Agency Attribution and Visuo-Spatial Perspective Taking

Jan Zwickel

Ludwig Maximilian University Munich

Department of Psychology

Leopoldstrasse 13

80802 Munich Germany

zwickel@psy.uni-muenchen.de

word count: 3602

Abstract

We tested whether processes that evoke agency interpretations and mental state attributions also lead to adoption of the actor's visuo-spatial perspective in the observer. Agency and mental state interpretations were manipulated by showing different film clips involving two triangles (the Frith-Happé animations). Participants made speeded spatial decisions while watching these films. The response to the spatial task could be either the same or different when given from the perspective of the participant versus the perspective of one of the triangles. Reaction times were longer when the perspectives of participants and triangles differed compared to when they were the same. This effect increased as the need to invoke agency interpretations for film understanding increased and in those films that have previously been shown to evoke mental state attributions. This demonstrates that processing of an agent's behavior co-occurs with perspective adoption, even in the case of triangles as actors.

Agency Attribution and Visuo-Spatial Perspective Taking

A crucial part of human life involves social interactions. To react adequately in these situations it is important to take the representation about the world of the interacting partner into account, for example, to understand what further information would be needed in a conversation, or to predict actions based on the assumed state of the other.

Therefore, it is not surprising to find that people are generally willing to represent the situation of others (Frith & Frith, 2006) and do so even if this involves representing painful stimulation (Jackson, Meltzoff, & Decety, 2005). The ability to correctly represent what someone else knows requires that the visuo-spatial perspective (VSP) of the other is taken into account to understand what the other can or cannot know (Aichhorn, Perner, Kronbichler, Staffen, & Ladurner, 2006). This can then be used as a starting state to predict how the other person feels or will act (Apperly, 2008).

That VSP-taking occurs spontaneously (independent of task requirements) in the presence of humans has been shown by Tversky and Hard (2009). Tversky and colleagues asked participants to describe the spatial relationship of two objects in a picture ("in relation to the bottle, where is the book?"). In one experimental condition, a human was seated behind the two objects and faced the observer. Therefore, the book was to the right of the bottle from the observer's perspective but to the left of the bottle seen from the perspective of the depicted person. One picture was taken while the male actor was reaching for the book, another picture when the actor was looking at the book but not reaching. The last picture showed the same situation without a human. When the pictures contained a human, observers often spontaneously described the location of the book from the view of the depicted person. This tendency was further increased when the word "placed" was added to the question ("in relation to the bottle, where is the book *placed*?") which according to the authors would draw attention to the action and thereby increase the effect. These results were interpreted as showing that participants spontaneously took the perspective of the depicted person to make sense of the situation.

Another demonstration of VSP-taking in the presence of humans can be found in the study of Thomas, Press, and Haggard (2006). In the experiment, participants faced either a human model or an object (a house). Participants' task was to report a tactile cue that could either be in an anatomically same or different position with respect to a visual cue presented at the human model or object. For example, a tactile cue to the participant's right arm could follow a visual cue at the model's right arm (anatomical same) or at the model's left arm (same side, seen from the participant's perspective). In the human model condition, participants were faster for anatomically same than different tactile-visual conditions demonstrating that the perspective of the model played a role when coding the visual stimuli. No difference between same and different situations was found in the object condition. One important difference between this and the current study was that only the current study did involve movements. We expected that objects that display movement patterns that lead to the attribution of agency (Johnson, 2003) would also lead to VSP-taking.

These reported studies show that VSP-taking occurs spontaneously in the presence of a human but it is unclear whether this is caused by the presence of a human body as Thomas et al. suggested or because the presence of a human is taken as a cue for the presence of an agent which then causes VSP-taking. The increase in VSP-taking when attention was drawn to the action in Tversky et al. suggests the latter because human actions can be interpreted as cues of agency.

The current study should shed light on this question. If VSP-taking occurs even if only non-human entities are present as long as these entities seem to be agents it would support the interpretation that VSP-taking in the above studies did not occur because of the presence of a human but because the picture of the human acted as a cue to the presence of an agent. According to this reasoning, detecting an agent co-occurs with VSP-taking.

