Cognition 117 (2010) 101–105

Contents lists available at ScienceDirect

Cognition

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

Brief article

Observing fearful faces leads to visuo-spatial perspective taking

Jan Zwickel [a,*], Hermann J. Müller [a,b]

[a] Ludwig Maximilians University Munich, Department of Psychology, Leopoldstrasse 13, 80802 Munich, Germany
[b] Department of Psychological Sciences, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK

ARTICLE INFO

Article history:
Received 28 July 2009
Revised 19 June 2010
Accepted 16 July 2010

Keywords:
Visuo-spatial perspective taking
Social cognition

ABSTRACT

A number of recent studies suggested that visuo-spatial perspective taking (VSPT) occurs spontaneously when viewing either a human body or an action by an agent. However, it remains unclear whether VSPT is caused by the observation of an (potential) action or occurs because the observer infers from certain cues that another mind is present (whether actions/action cues are displayed or not). To examine whether action cues are necessary for VSPT, we presented to participants human faces without a torso, that is: stimuli that indicated the presence of another mind, without providing action (effector) cues. Furthermore, we examined whether 'relevance' of understanding the other mind would influence VSPT, where relevance was manipulated by comparing the effects of observing a fearful versus a neutral facial expression on VSPT. Results showed that spontaneous VSPT occurs when a face with a fearful, but not with a neutral, expression is perceived. This indicates that spontaneous VSPT occurs (at least more robustly) under circumstances, where VSPT is of 'relevance' for understanding the situation. Furthermore, directly observing actions, or action cues, does not appear to be a prerequisite for VSPT.

1. Introduction

One important component in successful social behavior is to understand and predict the intentions and feelings of conspecifics. One crucial step in predicting others' behavior is to understand the state the other is in Apperly (2008), and one component of understanding may be to represent the world from the viewpoint of the other (Sodian & Thoermer, 2008; Weiskopf, 2005). Given the social importance of visuo-spatial perspective taking (VSPT), it is not surprising that there has been a growing interest in VSPT in humans (Belopolsky, Olivers, & Theeuwes, 2008; Frischen, Loach, & Tipper, 2009; Thomas, Press, & Haggard, 2006; Tversky & Hard, 2009; Zwickel, 2009).

A first step to demonstrate that the presence of other humans involved in a task changes the spatial coding of visual events has been taken by Sebanz, Knoblich, and Prinz (2003). Participants made right/left decisions in response to the color of a ring on a finger. If this task was performed together with another person, the irrelevant spatial pointing direction of the finger influenced the reaction time (RT); this was, however, not the case when no second participant responded.

The coding of visual events relative to observed human bodies was demonstrated in a study by Thomas et al. (2006). Participants saw a human sitting opposite and facing them. Their task was to detect a tactile stimulus applied to their own body. They responded faster when the tactile stimulus was preceded by a visual stimulus on an anatomically congruent body part of the model. As the facing condition permitted anatomical and specular congruency to be dissociated, the results showed that the visual stimuli were coded relative to the observed human body. By using houses as control stimuli, the authors could show that the effect depended on the presence of a human body.

The importance of actions for VSPT to occur was underlined by Frischen et al. (2009). In this study, participants had to reach to a target location while ignoring a visual

* Corresponding author. Tel.: +01 203 432 2518.
E-mail addresses: jan.zwickel@yale.edu, zwickel@psy.lmu.de (J. Zwickel).

0010-0277/$ - see front matter
doi:10.1016/j.cognition.2010.07.004

distractor stimulus. In such situations, reaches are typically delayed to targets appearing at locations, where there had been a distractor on the previous trial, with the biggest slowing found for distractors close to the participant (Tipper, Howard, & Houghton, 1998). The same selective slowing was found by Frischen et al. (2009) in a single-person condition. However, in another condition in which the reach on the previous trial had been performed by a facing person, the strongest slowing was observed for targets at distractor locations close to the facing individual. This was interpreted as showing that the reaching action was encoded from the perspective of the model (see Belopolsky et al. (2008), for a related observation).

