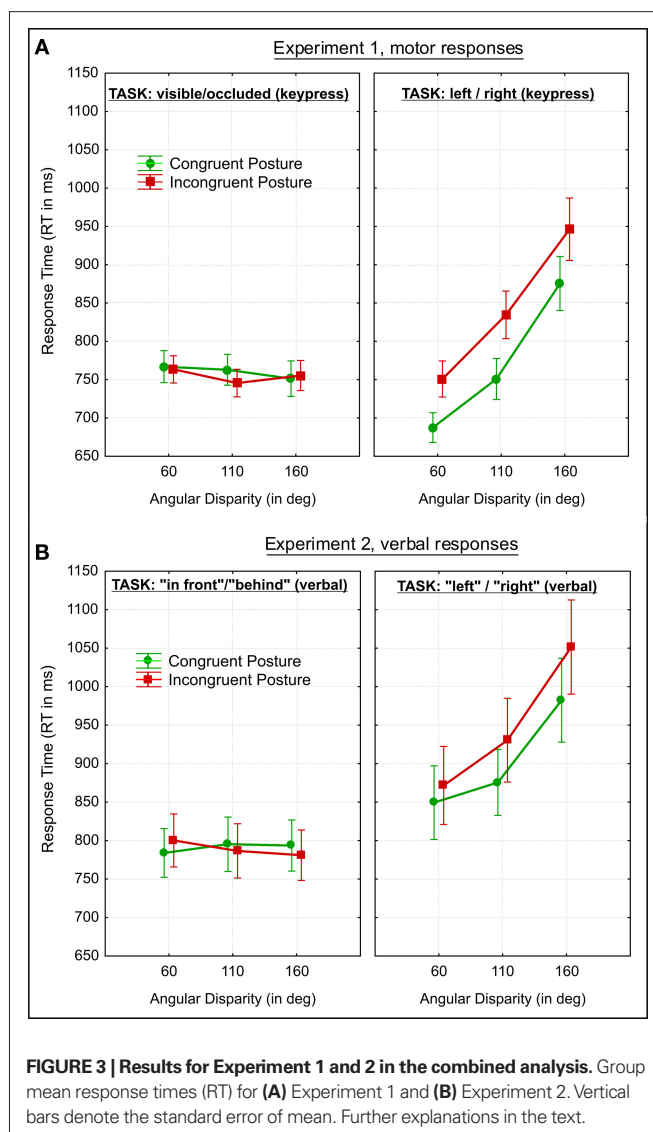
Sphericity was violated, so we employed a  $3 \times 2 \times 2$  MANOVA with the factors “angular disparity,” “body posture,” and “response” (left vs. right). The analysis revealed significant main effects of angle ( $F(2,21) = 25.6, p < 0.00001, \eta_p^2 = 0.661$ ) and of body posture ( $F(1,22) = 33.4, p < 0.00001, \eta_p^2 = 0.603$ ). Although all participants were right-handed we did not find significantly faster responses with the right index ( $p = 0.62$ ). Somewhat in contrast to our previous results (Kessler and Thomson, 2010) we did not find a significant interaction between angle and body posture either ( $p = 0.72$ ). As can be seen in **Figure 3A** (right graph) this is due a strong body posture effect at all angular disparities, even at the lowest of  $60^\circ$  (Newman–Keuls test  $p = 0.0003$ ).

### Analysis for VPT-1

Sphericity was not violated, so we employed a  $3 \times 2 \times 2$  ANOVA with the factors “angular disparity,” “body posture,” and “response” (visible vs. occluded). The only significant model term was the main effect of response ( $F(1,22) = 7.1, p = 0.014, \eta_p^2 = 0.245$ ), with “visible” being consistently faster than “occluded” judgments as shown in **Figure 4** (left graph) and conform to our predictions (Grabowski, 1999; Kessler, 2000; see Introduction).

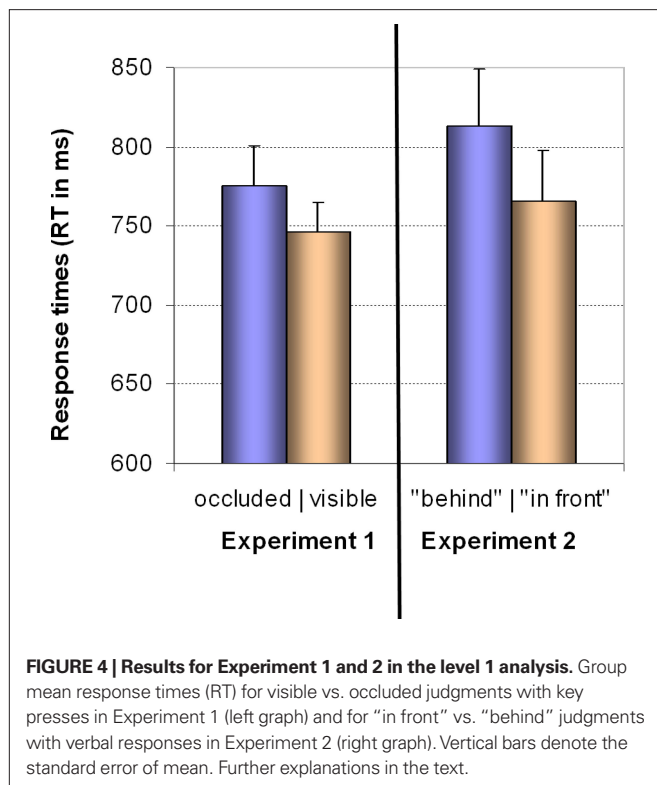
## DISCUSSION OF EXPERIMENT 1

We expected VPT-1 and VPT-2 to be qualitatively different processes, and we expected this to be reflected by distinct response patterns in relation to angular disparities as well as in relation to body posture. This was confirmed by the two significant interactions (task  $\times$  angle;



task  $\times$  body posture) in the combined analysis. Furthermore we did not find an asymmetry between spatially mapped left and right motor responses (VPT-2) while we did find that responses to visible targets were faster than to occluded targets (VPT-1). We replicated Michelon and Zacks’ (2006) findings showing that response times for VPT-2 increased with angular disparity, whereas response times for VPT-1 were not affected by angular disparity and remained constant. Interestingly, we found faster responses for visible than for occluded targets conform to our hypothesis which was extrapolated from asymmetries reported for “in front”/“behind” judgments (Herrmann et al., 1987; Bryant et al., 1992). We are not aware of any previous reports of such a finding regarding visibility judgments. This finding is a first hint that VPT-1 might indeed subservise “in front” and “behind” localizations.