We tested in Experiment 1 whether processes that invoke the interpretation of agency, would also lead to VSP-taking when no visual features of humans are present. To this end we showed short film clips to the participants. These films were taken from the Frith-Happé animations and depict a red and blue triangle moving in a self-propelled fashion across the screen (Abell, Happe, & Frith, 2000). It has been shown before that these films activate agency and theory of mind (ToM, see below) processes to a different degree (e.g., Abell et al., 2000;

Castelli, Happé, Frith, & Frith, 2000; Klein, Zwickel, Prinz, & Frith, 2009). Each film belongs to one of three categories: The films “billiard”, “drifting”, and “tennis” belong to the random (R) category. They contain no interaction between the triangles. For example, in “tennis” the triangles are bouncing back and forth in a rather uncoordinated way. According to typical descriptions of participants, the triangles are floating around without purpose.

Triangles in films of the goal-directed (GD) category respond to physical events of each other, e.g., follow each other continuously (chasing), or with stops in-between (leading), or move around each other in a symmetric way (dancing). The movement of the triangles in these films could be described as fulfilling a certain goal (“the triangles danced around each other”) but it was not necessary to attribute a specific mental state to the triangles (“one triangle wanted to dance with the other”).

Importantly, only films from the ToM category typically lead to descriptions according to which the triangles react to each other’s underlying mental states. One description of the ToM film “mocking” could be that the small triangle is mocking the big triangle behind it’s back, but when the big triangle turns and therefore can see the small one, the small one pretends to do something else. A description of “coaxing” could be that the big triangle is pushing the small triangle, which wants to stay inside, outside the house. Finally, the big triangle manages to move the small one outside. Snapshots of the film “surprising” are depicted in Figure 1. Understanding of ToM animations therefore required some attribution of mental states to the triangles.

-- Figure 1 about here –

VSP-taking was expected to occur during films from the GD category because the two triangles are perceived as agents. If the attribution of mental states would further increase the tendency of VSP-taking then this should result in larger effects during films of the ToM than GD category. VSP-taking was measured by asking participants to respond to dots occurring right or left of the red triangle with right and left key presses respectively. In the following, "right" and "left" always refer to the observer’s perspective, which was also how participants were instructed to respond. These dots occurred either while the red triangle's tip was pointing upward or downward. If it was pointing upward, the triangle can be said to have a spatial orientation that matches the orientation of the participant. In this case, right and left decisions are the same, whether or not participants adopt and respond relative to the spatial orientation of the triangle or their own. However, when the triangle is facing-down it may create interference because if the perspective of the triangle is taken the response is incongruent to the response from the participant's perspective (see Figure 2). If agency/mental state attribution co-occurs with a stronger tendency of VSP-taking, responses during downward pointing directions in GD/ToM films should lead to more interference. This interference in turn should be reflected in longer reaction times (RTs).

-- Figure 2 about here --

Experiment 1

Participants, Apparatus, Stimuli, and Design

Twenty-four participants with normal or corrected-to-normal vision (13 females, mean age = 25) were paid for participation and were naive with respect to the purpose of the study. An eyetracker was used to ensure that participants fixated the middle of the screen at the beginning of each movie. Responses were collected with the left and right buttons of a gamepad. Head to monitor distance was approximately 60 cm.

Nine film clips from the Frith-Happé animations (Abell et al., 2000) were taken and shortened to about 18 seconds while preserving the essential story line (Klein et al., 2009). All of these films, 20° in width and 16° in height, showed a red and blue triangle with heights of about 4° and 2° and widths of about 2° and 0.5°, respectively.

During every film presentation 6 time-points were randomly selected with the constraint that all time points were separated by at least 1.5s and that in half of them the red triangle was pointing upward and at the other half downward. At each time point a filled gray circle with a diameter of 0.5° appeared 2° right or left of the center of mass of the red triangle for 30ms. This constraint ensured that half of the dot presentations were in the same relative position to the red triangle seen from the observer's and triangle's perspective (congruent) and half were at different relative positions (incongruent). The short 30ms duration was chosen so that no triangle movement occurred during dot presentation. Figure 2 depicts the two congruency conditions. The side of dot appearance was determined pseudo-randomly.

Film presentation was organized in blocks of films from the same category. Block order was balanced across participants. In each block every film appeared ten times during a pseudo-random sequence. This resulted in a total of 90 trials across all three blocks. Within participants film category (R, GD, ToM) and congruency (incongruent, congruent) was varied.