That VSPT also affects verbal descriptions was shown by Tversky and Hard (2009), who asked participants to describe the location of objects in pictures. On some trials, participants spontaneously described the objects relative to a depicted person in the scene. This tendency was further increased by focusing on an action, which was manipulated by asking participants, for instance, 'in relation to the bottle, where does he place the book?' rather than 'in relation to the bottle, where is the book?'.

Importantly, the presence of a human body is not a necessary condition for spontaneous VSPT to occur, as shown by Zwickel (2009). In this study, VSPT occurred even in the case of triangle objects, as long as these exhibited movement patterns that elicit interpretations in terms of agency or mental states (Abell, Happé, & Frith, 2000). Here, participants were to respond as fast as possible to dots that were presented either to the right or to the left of a moving triangle. In one condition, the spatial decision was the same when made from the viewpoint of the participant and when made from the perspective of the moving triangle. For example, when the triangle was pointing upwards, a dot presented on the right side as seen from the observing participant was also on the right side as seen from the heading direction of the triangle. When the triangle was pointing downwards however, these decisions were in conflict. In this case, a dot presented to the right of the triangle as seen from the viewpoint of the participant was on the left relative to the heading direction of the triangle. This conflict led to slower responses. Slowing was significant only when the triangles moved in a way that typically leads to the attribution of agency, and it was more pronounced in animations in which mental-state attributions are typically made.

The study of Frischen et al. (2009) suggests that the observation of an action is sufficient for VSPT to occur in the presence of a human body, whereas the study of Zwickel (2009) indicates that perceiving an action (whether of a human or a triangle) is sufficient and the presence of a human body not necessary. Finally, the studies of Thomas et al. (2006) and Tversky and Hard (2009) show that, in the presence of a human body, VSPT occurs even when no actions are displayed, suggesting that the presence of a human body is sufficient for VSPT to occur. Note, however, that a torso, too, displays effectors for actions, and arms were made relevant in the detection task used by Thomas et al. (2006). Thus, it remains unclear whether VSPT occurs only when (potential) actions are observable or simply when the presence of another mind is detected. This is an important issue, as it concerns the feasibility of an essentially action-centered account of VSPT. The current study was designed to shed light on this question by examining whether a face stimulus alone (i.e., without a torso that could potentially display action cues) would engender VSPT.

A related question addressed in the current study was whether the relevance of taking the perspective of someone else would have an influence on spontaneous VSPT; that is, would detecting another mind alone be sufficient for triggering VSPT, or would some relevance of perspective taking be necessary? In the study of Zwickel (2009), VSPT did not appear to depend on the visual properties of the stimuli, but rather on the relevance of VSPT for story understanding (e.g., one triangle could not 'see' the other from its view point). Similarly, we expected that making the visuo-spatial perspective of the face more relevant would increase the likelihood of VSPT. Relevance was manipulated by comparing faces with emotionally neutral expressions to faces that exhibited a fearful expression (increasing the relevance of the depicted person's perspective). The emotion of fear was chosen because we assumed it would make VSPT relevant for understanding the cause of the depicted person's fear (see Putman, Hermans, & Honk, 2006).

2. Experiment

To measure VSPT, we compared performance in a task in which responses to spatial stimuli were the same when made from the perspective of the participant or from that of the observed face (congruent task), to a task in which responses from the perspective of the participant and that of the observed face differed (incongruent task). In the latter, occurrence of VSPT would lead to response conflict, which should slow RT performance.

In the incongruent task, participants decided whether a dot on the computer screen was displayed to the left or the right of the face. The correct response to this dot would differ when performed from the perspective of the observed face; from the perspective of the face, the responses would be reversed. In the congruent task, participants were to decide whether a dot was presented above or below the face. The presentation of a black rectangle of the same size as a face stimulus served as a baseline against which to test VSPT.

Furthermore, we wanted to examine the influence of attention on this effect, in particular, whether it would be necessary to direct attention to the emotion of the face for finding a difference between neutral and fearful faces. To manipulate attentional set, participants had to indicate after their spatial decision either whether the face (if present) had been a female or male face (gender instruction) or whether it had displayed a neutral or a fearful expression (emotion instruction).