We have also replicated and extended our previous VPT-2 findings (Kessler and Thomson, 2010) regarding the effect of the participants’ body posture by observing significantly faster RTs with a congruent than an incongruent posture at all angular disparities, even at  $60^\circ$ . We have confirmed that VPT-2 is embodied, in concordance with



our previous conclusion that it is the endogenous simulation of a body rotation (Kessler and Thomson, 2010). The novel finding in Experiment 1 was that VPT-1 was not affected by body posture and was therefore not at all embodied, or very differently than VPT-2. In Experiment 2 we set out to generalize these findings and conclusions to the language domain for establishing a direct link between the two levels of VPT and the two horizontal dimensions of projective prepositions ("in front"/"behind" vs. "left"/"right").

## EXPERIMENT 2

### MATERIALS AND METHODS

#### Participants

Fifteen female and 9 male volunteers, English native speakers, right-handed, with normal or corrected-to-normal vision, who were naive with respect to the purpose of the study, and who received payment or course credit took part in Experiment 2. Mean age was 22.3 years. One female participant had to be excluded from data analysis due to excessively slow response times. After the experiment the participant disclosed that she was diagnosed with a "slow processing" expression of dyslexia.

#### Stimuli, design, and procedure

Exactly the same stimuli and design were employed as in Experiment 1. Also the procedure was largely the same with the posture instruction at the beginning of each trial (cf. Figure 2), the 500 ms fixation cross, and the subsequent presentation of the experimental stimulus (cf. Figure 2). The main and essential difference was that we used a voice key (Logitech® headset in combination with DMDX software version 4 <http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm>) that provided us with voice onset times (RTs) as

well as with a recording of what had been said on a given trial. The voice onset threshold was tuned for each participant individually and we verified/adjusted the voice onset time on every single trial before proceeding with the analysis. Specifically, the recordings enabled us to determine for each response whether it had been correct and whether the voice key had been indeed triggered by the onset of the verbal response and not by any other acoustic event (e.g., smacking of lips) or if the voice onset had been missed by the voice key. If the automatic voice onset detection had been incorrect we re-measured the voice onset time with audio editing software (Music Editor v8.2.5). This verification procedure was essential as the initial phonemes varied across the four prepositions. To further validate our procedure for extracting the response times (RTs) we sampled 184 response audio files (46 files per preposition, 2 from each participant) from the total pool of responses and asked 8 independent Raters to determine the voice onset times in each of these 184 response files. For each Rater the difference between their RT for a given file and our manually adjusted voice key RT was calculated. We then calculated average differences for each preposition to see whether there was a specific bias in our data that could have affected the outcomes (e.g., "right" being systematically delayed compared to "left"). The average differences for each preposition were as follows (with minimum and maximum across Raters in brackets): behind = -4.79 ms (min. = -12.22; max. = 15.48), in front = -1.78 ms (min. = -12.39; max. = 15.78), left = -5.68 ms (min. = -14.35; max. = 15.26), right = -2.68 ms (min. = -11.07; max. = 22.83). None of the Raters revealed specific distortions for a particular preposition. This validation reveals that our procedure worked well, but also, that it was not possible to determine the voice onset with millisecond accuracy, which we have to take into consideration when interpreting our results. Most importantly for the interpretation of our data, no bias was observed for the RTs of a particular preposition to be systematically over- or underestimated in relation to the others to an extent that could explain our effects.

Error rates were particularly low in this Experiment (see Table 1). We analyzed RTs only for the correct responses and we employed individual medians for each condition to reduce distortions by outlier RTs. The two tasks ("in front"/"behind" vs. "left"/"right") were presented again in alternating miniblocks of 24 trials each for maximum comparability to Experiment 1. At the beginning of each miniblock participants were instructed about the two verbal alternatives they were expected to use: "left"/"right" or "in front"/"behind." Again two miniblocks of 6 trials each were administered for practising VPT-2 and VPT-1, respectively, before the 336 experimental trials were presented.

## RESULTS

As for Experiment 1 we conducted three analyses. In the first we compared the two tasks ("in front"/"behind" vs. "left"/"right") together with the factors "body posture" (congruent vs. incongruent) and "angular disparity" (60°, 110°, 160°). We then conducted two separate analyses for each task, where we included the two possible responses as an additional factor: "left" vs. "right" and "in front" vs. "behind," respectively. This allowed us to test for asymmetries between the poles of each dimension separately. Error Rates were particularly low and are shown in Table 1.

### Combined analysis

Sphericity was violated, so we employed a  $2 \times 3 \times 2$  MANOVA with “task,” “angular disparity,” and “body posture” as factors. As in Experiment 1 the analysis revealed significant main effects of all three factors ( $p < 0.01$ ), and the expected interactions between task and angular disparity ( $F(2,21) = 18.1, p < 0.0001, \eta_p^2 = 0.595$ ) as well as between task and body posture ( $F(1,22) = 21.33, p < 0.001, \eta_p^2 = 0.492$ ). In contrast to Experiment 1 also the three-way interaction between task, angular disparity, and body posture reached significance ( $F(1,21) = 5.9, p < 0.01, \eta_p^2 = 0.197$ ). The latter was mainly due to a gradually increasing posture effect with angular disparity for “left”/“right”: The body posture effect reached significance at  $160^\circ$  and  $110^\circ$  (Newman–Keuls test  $p < 0.001$ ) but not quite at  $60^\circ$  ( $p = 0.07$ ). “In front”/“behind,” on the other hand, revealed a strikingly similar pattern to Experiment 1 with no RT increase with angular disparity and no posture effect (Figure 3, left graphs). The interaction effects will be analyzed further in relation to each task separately in the next two sub-sections (Analysis for “Left”/“Right” and Analysis for “in Front”/“Behind”).