Procedure

Participants were instructed to watch the films attentively to report on their content later. Additionally, they should respond as fast as possible to the relative location of the dot, seen from *their* view. Participants were asked to press the right button, if the dot appeared to the right of the triangle and the left button, if it appeared on the left. Every trial started with a fixation cross (1°) at the center of the screen that was replaced by a film after 500ms of fixation. Importantly, all films were presented without breaks so that they only differed from films of the study by Klein et al. in the dots that appeared.

In addition to the three experimental blocks, a training block with three other films was run. After each block, participants reported what they had seen during the last presentations. These responses were only required to ensure that participants watched the films attentively and not further analyzed.

Data Analysis

RT was measured from onset of the dot until a button was pressed. First, dot presentations with RTs larger than or equal to 1500ms were excluded (*no responses*). Next, wrong responses were excluded (*wrong responses*). Finally, all responses that differed in RT by more than 2 standard deviations from the mean of the participant were not analyzed (*unfocused responses*). RTs were subsequently averaged within participants for each film category and congruency condition separately. The differences in RTs between incongruent and congruent decisions (congruency effect) for each film category were then subjected to a repeated measures ANOVA with the factor film category (R, GD, and ToM) and followed-up by paired t-tests.

Results and Discussion

The percentages of excluded trials were 8.49%, 4.96%, and 3.59% for no, wrong, and unfocused responses. The congruency effect increased from the R to the GD to the ToM condition (see Figure 3 and Table 1), which was statistically reflected in a main effect of film category ($F(2, 46) = 12.38$, $p < .01$, $\eta^2 = .35$). As can be seen from the 95% confidence intervals of Figure 3 only the GD and ToM conditions led to a significant congruency effect. Finally paired Bonferroni-corrected t-tests revealed a significantly higher congruency effect in the ToM than GD condition ($t(23) = 3.27$, $p_2 < .01$) and no difference between the GD and the R conditions ($t(23) = 1.81, p_2 > .10$). The number of wrong responses was higher in the incongruent than the congruent conditions and did not increase with faster RTs, this is evidence against a speed-accuracy trade off.

Comparing the mean RTs for the GD and ToM conditions in Table 1 seems to suggest that the congruency effect is also caused by a decrease in RTs in the congruent condition and therefore that VSP adoption not only leads to slower RTs in incongruent conditions but also to faster RTs in congruent conditions. As in the congruent ToM condition the response from the perspective of the participant and from the adopted perspective were the same as such the faster of the two could have determined the RT. Post-hoc we calculated an ANOVA on the RTs for the incongruent and congruent conditions separately, and found an effect of film condition for the incongruent conditions ($F(2, 46) = 4.18$, $p < .05$, $\eta^2 = .15$) but not for the congruent conditions ($F < 1$).

Additionally, interpreting this difference between the GD and ToM films for the congruent condition alone assumes that there is no general difference between RTs in the GD and ToM conditions. Making inferences in the GD condition, however, could be more difficult than mentalising in the ToM conditions and this would lead to generally elevated RTs in the GD condition. The higher error rates in the GD condition support this interpretation. This difference in RT between film categories is controlled for when using the difference between incongruent and congruent conditions within each film condition.

-- Figure 3 about here --

This pattern of results suggests that people spontaneously adopt the perspective of an agent and code responses also relative to this perspective. This VSP-taking was increased for films with mentalising content (see General Discussion). However, the current experiment did not exclude the possibility that these response differences were caused by superficial visual differences in the animations. To exclude this alternative interpretation Experiment 2 was performed.

-- Table 1 about here --

Experiment 2

With Experiment 2, we wanted to rule out that superficial visual differences between the animations, e.g., different positions of the triangle, caused the differences in response times in Experiment 1. Therefore twenty-five different participants saw the same triangle-dot situations from Experiment 1, i.e., each participant of Experiment 2 was matched to one participant of Experiment 1. Crucially, because only still pictures were shown agency attribution and mentalising should not occur. Therefore, if a congruency effect would still occur it would suggest that the effect in Experiment 1 was caused by differences in visual difficulty. However, observing no congruency effect would support the interpretation of Experiment 1 that it was indeed the attribution of agency or mental states that caused the congruency effect.

