Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Visualizing Theory of Mind with Multiple Intrinsic Frames of Reference

Permalink

https://escholarship.org/uc/item/1dn3z2fp

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 36(36)

ISSN

1069-7977

Authors

Liang, Chen
Sun, Yanlong
Wang, Hongbin

Publication Date

2014

Visualizing Theory of Mind with Multiple Intrinsic Frames of Reference

Chen Liang (chen.liang@uth.tmc.edu)
University of Texas Health Science Center at Houston
7000 Fannin Suite 600, Houston, TX 7030 USA

Yanlong Sun (ysun@tamhsc.edu)
Hongbin Wang (hwang@tamhsc.edu)
Texas A&M University Health Science Center
2121 West Holcombe Blvd, Suite 1109, Houston, TX 77030 USA

Abstract

Research in the field of spatial cognition has advocated a frame of reference (FOR) -based cognitive representation system to account for human's spatial reasoning and navigation capacities. It has been argued that such mental models may also contribute to the underlying mechanisms of Theory of Mind (ToM). In the present study, we investigated how people made rapid judgments about the number of visible objects from their own perspectives (egocentric frame of reference, EFOR) and from others' perspectives (intrinsic frame of reference, IFOR). We examined both behavioral and eye tracking responses, and the results suggest that a FOR-based representation system promotes the efficiency and flexibility of ToM functions. Our findings support the notion of a possible conceptual link between spatial and social cognitive processes.

Keywords: theory of mind; visual perspective taking; intrinsic frame of reference; eye movement

Introduction

Researchers have become increasingly interested in bridging the conceptual gaps between "Theory of mind" (ToM), which examines people's ability to judge other's intentions, beliefs and mental states (Frith & Frith, 2012), and spatial cognition, which examines people's ability to reason about spatial relationships and organize spatial representations. Frame of reference (FOR) has been used to account for the mechanisms underlying spatial cognition in terms of processing spatial representations and their relationships (May & Klatzky, 2000; Shelton & McNamara, 2001; Sun & Wang, 2010, 2014; Tamborello, Sun, & Wang, 2012; H. Wang, Johnson, & Zhang, 2001). In ToM judgments, investigations of egocentric visual perspective and allocentric visual perspective have also hinted at the relevance of a FOR-based internal representation system (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010; Samson, Apperly, Kathirgamanathan, & Humphreys, 2005).

Depending on the reference point of a spatial representation, three classes of FORs have been proposed: egocentric frame of reference (EFOR, self-centered), allocentric frame of reference (AFOR, world-centered), and intrinsic frame of reference (IFOR, anchored to other person or object) (Chen & McNamara, 2011; Sun & Wang, 2014; H. Wang, Johnson, Sun, & Zhang, 2005; R. Wang & Spelke, 2002). Multiple FORs are often used together when people judge complex spatial relationships. Neurological studies reveal distinct processes of generating segmented FOR-based internal representations and transformation among these representations (Colby, Duhamel, & Goldberg, 1995; Duhamel, Colby, & Goldberg, 1991; Samson et al., 2005). This permits a computational description of how the human brain processes spatial information in different contexts, including (1) coexistent, yet temporally discrete (merge in secession) representations, and (2) reconciliation and transformation between multiple representations in different FORs. For instance, when people judge spatial relationships from other's perspective (perspective taking), EFOR-based and IFOR-based representations would interact with each other in order to resolve potential incongruences (Kessler & Rutherford, 2010; Kessler & Thomson, 2010). Moreover, a growing body of research suggest that low-level spatial representations in spatial tracking, predictive encoding and attention shifting are essential in supporting sophisticated abilities during social interactions (Corbetta, Patel, & Shulman, 2008; Frith & Frith, 2012; Mitchell, 2006; Perner, Mauer, & Hildenbrand, 2011).

A FOR-based account of ToM may provide a fresh approach to understanding the intrinsic nexuses between ToM and spatial cognition. Particularly, one's ability to infer others' views, intensions, and beliefs occurs as one adopts others' perspectives (Gallagher & Frith, 2003). This process requires the inhibition of one's egocentric perspective, so as to make someone else's perspective more accessible (Samson et al., 2010). In a complex task environment, inhibitory competitions may exist not only between one's EFOR and IFOR representations, but also between multiple IFOR representations. In addition, active maintenance of multiple representations is required when the spatial layouts of the task environment are evolving dynamically over time (Morton & Munakata, 2002; Perner & Ruffman, 2005). Recently, it has been proposed that the maintenance of FOR-based representations is driven by expectation towards an efficient and flexible partitioning of the spatio-temporal statistics in the task environment (Sun & Wang, 2014). By comparing the results from a direction-pointing spatial task (Tamborello et al., 2012) and a false-belief task (Onishi & Baillargeon, 2005), Sun and Wang (2014) argued that FOR-based spatial representations are the common factor in both tasks. Therefore, spatial and social abilities may share a common origin at the level of spatio-temporal association and predictive encoding, and, FOR-based spatial representations may provide building blocks for general ToM abilities.

If FOR-based representations can indeed account for the performance in the false-belief task, it remains to be demonstrated that the same mechanism might be also at work in other types of ToM tasks, such as those involving number cognition. Here we report a study using a recently developed task, in which participants perform rapid ToM judgments from the perspectives of themselves or a computer-generated avatar (McCleery, Surtees, Graham, Richards, & Apperly, 2011; Samson et al., 2010). We modified the task so that conflicts and competitions may exist not only between the self-perspective (i.e., EFOR representation) and the avatar-perspective (i.e., IFOR representation), but also between different avatar-perspectives (i.e., multiple IFOR representations). This change allowed us to investigate more complex interactions between multiple conflicting FORs. More importantly, it provided a direct comparison between a typical ToM task and a spatial reasoning task so that we could examine the common spatial representations in both tasks.