### Analysis for “left”/“right”

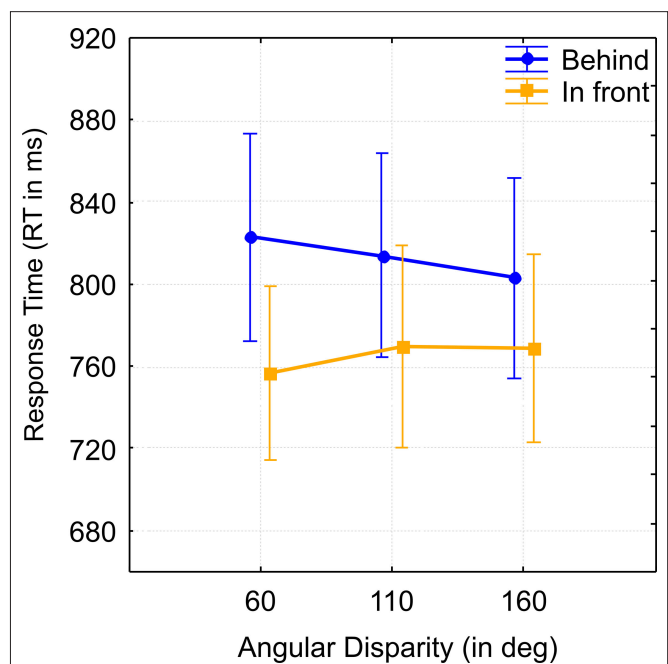
Sphericity was violated, so we employed a  $3 \times 2 \times 2$  MANOVA with the factors “angular disparity,” “body posture,” and “response” (“left” vs. “right”). The analysis revealed significant main effects of angle ( $F(2,21) = 22.1, p < 0.0001, \eta_p^2 = 0.619$ ) and of body posture ( $F(1,22) = 19.1, p < 0.001, \eta_p^2 = 0.498$ ). Conform to Experiment 1 we did not observe significantly faster responses for either “left” or “right” responses ( $p = 0.14$ ). However, in Experiment 2 RTs were determined based on vocal responses and although our procedures for determining voice onsets worked well on average (see Stimuli, Design, and Procedure), increased variability could have masked a true difference between left and right.

### Analysis for “in front”/“behind”

Sphericity was not violated, so we employed a  $3 \times 2 \times 2$  ANOVA with the factors “angular disparity,” “body posture,” and “response” (“in front” vs. “behind”). The main effect of response reached significance ( $F(1,22) = 16.7, p < 0.001, \eta_p^2 = 0.432$ ), as shown in Figure 4, right part of the graph. The interaction between response and angle also reached significance ( $F(1,22) = 3.5, p < 0.05, \eta_p^2 = 0.137$ ) suggesting that the difference between “in front” and “behind” judgments was consistent but decreased across angular disparities (see Figure 5). “In front” judgments were overall faster than “behind” judgments conform to the findings in Experiment 1 for VPT-1 by means of key presses, with “visible” being faster than “occluded” judgments (Figure 4, left graph).

## DISCUSSION OF EXPERIMENT 2

RTs for “in front” were consistently faster than for “behind” which corresponded to the asymmetry between “visible” and “occluded” responses in Experiment 1. Angular disparity also had an effect in form of an interaction but very differently compared to the VPT-2 tasks in both Experiments. Here RTs for “behind” decreased with angle while RTs for “in front” increased. This may be explained by taking into consideration that at  $60^\circ$  targets that are “in front” from the avatar’s perspective are also closer to the observer than targets that are “behind,” which are further away. This is reversed at



**FIGURE 5 | Results Experiment 2 in the level 1 analysis.** Group mean response times (RT) for “in front” vs. “behind” judgments in Experiment 2 at each angular disparity, which reflects the significant interaction between angular disparity and preposition. Vertical bars denote the standard error of mean. Further explanations in the text.

$160^\circ$  where targets that are “behind” are closer to the observer than targets that are “in front.” This seems to suggest that the observer’s fixed viewpoint plays a role in determining the inter-object relationships that lead to “in front”/“behind” judgments. This further underpins the qualitative difference to “left”/“right” judgments, where the observer’s viewpoint is mentally shifted.

In concordance with Experiment 1 body posture did not modulate RTs for “in front”/“behind” judgments. In total we observed a strikingly similar pattern between “in front”/“behind” and visible/occluded judgments (Figure 3 left column, Figure 4), which conforms to our prediction that VPT-1 subserves both types of judgments and is independent of response modality.

Verbal “left”/“right” judgments were significantly influenced by angular disparity conform to Experiment 1, our previous findings (Kessler and Thomson, 2010) and to reports by others (Kozhevnikov and Hegarty, 2001; Hegarty and Waller, 2004; Zacks and Michelon, 2005; Kozhevnikov et al., 2006). Most importantly, this task was also strongly affected by body posture. The pattern here also revealed a significant interaction between angle and posture conform to our previous findings (Kessler and Thomson, 2010).

We conclude that we accomplished our goal to generalize the data pattern obtained for the two VPT tasks with key presses (motor response) in Experiment 1 to verbal responses using spatial prepositions in Experiment 2. However, if one visually compares the RT patterns across Experiments (Figure 3) then it seems that RTs for VPT-2 (“left”/“right”) are generally increased in Experiment 2 compared to Experiment 1 (Figures 3A,B, right graphs). In order to statistically substantiate this observation we conducted a direct comparison between the two Experiments.



## COMPARISON BETWEEN EXPERIMENT 1 AND 2

We compared the two Experiments by means of a MANOVA (Sphericity was violated) employing the within subjects factors “task,” “angular disparity,” “body posture” and the between factor “experiment.” The results reflected the experiment-specific findings with overall significant effects of angular disparity ( $F(2,43) = 46.3$ ,  $p < 0.00001$ ,  $\eta_p^2 = 0.635$ ), body posture ( $F(1,44) = 25.8$ ,  $p < 0.00001$ ,  $\eta_p^2 = 0.369$ ), task  $\times$  angle ( $F(2,43) = 44.8$ ,  $p < 0.00001$ ,  $\eta_p^2 = 0.626$ ), task  $\times$  posture ( $F(1,44) = 51.7$ ,  $p < 0.00001$ ,  $\eta_p^2 = 0.540$ ), and task  $\times$  angle  $\times$  posture ( $F(2,43) = 4.5$ ,  $p < 0.02$ ,  $\eta_p^2 = 0.072$ ). The major difference between the two experiments was reflected in a significant interaction between task and experiment ( $F(1,44) = 4.9$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.101$ ) shown in **Figure 6**. The interaction was due to a larger difference in overall response times for VPT-2 between the two Experiments (Experiment 2 slower than Experiment 1) than for VPT-1.