Participants, Apparatus, Stimuli, Design, Procedure, and Analysis

Twenty-five participants (17 females, mean age = 30) took part in Experiment 2. The number of participants was increased to replace one participant with more than 80% wrong responses. All other conditions were as reported for Experiment 1, except that no eye tracking was used and in each trial only six still pictures (the same decision situations as in Experiment 1) were shown. Each presentation started with a display of the situation without the dot. After a random interval of 500 - 1000ms the dot was added and disappeared as in Experiment 1 after 30ms. Participants responded as fast as possible with the "s" key if the dot occurred left of the triangle and with the "l" key if the dot occurred to the right.

Results and Discussion

Exclusion rates were 0.76%, 2.93%, 3.74% this time. RTs were considerably lower than in Experiment 1 which was probably caused by the additional task of understanding the story and the higher temporal uncertainty of the dot in Experiment 1. Importantly, RTs were nearly identical in every condition (see Tabel 1). This was corroborated by a non-significant ANOVA ($F(2, 46) = 1.90$, $p > .10$) and a significant difference in effect size between the two experiments ($Z = 2.88$, $p < .01$)[1]. Therefore, when no story was provided, the same triangle-dot situations as in Experiment 1 did not lead to a difference between the film conditions. Also the number of wrong responses differed only slightly between the conditions.

One could argue that the generally faster RTs in Experiment 2 did not allow for any perspective effect to take place. This seems rather unlikely given that the triangles, without the dot were always presented for at least 500ms and this should have been enough time for VSP adoption to take place. However, Experiment 2 rules out the alternative interpretation that the effect found in Experiment 1 was caused by low-level visual differences.

General Discussion

In Experiment 1, participants were presented with films that did or did not invoke agency attribution to a geometric figure. During films in which agency attribution occured participants were expected to adopt the VSP of the agent. Adoption of the triangle's perspective should lead to interference when the red triangle pointed downward and the triangle's and participant's perspectives differed but not when the triangle pointed upward. This conflict would require time to inhibit the inappropriate response of the adopted perspective. Similarly, when attribution of agency occurred in congruent conditions the activation of two congruent responses lead to faster RTs. The largest congruency effect was found in the ToM condition. Experiment 2 ruled out the possibility that the findings from Experiment 1 were caused by some superficial differences between visual properties among the three categories of films.

One interpretation of this could be that mental state attribution adds to the effect of agency attribution and therefore leads to stronger VSP-taking. An alternative interpretation is that the strength of agency attribution increased in the ToM compared to the GD condition. According to this interpretation VSP-taking is associated with agency attribution alone. The current experiments did not rule out this possibility. However, the former interpretation seems more likely because differences between ToM and GD films have been typically described in terms of mentalising but not agency (e.g., Abell et al., 2000). Further, even though a significant effect of VSP-taking was found for the GD but not for the R films, VSP-taking during GD films was not significantly larger than during R films. In contrast, VSP-taking was significantly larger during ToM films which supports the importance of mentalising for VSP-taking.

Response slowing in the incongruent GD and ToM conditions was not caused by disposing the perspective of the participant in total, because in this case no conflict would occur. In contrast, these results reflect the co-activation of two VSPs, the adopted perspective and the perspective of the participant.

The fact that VSP-taking occurred also during triangle animations shows that VSP-taking is not dependent on the presence of a human body as in Thomas et al. (2006) where no movements were shown. In the presence of agency cues VSP-taking in the current study occurred spontaneously, similar to Thomas et al. (2006) and Tversky and Hard (2009), even if it was not required by the task. The spontaneous involvement of VSP processes during the ToM condition suggests that mental perspective taking does not only involve abstract reasoning about the other's state but also entails adopting the actual visuo-spatial perspective. It will be a question of further research to determine what will be the minimal requirements of stimuli contents to evoke VSP-taking and whether the current films are a representative sample of these contents.

Acknowledgement

I would like to thank Uta Frith and Francesca Happé for making these animations available to me and Ulrike Hahn, Uta Frith, and two anonymous reviewers for helpful suggestions on an earlier version of the manuscript. I am also in debt to Jared Pool, and Katrin Rebekka Scharpf for helpful comments on the topic. This work was supported within the Deutsche Forschungsgemeinschaft (DFG) excellence initiative "CoTeSys".