If the mere presence of a face led to VSPT, one would expect stronger VSPT effects in the neutral-face compared to the baseline condition, evidenced by a selective response slowing on incongruent, but not congruent, trials. The influence of VSPT relevance was tested by comparing the fearful face against the neutral-face condition. The influence of relevance should be observed as a (particularly marked) response slowing on incongruent trials in the fearful-face condition. Finally, an influence of attentional set was expected to become manifest in stronger VSPT in the emotion instruction condition.

2.1. Method

2.1.1. Participants

Twenty-two participants took part in the experiment (mean age 25 years, 11 females). Two of them were recruited as replacements for two participants who failed to follow the task instruction.

2.1.2. Apparatus and stimuli

Stimuli were presented on a 19" computer screen (85-Hz refresh rate) positioned 50 cm in front of the participants. Responses were collected from a standard keyboard. Twenty-four edited versions of the Karolinska faces (Lundqvist & Litton, 1998) were used (12 females, 12 males), with their hair removed and the (colored) faces grey-scaled and presented against a black rectangular (4° in width and 6° in height) background. These stimuli were presented on a white screen surround. Twelve of the images displayed a neutral and twelve a fearful expression. Filled black rectangles of the same size as the faces served as baseline stimuli.

2.1.3. Design and procedure

Half of the participants started with the gender instruction and then switched to the emotion instruction, and vice versa for the other half. Each part started with an instruction screen detailing the task, followed by 20 training trials and 288 experimental trials. The instructions told participants to (internally) make their dot location and gender or, respectively, emotion decision at the same time, rather than postponing the second task. This was further encouraged by presenting a (checker pattern) mask after the first response. Trial order was pseudo-randomized with the constraint that each face/baseline stimulus and dot location combination occurred with the same frequency.

Trials started with the display of a face stimulus (or a black rectangle in the baseline condition). After 500 ms, a dot (.5° in diameter) appeared 1° to the right, left, top, or bottom of the stimulus for 35 ms. Participants were to respond as fast as possible by pressing the 'k' key to a dot on the left and the 'l' key to a dot on the right. The 'e' key was to be pressed to dots appearing above and the 'd' key to dots below the face. Upon this response, the face stimulus was replaced by a mask, which was surrounded by the German words for 'don't know', 'nothing', 'male'/'fearful', and 'female'/'neutral' in the gender/emotion task, respectively. The individual words were placed randomly 13° to the right, left, top, and bottom of the face stimulus. Participants selected one of the alternatives by pressing the key that matched the direction of the selected word ('k', 'l', 'e', or 'd'). After the response, the screen went blank and a new trial started.

2.1.4. Data analysis

RT was measured from the onset of the dot. For analysis, trials on which the dots were responded to with the wrong key were removed, as were all trials with RTs longer than 1500 ms and, respectively, shorter than 150 ms (3% of the trials). The remaining RTs in a particular experimental condition were filtered according to a 3-standard-deviations procedure (5% of the trials).

As we were mainly interested in the differential effect of the face stimuli on the left/right and top/bottom decision tasks, rather than in absolute RT values, we calculated a direct measure of this, namely DDE (=dimension difference effect), that is, the difference in (mean) RTs for left/right trials minus top/bottom trials, computed separately for each face condition and participant. Higher DDE values in one face condition compared to another would indicate that RTs were relatively prolonged for left/right trials compared to top/bottom trials. Furthermore, we subtracted the (mean) baseline DDEs from the neutral- and fearful-face conditions to derive a quantitative measure for VSPT (VSPT value). Finally, these VSPT values were examined in a repeated-measures ANOVA with the factors instruction (gender, emotion) and face (neutral, fearful).

2.2. Results

Table 1 details the mean RTs for each condition. As can be seen, RTs increased from the baseline to the neutral-face condition for both left/right and top/bottom decisions. However, only RTs for left/right decisions increased from neutral to fearful faces, while there was hardly any difference between neutral- and fearful-faces for top/bottom decisions. This was the case whether the participants had to focus on the gender or the emotion of the faces. These observations were corroborated by a significant main effect of face ($F(1, 19) = 5.46$, $p < .05$, $\eta^2 = .22$), with no main effect of instruction ($F(1, 19) < 1$) and no interaction between instruction and face ($F(1, 19) = 2.38$, $p > .10$). Table 2 reports the mean VSPT values and 95% confidence intervals for the neutral and fearful-face conditions; Fig. 1 presents the mean VSPT values for the two face conditions. As can be seen from the 95% confidence intervals, only the fearful-face condition differed significantly from baseline ($t(19) = 3.43$, $p < .05$), but not the neutral-face condition ($t(19) = 0.08$, $p > .10$).