We also compared each task (VPT-1 and VPT-2) separately between the two Experiments but did not find any significant effect of “experiment.” It is, however, noteworthy that in the VPT-1 analysis the interaction between response (“in front” + “visible” vs. “behind” + “occluded”) and angular disparity reached significance across both experiments ( $F(2,88) = 6.2$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.124$ ), suggesting that the corresponding interaction reported for Experiment 2 (see **Figure 5**) might generalize across modalities.

We compared Error Rates (ER) between the two Experiments and the two tasks in a  $2 \times 2$  ANOVA with “task” as within and “experiment” as between subjects factors (as shown in **Table 1**, some conditions in Experiment 2 did not reveal any variance at all, so the full design could not be employed). The main effect

of “experiment” reached significance ( $F(1,44) = 20$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.313$ ) suggesting that error rates were lower with verbal responses (cf. **Table 1**).

Overall the comparison between Experiments further corroborated our conclusion that VPT-1 and VPT-2 were employed independently of response modality, yet, in addition we found an asymmetry between VPT-1 and VPT-2 across modalities that indicates that overall RTs are more strongly increased for VPT-2 than for VPT-1, when responses were made verbally (**Figure 6**). Error rate differences only reflected a main effect of experiment with verbal responses being more accurate, so the observed interaction with RTs is not likely to reflect a general speed-accuracy-trade-off.

## GENERAL DISCUSSION

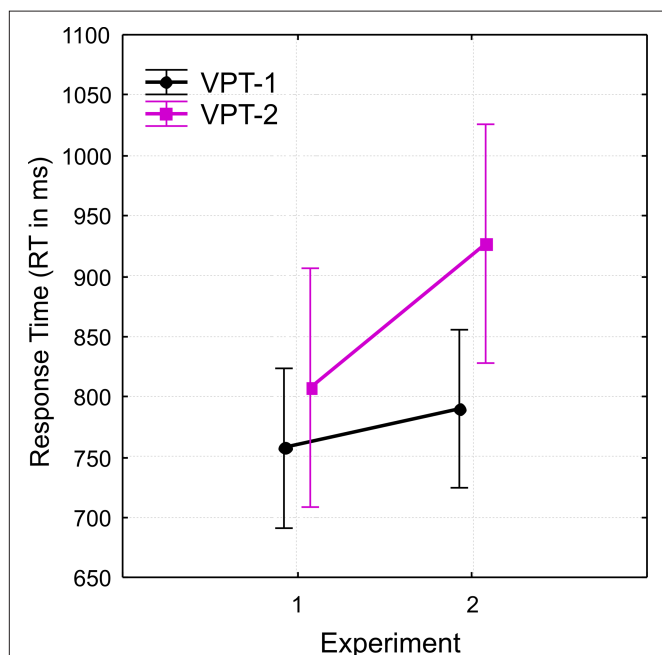
We found revealing commonalities and differences across the two experiments. Disregarding response modality (motor vs. verbal), within each experiment we found a qualitatively distinct pattern between the two tasks, while the pattern was remarkably similar for each task across the two experiments (i.e., modalities). It is important to emphasize that these results confirm our hypothesis that the distinct patterns we observed for VPT-1 and VPT-2 regarding angular disparity and embodiment do not depend on response modality, which contrasts with some findings reported in the context of spatial updating (de Vega and Rodrigo, 2001). In the following we discuss the implications of our results in detail.

### VPT-1 AND “IN FRONT”/“BEHIND”

One of our most striking results was that the response time patterns for visible/occluded and for “in front”/“behind” judgments were almost identical (cf. **Figure 3**, left column, **Figure 4**), considering that different individuals participated in the two experiments, that the task instructions differed (visibility vs. spatial location judgments), and that different response modalities were employed. The most prominent features in common were the absence of an increasing cost across angular disparities as well as the absence of an embodiment effect, i.e., of the participants’ body posture (congruent vs. incongruent). These strong similarities support our hypothesis that identical processes subserve the two types of judgments, yet, our final piece of evidence is also the most convincing: In both experiments those trials, where the target was located closer to the avatar than the occluder, i.e., requiring “visible” or “in front” responses, respectively, were processed consistently faster than trials where the occluder was between the avatar and the target, i.e. in “occluded” or “behind” trials. In total we believe that this provides conclusive evidence that in scenarios where the spatial prepositions “in front” and “behind” have to be determined from the perspective of another person, level 1 perspective taking (VPT-1) is the employed cognitive process. Accordingly, our findings close a gap between cognitive and developmental psychology on the one hand and psycholinguistics on the other.

### VPT-2 AND “LEFT”/“RIGHT”

With respect to VPT-2 we were able to replicate the embodiment effect in form of posture congruence effect with verbal responses, which we had previously observed with motor responses. This confirms that VPT-2 subserves “left”/“right” judgments (e.g., Michelon



**FIGURE 6 | Comparing the two Experiments.** Group mean response times (RT) for VPT-1 vs. VPT-2 and for each Experiment (1 vs. 2) reflecting the significant interaction between “experiment” and “task.” Vertical bars denote the standard error of mean. Further explanations in the text.

and Zacks, 2006) and that disregarding response modality the default strategy consists of mentally simulating a body rotation (cf. Kessler and Thomson, 2010).

However, **Figure 6** also shows a systematic difference between Experiments 1 and 2. With verbal “left”/“right” responses VPT-2 take longer, i.e., RTs seem to be shifted up by more than 100 ms. It is puzzling that verbal “left”/“right” judgments come at such an increased cost compared to motor responses. “In front”/“behind” judgments do not seem to suffer such a cost compared to visibility judgments (visible/occluded). Also, a general speed-accuracy-trade-off does not seem to account for the particularly elevated costs of VPT-2 with a verbal response (see Comparison Between Experiment 1 and 2). We propose that for “left”/“right” judgments language processes, e.g., in form of lexical access, take longer compared to “in front”/“behind,” possibly due to their larger ambiguity (e.g., Coventry and Garrod, 2004; e.g., Levelt, 1996; Grabowski, 1999). This ambiguity is reflected in larger RTs for verbal “left”/“right” responses even at low angular disparities (compare **Figure 3B**, right graph), and possibly even at 0° angular disparity as Grabowski (1999) suggested.