References

Abell, F., Happe, 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://dx.doi.org/10.1016/S0885-2014%2800%2900014-9

Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *Neuroimage, 30*(3), 1059--1068.

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. 10.1016/j.cognition.2007.07.019

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage, 12*(3), 314-325. 10.1006/nimg.2000.0612

Frith, C. D., & Frith, U. (2006). How we predict what other people are going to do. *Brain Research, 1079*(1), 36-46. 10.1016/j.brainres.2005.12.126

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *Neuroimage, 24*(3), 771-779. http://dx.doi.org/10.1016/j.neuroimage.2004.09.006

Johnson, S. C. (2003). Detecting agents. *Philos Trans R Soc Lond B Biol Sci, 358*, 549-559. 10.1098/rstb.2002.1237

Klein, A., Zwickel, J., Prinz, W., & Frith, U. (2009). Animated triangles: An eye tracking investigation. *Quarterly Journal of Experimental Psychology, 62*(6), 1189-1197. 10.1080/17470210802384214

Rosenthal, R. (1997). *Meta-analytic procedures for social research*. London: Sage Publications.

Thomas, R., Press, C., & Haggard, P. (2006). Shared representations in body perception. *Acta Psychologica, 121*(3), 317-330. 10.1016/j.actpsy.2005.08.002

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition, 110*(1), 124-129. 10.1016/j.cognition.2008.10.008

Table 1

Mean (M) and Standard Errors (SE) of RT (ms) and Wrong Response Rate (Number of Wrong Responses Divided by the Number of Responses) are Reported Across Participants for the Three Film Categories (R, GD, and ToM) as a Function of Congruency (Incongruent, Congruent).

	RT				Wrong Response Rate			
	Incongruent		Congruent		Incongruent		Congruent	
Condition	M	SE	M	SE	M	SE	M	SE
				Experiment 1				
R	523	15	518	15	3.54	.76	2.33	.47
GD	538	16	525	16	10.08	1.47	5.83	.97
ToM	545	15	514	15	6.08	1.06	1.63	.42
				Experiment 2				
R	317	8	318	8	3.21	.55	3.08	.62
GD	321	8	318	9	3.21	.67	2.83	.56
ToM	323	8	324	8	2.67	.50	2.17	.45

Figures

Figure 1. Still pictures of the ToM film “Surprising” for different time points. The small triangle knocks on a door and hides behind it while the big triangle is looking out of the house. After the big triangle has moved in again, it repeats its knocking and hiding behavior, slips in and surprises the big triangle.

Figure 2. Examples of dot presentation in incongruent and congruent situations. In the picture shown on the left, the correct response would be “right” from the perspective of the participant, but “left” from the perspective of the triangle. In the picture on the right side, the correct response would be “right” from both perspectives.

Figure 3. Difference in mean RT between incongruent and congruent conditions as a function of film category. Whiskers indicate 95% confidence intervals.