3. Discussion

Participants had to make a spatial decision in response to the eccentric location of a briefly presented dot. When they simultaneously looked at a fearful face in the screen center, they co-represented the visuo-spatial perspective of the face. This led to slower responses under task conditions in which the responses from the adopted and from the participant's perspective were in conflict. Accordingly, VSPT values differed significantly from baseline only for fearful faces, but not for neutral faces. The size of the VSPT effect, of 26 ms, was virtually identical to that reported by Zwickel (2009) for the difference in VSPT between 'theory-of-mind' and random conditions. The stability of VSPT effects across different paradigms is reassuring.

One caveat remains. Emotion and dot location judgments might share some kind of polarity correspondence (e.g., Proctor & Cho, 2006). Every condition required the

Table 1
Mean RTs, and associated standard errors (in brackets), in milliseconds, for each instruction (gender, emotion) and face (baseline, neutral, fear) condition as a function of task (left/right = horizontal, top/bottom = vertical) and the dimension difference effect (DDE). Values averaged across the two task conditions are also provided (average). Note that due to reporting rounded values, DDE values do not have to match the difference between Horizontal and Vertical values exactly.

Face	Gender			Emotion			Average		
	Horizontal	Vertical	DDE	Horizontal	Vertical	DDE	Horizontal	Vertical	DDE
Baseline	587 (32)	636 (34)	−49 (20)	599 (34)	643 (32)	−44 (14)	593 (32)	640 (29)	−46 (15)
Neutral	613 (37)	648 (35)	−35 (21)	640 (34)	699 (40)	−59 (15)	627 (34)	674 (33)	−47 (15)
Fearful	625 (39)	648 (35)	−23 (21)	675 (43)	692 (36)	−17 (17)	650 (38)	670 (32)	−20 (17)

Table 2
Mean VSPT and 95% confidence interval values (lower and upper limits) in ms for each instruction (gender, emotion) and face (neutral, fear) condition. Note that due to reporting rounded values, VSPT values do not have to match the differences between the face condition and baseline when calculated from Table 1 exactly.

Face	Gender			Emotion		
	VSPT	Lower	Upper	VSPT	Lower	Upper
Neutral	14	−6.12	33.47	−15	−43.54	13.72
Fearful	26	0.41	51.96	27	1.87	51.26

same number of left and right judgments. However, if the gain in reaction time due to correspondence and the loss due to non-correspondence were asymmetric, this could make the interpretation of our results more complex. To exclude any influence of polarity correspondence on our results, we included the side of dot presentation (left/right) as a factor. If polarity of emotion and dot location played a role, an interaction between face (neutral/fearful) and dot location (left/right) on RTs should be obtained. However, side of presentation did not interact with emotion (all interactions involving side of dot location and face emotion: $ps > .50$).

By comparing performance in tasks that should or should not be affected by VSPT, we showed that VSPT does not occur spontaneously in response to a face showing a neutral expression; rather, it occurs when the emotion of fear is displayed. This pattern was obtained independently of whether or not attention was directed to emotion by the task instruction. The absence of a neutral-face effect on VSPT in these situations indicates that it is not simply attention being directed to the emotional expression of a face, but rather the presence of a fearful expression that engenders spontaneous VSPT.