### VPT-1 VS. VPT-2

We have replicated Michelon and Zacks’ (2006) findings that VPT-1 and VPT-2 reveal qualitatively different response time patterns in relation to increasing angular disparity: While VPT-1 was not affected in any significant way, VPT-2 showed a significant increase in response time with increasing angular disparity. Our results in total corroborate the notion that the two types of perspective taking are based on two qualitatively different processes. While we replicated our previous findings (Kessler and Thomson, 2010) showing that VPT-2 strongly depends on the congruence of the participants’ body posture with the direction in which the avatar is seated, we showed for the first time that no such pattern is observed for VPT-1. We conclude in concordance with Kessler and Thomson (2010) that VPT-2 is the endogenous (self-initiated) emulation or simulation of a body rotation into another perspective.

In contrast, VPT-1 seems to involve a process that determines object locations in relation to the line-of-sight of another person (Michelon and Zacks, 2006). This could be regarded as some sort of embodied representation in its own right: Gaze is a very strong social cue that has been related to processes of motor resonance (for a review Frischen et al., 2007). Perceiving someone’s gaze is processed in a brain network that overlaps with gaze control (Grosbras et al., 2005). This would support the notion of embodiment of VPT-1 in the form of motor resonance, which, however, does not rely on deliberate movement simulation like VPT-2 (see Kessler and Thomson, 2010, Experiment 4 for details of this distinction). Michelon and Zacks (2006, Experiment 3) investigated VPT-1 in the absence of an avatar, that is, without an external “gaze”-anchor for establishing the line-of-sight. Humans can easily imagine a virtual line-of-sight and solve the task. The interesting question is whether such a process would be implemented in part by cortical gaze control areas. This could possibly extend the notion of embodiment of VPT-1 toward deliberate simulation in gaze coding areas. However, based on our current findings we conclude that even if VPT-1 involves gaze simulation, this form of simulation is very different from movement simulation during VPT-2.

### Implications for the neural substrates of VPT-1 and VPT-2

We propose the following hypotheses regarding the neural substrates of VPT-1 and VPT-2, which seem most compatible with the currently available data. VPT-2 is either a form of action simulation that essentially involves action control areas in the posterior frontal cortex together with body schema representations in the parietal lobe and possibly vestibular information represented in the Insula, or VPT-2 is a form of action emulation, where the perceptive and proprioceptive outcomes of the transformation are generated without the need for a full movement simulation that contains all the intermediate steps “to get there.” Although this distinction is rather gradual in our case an emulation process during VPT-2 might not require action control areas to be involved, while essentially relying on the emulation of representations in body schema and other proprioceptive areas (e.g., vestibular representations in the Insula).

Parietal areas and in particular the temporo-parietal junction (TPJ) indeed seem to be a prominent part of the physiological substrate of VPT-2 (Zacks and Michelon, 2005; Arzy et al., 2006; Keehner et al., 2006). For example Arzy et al. (2006) and Blanke et al. (2005) reported that the TPJ was related to disembodied processing (the self was imagined outside the body) and Samson et al. (2005) reported difficulties with VPT-2 after lesions in this area. Recent ERP mapping results corroborated the importance of the TPJ and further implicated posterior frontal areas (Schwabe et al., 2009). Zacks and Michelon (2005) also concluded that VPT-2 recruits posterior frontal areas that code for body movements but Wraga et al. (2005) questioned the involvement of full action simulation in VPT-2 compared to mental object rotation. Hence, for VPT-2 strong evidence exists in support of the temporo-parietal junction being part of the neural substrate, while posterior frontal and vestibular/insular (e.g., Blanke and Thut, 2007; Grabherr et al., 2007) contributions are still being debated.

To our knowledge no conclusive neuroimaging results exist for VPT-1, but in the light of our results simulation/emulation networks are an unlikely neural substrate. We predict that the dorsal between-objects pathway in parietal cortex would be the main processing substrate together with the ventral object/person recognition system in the temporal lobe (e.g., Goodale and Milner, 1992; Ungerleider and Haxby, 1994), but would not require additional processing input from action control or proprioceptive areas.

### Implications for evolution and development

The qualitative distinction between VPT-1 and VPT-2 we observed in our data is reflected in onto- and phylogenetic differences. As we described in the Introduction, VPT-1 develops around the age of 2 years and autistic children do not seem to experience particular difficulties with this task (Leslie and Frith, 1988; Baron-Cohen, 1989). VPT-2 tends to develop around 4–5 years (Gzesh and Surber, 1985; Hamilton et al., 2009), but not in children diagnosed with autism spectrum disorder (ASD) (Hamilton et al., 2009). Primates seem capable of certain forms of VPT-1 but not at all of VPT-2 (Tomasello et al., 2005). This suggests that VPT-2 is the more advanced cognitive process, which is consistent with our RT results where VPT-2 was slower, and hence, the more demanding cognitive process than VPT-1.

We concluded (see VPT-1 vs. VPT-2, Implications for the Neural Substrates of VPT-1 and VPT-2) that VPT-2 is the endogenous (self-initiated) emulation or simulation of a body rotation into another perspective, while VPT-1 does not seem to rely on such a simulation

process. Instead VPT-1 seems to involve a process that determines object locations in relation to the line-of-sight of another person. Although movement simulation seems like a simple operation at first glance, it minimally requires the awareness of two pieces of information. Firstly, the knowledge that someone else may have a very different view of the world which requires some form of alignment before it can be fully understood, and that, secondly, one does not always have to physically change one's location in order to achieve such an alignment; instead, one can simply imagine it. While we seem to share the first step with apes who change their location to align themselves with humans or conspecifics (Tomasello et al., 1998; Brauer et al., 2005), the second stage seems to be uniquely human. The closer one looks the more impressive this latter achievement actually becomes: we use a skill that evolved millions of years ago, i.e., moving, decouple the actual execution of the movement from its planning, control, and sensory transformation and employ the abstract movement representation for a mental simulation of the representational consequences.