Figure 1

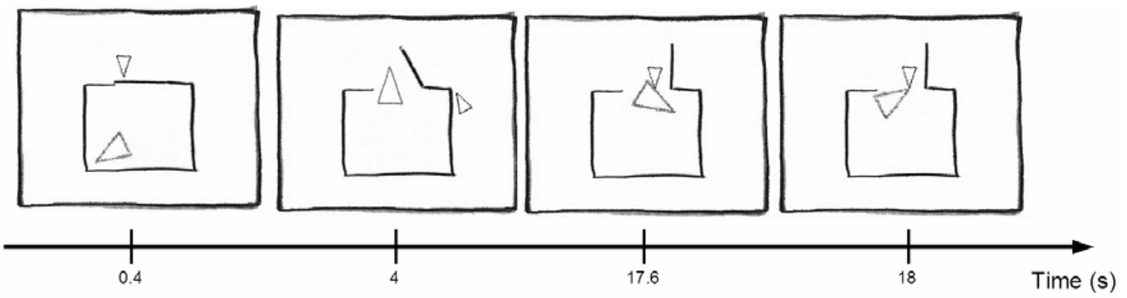


Figure 2

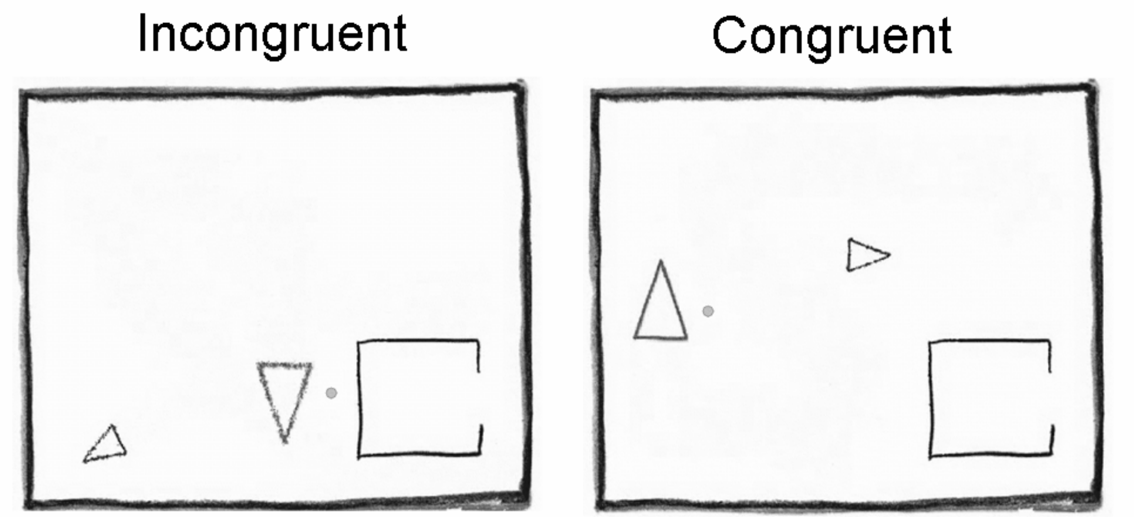
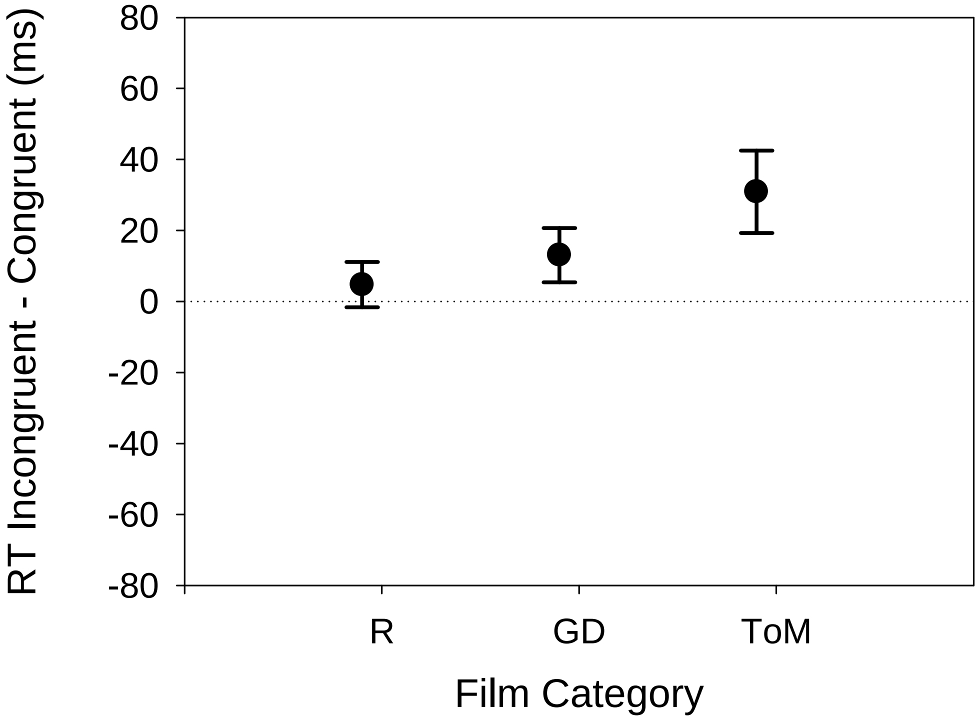


Figure 3,


80
60
40
20
0
-20
-40
-60
-80
RT Incongruent - Congruent (ms)
R
GD
ToM
Film Category

Footnotes

[1] Calculation was done as suggested in Rosenthal (1997).