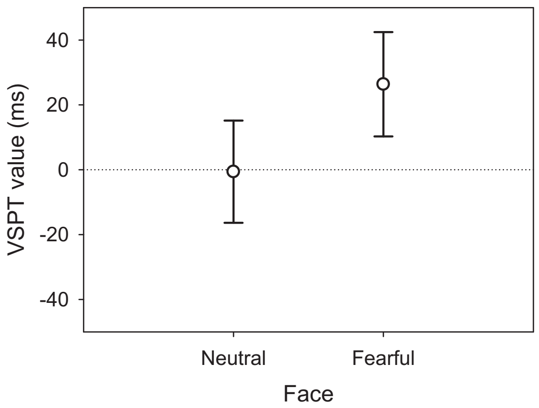

Fig. 1. Mean visuo-spatial perspective taking (VSPT) values, and associated 95% confidence intervals, for neutral and fearful faces. Positive values denote more VSPT compared to the baseline condition; negative values would indicate less VSPT compared to the baseline; and values near zero denote the same degree of VSPT as in the baseline condition.

The present findings suggest that it is not the observation of actions, or action cues, that leads to VSPT. Rather, VSPT occurs because of the detection of a mental state. In this view, action cues are only one kind of cue to attributing a mind and faces another. However, simply detecting the presence of another mind does not automatically trigger VSPT. Rather, some kind of relevance of VSPT appears to be required. Presumably, the display of actions or action effectors/cues in earlier studies established this necessary relevance. In the present study, we manipulated relevance by showing a fearful (vs. neutral) facial expression. One might speculate that when fear is observed in the other, VSPT supports detecting the source of the fear by taking into account what the other is seeing at the moment. Note, though, that the tendency for VSPT to occur might be enhanced by action processes *in the observer* that are invoked by observing an other's emotional (e.g., fearful) expression (this was suggested to us by an anonymous reviewer). On this view, VSPT would be reinforced (or even triggered in the first instance) by the *observer* preparing for some action in response to the other (e.g., a flight reaction in response to a fearful face), even though no direct action is *observed*. Further work is necessary to elaborate the role of such invoked action processes in VSPT.

Acknowledgement

This work was supported by a German Research Foundation (DFG) grant EC 142, Excellence Cluster "Cognition for Technical Systems (CoTeSys)".

References

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development, 15*(1), 1–16. <http://www.sciencedirect.com/science/article/B6W47-41GWN8R-1/1/732d0f8c8a>.

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*(1), 266–283. <http://dx.doi.org/10.1016/j.cognition.2007.07.019>.

Belopolsky, A. V., Olivers, C. N. L., & Theeuwes, J. (2008). To point a finger: Attentional and motor consequences of observing pointing movements. *Acta Psychologica, 128*(1), 56–62. <http://dx.doi.org/10.1016/j.actpsy.2007.09.012>.

Frischen, A., Loach, D., & Tipper, S. P. (2009). Seeing the world through another person's eyes: Simulating selective attention via action observation. *Cognition, 111*(2), 212–218. <http://dx.doi.org/10.1016/j.cognition.2009.02.003>.

Lundqvist, D., & Litton, J. E. (1998). The averaged Karolinska directed emotional faces. CD-ROM.

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin, 132*(3), 416–442. <http://dx.doi.org/10.1037/0033-2909.132.3.416>.

Putman, P., Hermans, E., & van Honk, J. (2006). Anxiety meets fear in perception of dynamic expressive gaze. *Emotion, 6*(1), 94–102. <http://dx.doi.org/10.1037/1528-3542.6.1.94>.

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition, 88*(3), B11–B21.

Sodian, B., & Thoermer, C. (2008). Precursors to a theory of mind in infancy: Perspectives for research on autism. *Quarterly Journal of Experimental Psychology, 61*(1), 27–39. <http://dx.doi.org/10.1080/17470210701508681>.

Thomas, R., Press, C., & Haggard, P. (2006). Shared representations in body perception. *Acta Psychologica, 121*(3), 317–330. <http://dx.doi.org/10.1016/j.actpsy.2005.08.002>.

Tipper, S. P., Howard, L. A., & Houghton, G. (1998). Action-based mechanisms of attention. *Philosophical Transactions: Biological Sciences, 353*(1373), 1385–1393.

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition, 110*(1), 124–129. <http://dx.doi.org/10.1016/j.cognition.2008.10.008>.

Weiskopf, D. A. (2005). Mental mirroring as the origin of attributions. *Mind & Language, 20*(5), 495–520. <http://dx.doi.org/10.1111/j.0268-1064.2005.00297.x>.

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