According to these considerations VPT-2 should develop after certain forms of mental simulation have been mastered by a given individual, specifically the skill and the awareness that one can actually imagine oneself deliberately in another location, i.e., outside our own body. The next step would be to employ such a mental operation to imagine someone else's perspective. This can be regarded as a prototype of theory of mind (ToM) where an individual infers the mental states of another person. In support of this claim Hamilton et al. (2009) found that only ToM score, but not verbal skills or mental object rotation performance, predicted VPT-2 ability in typically developing and in children diagnosed with autism. While mental object rotation performance was not impaired in the autism group, they had significantly lower ToM scores and were significantly impaired on the VPT-2 task.

So what might go wrong in autism? Can autistic individuals mentally simulate movements without actually executing them? Can they imagine themselves outside their bodies? If the answers are yes, is the transfer of this skill onto inferring someone else's mental states amiss or hampered? To our knowledge there is no report of adults diagnosed with autism that describes their ability or disability to conduct VPT-2 and the particular strategies they employ in case they have mastered this ability at some point of their individual development. A posture change like the one we employed here could reveal if autistic participants use movement simulation for VPT-2 in the same way as typical participants, or whether they have learned to employ a mental object rotation (OR) strategy instead, since OR does not seem to impose any difficulties for them (e.g., Hamilton et al., 2009). Kessler and Thomson (2010) showed that with typical participants VPT-2 as well as OR increased with angular disparity, yet, that the posture manipulation allowed for a qualitative distinction between VPT-2 and OR.

#### IMPLICATIONS FOR SPATIAL LANGUAGE: GROUNDED, SITUATED, AND EMBODIED PROCESSING

Both processes of perspective taking result in higher-level cognitive representations about other people's views of the world that are strongly grounded in the context of the specific situation (cf. Harnad, 1990; Barsalou, 2008; Myachykov et al., 2009). Specifically, the location and orientation of the avatar and the location of the target in relation to the avatar as well as to the occluder (or relatum in general) provide

the situational input to the VPT processes. These transform the input into an abstract cognitive representation that is a direct outcome of the situational constraints and, hence, grounded in the situation.

However, based on our results we propose that for determining the situationally grounded meaning of projective spatial prepositions, "left" and "right" involve an effortful, embodied process of movement simulation that is equivalent to VPT-2, while "in front" and "behind" are determined quite effortlessly based on the generation of a "line-of-sight" and the resulting spatial relations; equivalent to VPT-1. Grabowski (1999) suggested an anthropomorphological definition of the semantics of these four spatial prepositions proposing that the human line-of-sight would determine the "front" and as a result the "behind" from a given perspective<sup>1</sup>. This is in full agreement with the notion we propose here (also Kessler, 2000), namely that both spatial dimensions and the related four prepositions are not only grounded in the situational constraints but are also related to the human perceptual apparatus and body. According to this notion front/behind are related to asymmetries of the human body (eyes define the front), while left/right relate to the symmetrical sides of our bodies. As a consequence the front of any other perspective can be easily determined or imagined, which automatically provides a line-of-sight and the "front" pole within the visible area in relation to a relatum (in the experiments presented here this was the occluder in the middle of the table, and in **Figure 1** it is the tree). Next, "behind" can be directly determined as the opposite pole, which is occluded from the view of the target perspective. Hence, the first horizontal dimension can be easily determined based on features and inter-object relations within the presented scene resulting in a grounded and situated "meaning" of the respective prepositions. Furthermore, our results suggest that this process is identical to VPT-1.

Due to the symmetry of our bodies "left" and "right" cannot be determined based on visual features, but have to be determined with more effort, namely, by means of a perspective transformation into the target perspective. Our results have confirmed that this transformation process is identical to level 2 perspective taking (VPT-2), which we now know is the simulation or emulation of a body rotation.

In terms of embodiment and grounding we propose that "in front"/"behind" are essentially grounded in relation to the visible features of the human body but that the extent to which embodiment – in a more powerful sense, conform to Wilson's sixth meaning of embodied cognition (2002) – is still part of the ongoing processing (e.g., simulation within gaze control areas) is an open empirical question as pointed out in a previous section (VPT-1 vs. VPT-2). "Left"/"right," in contrast, essentially rely on movement simulation/emulation when determined from a perspective that is misaligned with the egocentric viewpoint. These distinctions between embodied, grounded, and situated processing are in concordance with Myachykov et al.'s (2009) notion proposed in the context of number representations. Our care in distinguishing the "when," "how much," and "in what way" of embodiment of VPT-1 and VPT-2 and in finding different answers for the two forms of perspective taking is further in agreement with Chatterjee's (2010) recent call for more rigorous distinctions when claiming embodiment of cognitive processes.

<sup>1</sup>Grabowski suggests several important distinctions, e.g., "inside perspective" vs. "outside perspective" Here we simply focus on the main aspect of this notion, namely the relation to the human body as an anchor for spatial dimensioning.

### **Cognitive effort: implications for selecting a frame of reference**

In congruence with the considerations in the previous section (Implications for Spatial Language: Grounded, Situated, and Embodied Processing), Kessler (2000) devised a model for the interpretation of spatial prepositions in complex situations with up to three competing frames of reference (FOR). For each FOR (or potential perspective) a basic direction, so-called “anchor” direction, is extracted based on the visual features of that perspective (e.g., face of another person or front of an intrinsically directed object like a car; see Grabowski, 1999 and Levelt, 1996, for discussions). This anchor direction is equivalent to “in front” (i.e., visible) while “behind” (i.e., occluded) requires a single additional step in the opposite direction. This prioritization of the front pole predicts an advantage in speed over the opposite pole, which is exactly what we found disregarding response modality (i.e., “visible” and “in front” are processed faster than “occluded” and “behind”). Furthermore such an “anchor” direction can then be employed to determine the left and right poles in the orthogonal direction: the egocentric gaze direction (egocentric anchor) is rotated into the target anchor and the egocentric left and right are mapped onto the target left and right of that perspective. This is essentially a model of perspective transformation and Kessler (2000, pp. 135, 208–210) pointed out the similarity between the model features and a gradual transformation of a body representation for action control. Our recent findings (Kessler and Thomson, 2010) and the results presented here support this hypothesis.

Hence, while the situated meaning of all four projective prepositions is strongly grounded within the spatial configuration of the situation, only localizations employing “left” or “right” necessitate a movement simulation. While one could point out that a line-of-sight could also be regarded as an embodied basis that subserves VPT-1 (see VPT-1 vs. VPT-2), our findings reveal that only the movement simulation for VPT-2 comes at a cost that increases with disparity, i.e., the distance of the simulated movement, and is modulated by the congruence of proprioceptive information (i.e., body posture). Such a qualitative distinction especially in terms of cognitive effort would have strong implications for the use of spatial prepositions in specific situations, assuming that cognitive systems tend to minimize their effort. While there is no increased effort involved in determining “in front” or “behind” from the perspective of someone sitting opposite to ourselves, it implies a mental self-rotation effort to adopt this person’s perspective for “left” or “right.” Therefore we might willingly adopt the frame of reference (FOR) of the other person for the use of “in front” and “behind” while we might be reluctant to do the same for “left” and “right” due to the involved cognitive effort (Kessler, 2000). Additional factors within the situational context play a role in our “willingness” to take on the effort or not, as has been shown for socio-emotional factors (Graf, 1994; Levelt, 1996; Grabowski and Miller, 2000; Kessler, 2000; Coventry and Garrod, 2004), but also for more implicit influences such as an action-related topic of the situation (Tversky and Hard, 2009).

In contrast, even the most situated and grounded psycholinguistic accounts (e.g., Grabowski, 1999; Grabowski and Miller, 2000) still implicitly assume that first and foremost a frame of reference (FOR) is chosen (e.g., egocentric, partner-centered, or intrinsic),

and then all relative dimensions of space (i.e., “left”/“right” AND “in front”/“behind”) are determined based on this particular FOR, thus, implying that the identical cognitive transformation into that FOR (i.e., “origo projection,” Grabowski, 1999; Grabowski and Miller, 2000) underlies all dimensional relations (i.e., “in front”/“behind” AND “left”/“right”). In contrast, we have confirmed the hypothesis of Kessler (2000) that two qualitatively different cognitive processes (VPT-1 vs. VPT-2) are involved. Accordingly, we predict that different FORs may be chosen for the two horizontal dimensions (“in front”/“behind” vs. “left”/“right”) in situations, where the two underlying processes (VPT-1 and VPT-2, respectively) significantly differ in terms of cognitive effort. In essence we propose a dissociation between semantic conceptions of spatial prepositions and the psychology of their situated, grounded, and embodied use.

### **CONCLUSIONS**

Based on our findings we draw the following three major conclusions. Firstly, level 1 (VPT-1) and level 2 (VPT-2) perspective taking are qualitatively different cognitive processes, especially with respect to embodiment. VPT-2 is the mental simulation of a body movement where the effort increases with mental distance to cover (angular disparity effect) and with the level of body posture incongruence. VPT-1 relies on determining visibility of the target by imagining the line-of-sight from a given perspective. This process does not depend on mental distance (no angular disparity effect) or on movement simulation (no body posture effect). The finding that visibility is faster to judge than occlusion further supports the assumed process. VPT-2 is the more effortful and, hence, the more sophisticated cognitive process, which is consistent with the available developmental, comparative, and neuroimaging data.

Secondly, both VPT processes are applied in their essential form with key presses as well as with verbal responses in form of prepositions. We therefore conclude that VPT-1 is the cognitive process that subserves verbal localizations using “in front” and “behind” from a perspective other than the egocentric and that VPT-2 is the cognitive process that subserves verbal localizations using “left” and “right” from a perspective other than the egocentric. In the latter case, however, verbal responses come at an additional cost compared to key presses, which might reflect higher ambiguity. Both VPT processes result in a grounded and situated “meaning” of the prepositions they subserves, but only VPT-2 is embodied in the form of movement simulation.

Thirdly, the difference in cognitive effort associated with VPT-2 compared to VPT-1 implies that in specific situations language users will prefer a different, less “effortful,” frame of reference for “left”/“right” than for “in front”/“behind.” This prediction is at odds with the assumption of a single and general “origo projection” process for the psychological use of spatial prepositions.

### **ACKNOWLEDGMENTS**

We would like to thank Anna Borghi and the two Referees for their very helpful comments for improving our manuscript. This research was supported by ESRC/MRC funding (RES-060-25-0010) awarded to Klaus Kessler. Finally we would like to thank the eight Raters for their help with validating our procedures.

## REFERENCES

- Arzy, S., Thut, G., Mohr, C., Michel, C. M., and Blanke, O. (2006). Neural basis of embodiment: distinct contributions of temporoparietal junction and extrastriate body area. *J. Neurosci.* 26, 8074–8081.
- Baron-Cohen, S. (1989). Perceptual role taking and protodeclarative pointing in autism. *Br. J. Dev. Psychol.* 7, 113–127.
- Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.
- Blanke, O., Mohr, C., Michel, C. M., Pascual-Leone, A., Brugger, P., Seeck, M., Landis, T., and Thut, G. (2005). Linking out-of-body experience and self processing to mental own-body imagery at the temporoparietal junction. *J. Neurosci.* 25, 550–557.
- Blanke, O., and Thut, G. (2007). “Inducing out-of-body experiences,” in *Tall Tales: Popular Myths About the Mind and Brain*, ed. S. Della Sala (Oxford: Oxford University Press), 425–439.
- Brauer, J., Call, J., and Tomasello, M. (2005). All great ape species follow gaze to distant locations and around barriers. *J. Comp. Psychol.* 119, 145–154.
- Bryant, D. J., Tversky, B., and Franklin, N. (1992). Internal and external spatial frameworks for representing described scenes. *J. Mem. Lang.* 31, 74–98.
- Call, J., and Tomasello, M. (1999). A non-verbal false belief task: The performance of children and great apes. *Child Dev.* 70, 381–395.
- Chatterjee, A. (2010). Disembodying cognition. *Lang. Cogn.* 2, 79–116.
- Coventry, K. R., and Garrod, S. C. (2004). *Saying, Seeing and Acting: The Psychological Semantics of Spatial Prepositions*. Hove, New York: Psychology Press.
- Davidson, M. L. (1972). Univariate versus multivariate tests in repeated-measures experiments. *Psychol. Bull.* 77, 446–452.
- de Vega, M., and Rodrigo, M. J. (2001). Updating spatial layouts mediated by pointing and labelling under physical and imaginary rotation. *Eur. J. Cogn. Psychol.* 13, 369–393.
- Flavell, J. H., Green, F. L., and Flavell, E. R. (1986). Development of knowledge about the appearance-reality distinction. *Monogr. Soc. Res. Child Dev.* 51, i–v, 1–87.
- Frischen, A., Bayliss, A. P., and Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychol. Bull.* 133, 694–724.
- Frith, C. D., and Frith, U. (2007). Social cognition in humans. *Curr. Biol.* 17, R724–R732.
- Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25.
- Grabherr, L., Karmali, F., Bach, S., Indermaur, K., Metzler, S., and Mast, F. W. (2007). Mental own-body and body-part transformations in micro-gravity. *J. Vestib. Res.* 17, 279–287.
- Grabowski, J. (1999). A uniform anthropomorphological approach to the human conception of dimensional relations. *Spat. Cogn. Comput.* 1, 349–363.
- Grabowski, J., and Miller, G. A. (2000). Factors affecting the use of dimensional prepositions in German and American English: object orientation, social context, and prepositional pattern. *J. Psycholinguist. Res.* 29, 517–553.
- Graf, R. (1994). *Selbstrotation und Raumreferenz: Zur Psychologie partnerbezogenen Lokalisierens (Self-rotation and Spatial Reference: The Psychology of Partner-Centred Localisations)*. Frankfurt: Peter Lang.
- Grosbras, M. N., Laird, A. R., and Paus, T. (2005). Cortical regions involved in eye movements, shifts of attention, and gaze perception. *Hum. Brain Mapp.* 25, 140–154.
- Gzesh, S. M., and Surber, C. F. (1985). Visual perspective-taking skills in children. *Child Dev.* 56, 1204–1213.
- Hamilton, A. F. D., Brindley, R., and Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition* 113, 37–44.
- Harnad, S. (1990). The symbol grounding problem. *Physica D* 42, 335–346.
- Hegarty, M., and Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence* 32, 175–191.
- Herrmann, T., Bürkle, B., and Nirmaier, H. (1987). Zur hörerbegrenzten Raumreferenz: Hörerposition und Lokalisationsaufwand (Listener-oriented spatial reference: Listener position and localisation effort). *Sprache Kogn.* 6, 126–137.
- Huttenlocher, J., and Presson, C. C. (1973). Mental rotation and the perspective problem. *Cogn. Psychol.* 4, 277–299.
- Keehner, M., Guerin, S. A., Miller, M. B., Turk, D. J., and Hegarty, M. (2006). Modulation of neural activity by angle of rotation during imagined spatial transformations. *Neuroimage* 33, 391–398.
- Kessler, K. (2000). *Spatial Cognition and Verbal Localisations: A Connectionist Model for the Interpretation of Spatial Prepositions*. Wiesbaden: Deutscher Universitäts-Verlag.
- Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88.
- Kozhevnikov, M., and Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Mem. Cognit.* 29, 745–756.
- Kozhevnikov, M., Motes, M. A., Rasch, B., and Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Appl. Cogn. Psychol.* 20, 397–417.
- Leslie, A. M., and Frith, U. (1988). Autistic childrens understanding of seeing, knowing and believing. *Br. J. Dev. Psychol.* 6, 315–324.
- Levelt, W. J. M. (1996). “Perspective taking and ellipsis in spatial descriptions,” in *Language and Space*, eds. P. Bloom, M. A. Peterson, L. Nadel, and M. F. Garret (Cambridge, MA: A Bradford Book), 77–108.
- Levine, M., Jankovic, I. N., and Palij, M. (1982). Principles of spatial problem solving. *J. Exp. Psychol. Gen.* 111, 157–175.
- Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102, 59–70.
- Michelon, P., and Zacks, J. M. (2006). Two kinds of visual perspective taking. *Percept. Psychophys.* 68, 327–337.
- Myachykov, A., Platenburg, W. P. A., and Fischer, M. H. (2009). Non-abstractness as mental simulation in the representation of number. *Behav. Brain Sci.* 32, 343–344.
- O'Brien, R. G., and Kaiser, M. K. (1985). Manova method for analyzing repeated measures designs – an extensive primer. *Psychol. Bull.* 97, 316–333.
- Pack, A. A., and Herman, L. M. (2006). Dolphin social cognition and joint attention: our current understanding. *Aquat. Mamm.* 32, 443–460.
- Retz-Schmidt, G. (1988). Various views on spatial prepositions. *AI Mag.* 9, 95–105.
- Samson, D., Apperly, I. A., Kathirgamanathan, U., and Humphreys, G. W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain* 128(Pt 5), 1102–1111.
- Scheumann, M., and Call, J. (2004). The use of experimenter-given cues by South African fur seals (*Arctocephalus pusillus*). *Anim. Cogn.* 7, 224–230.
- Schwabe, L., Lenggenhager, B., and Blanke, O. (2009). The timing of temporoparietal and frontal activations during mental own body transformations from different visuospatial perspectives. *Hum. Brain Mapp.* 30, 1801–1812.
- Tomasello, M., Call, J., and Hare, B. (1998). Five primate species follow the visual gaze of conspecifics. *Anim. Behav.* 55, 1063–1069.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691; discussion 691–735.
- Tversky, B., and Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition* 110, 124–129.
- Ungerleider, L. G., and Haxby, J. V. (1994). ‘What’ and ‘where’ in the human brain. *Curr. Opin. Neurobiol.* 4, 157–165.
- Vasey, M. W., and Thayer, J. F. (1987). The continuing problem of false positives in repeated measures anova in psychophysiology – a multivariate solution. *Psychophysiology* 24, 479–486.
- Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636.
- Wraga, M., Shephard, J. M., Church, J. A., Inati, S., and Kosslyn, S. M. (2005). Imagined rotations of self versus objects: an fMRI study. *Neuropsychologia* 43, 1351–1361.
- Zacks, J. M., and Michelon, P. (2005). Transformations of visuospatial images. *Behav. Cogn. Neurosci. Rev.* 4, 96–118.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 September 2010; accepted: 10 November 2010; published online: 06 December 2010.

Citation: Kessler K and Rutherford H (2010) The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Front. Psychology* 1:213. doi: 10.3389/fpsyg.2010.00213

This article was submitted to *Frontiers in Cognition*, a specialty of *Frontiers in Psychology*.

Copyright © 2010 Kessler and Rutherford. This is an open-access article subject to an exclusive license agreement between the authors and the *Frontiers Research Foundation*, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.