Experiment 1

Methods

Participants Twenty-one participants (aged from 18 to 50, mean age: 30.2 years), composed of graduate students and staff from the University of Texas Health Science Center at Houston, were recruited to participate in the experiment. They received gift cards in return for their participation.

Stimuli The visual stimuli consisted of a picture showing two avatars (“Michael” and “Rachel”) standing in a room facing each other (Figure 1). Avatars’ relative positions exchanged randomly across trials. A certain number of red discs (randomly chosen from 1, 2, and 3) were displayed in the room such that two avatars would see either (1) the same number of discs (IFOR-IFOR consistent condition) or (2) different numbers of discs (IFOR-IFOR inconsistent condition).

Before the display of the visual stimuli, participants were prompted with a spoken sentence (by a male voice in English). Example sentences were, “Michael sees N” (50% of the trials) or “Rachel sees N” (50% of the trials), where N could be 1, 2, or 3 with equal probabilities. The spoken sentence either correctly (Match condition) or incorrectly (Mismatch condition) described the visual stimuli from the prompted avatar’s perspective. Half of the trials were matched and half of the trials were mismatched.

Procedure Each trial began with a fixation cross in the center of the screen. After 600 ms, a spoken sentence lasting around 2000 ms was presented. Following the spoken sentence, the screen maintained a fixation cross in the center with randomized duration of 150, 250, and 350 ms. Next, a probe picture showing a lateral view of the room appeared on the screen. The participants’ task was to indicate, as quickly and accurately as possible if the picture matched the spoken sentence they just heard, by pressing a response key. Once a response was made, either the picture disappeared and the trial ended or the trail ended after the time of limit of 1500 ms had been reached. The response time was recorded with the zero reading locked with the onset of the picture. The inter trial interval was 1000 ms.

Results

Data for one participant were removed from the analysis due to low accuracy (<50%). Twenty participants' data were included in the data analysis. Data for included participants showed an average accuracy of 92.7% (SD = 0.3%).

We conducted a repeated measure analysis of variance (ANOVA) with IFOR-IFOR Consistency (Consistent vs. Inconsistent) and Matching (Match vs. Mismatch) as the within subject variables and reaction time (RT) and accuracy as the dependent variables (Figure 2). Only trials with correct responses and response time within 1500 ms (98.2% of the total trials) were included in the analyses.

IFOR-IFOR Consistency × Matching interaction reached statistical significance, *F*(1, 19) = 9.02, p < .01, η_p^2 = .29. The ANOVA revealed a main effect of IFOR-IFOR Consistency, *F*(1, 17) = 274.30, p < .001, η_p^2 = .97, as well as a main effect of Matching, *F*(1, 18) = 35.67, p < .001, η_p^2 = .84. In the error analysis, time-out trials (1.8%) were counted as erroneous trials. The ANOVA revealed a main effect of IFOR-IFOR Consistency, *F*(1, 19) = 22.6, p < .001, η_p^2 = .48.

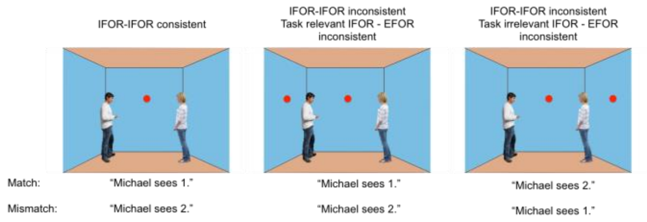

Figure 1: Examples of the visual stimuli used in the experiments. The visual stimulus (dimension: 504 × 315 pixels) was presented in the center of the computer screen (dimension: 1024 × 768 pixels) with a white canvas as background.

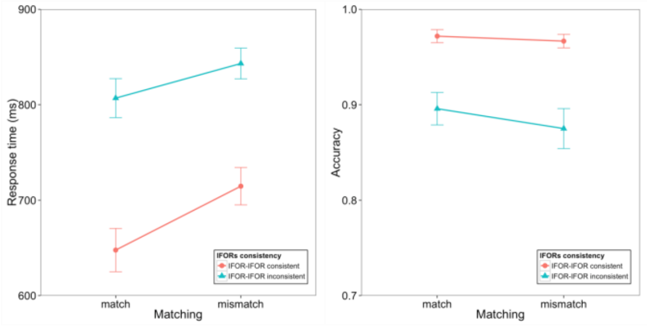

Figure 2: Experiment 1 response times (left) and accuracy (right) by IFOR-IFOR Consistency and Matching conditions. Error bars depict standard errors of the mean.

In order to investigate the effect of IFOR-EFOR conflicts on participants' RTs and accuracy, we conducted another repeated measures ANOVA. Note that previously, a 2 × 2 within subject design was used with IFOR-IFOR Consistency (Consistent vs. Inconsistent) × Matching (Match vs. Mismatch) as the independent variables. IFOR-EFOR conflicts appeared only in IFOR-IFOR inconsistent condition, in which half of the IFOR-IFOR inconsistent trials were in task relevant IFOR-EFOR conflict and second half in task irrelevant IFOR-EFOR conflict. With Matching (Match vs. Mismatch) remaining the same, we have four conditions: (1) Task relevant IFOR-EFOR Inconsistent - Match (TR-M), (2) Task relevant IFOR-EFOR Inconsistent - Mismatch (TR-MM), (3) Task irrelevant IFOR-EFOR Inconsistent - Match (TIR-M), and (4) Task irrelevant IFOR-EFOR Inconsistent - Mismatch (TIR-MM). Using this design, we employed another round of repeated measures ANOVA with the 2 × 2 array proposed above. See Figure 3.

The ANOVA revealed a main effect of IFOR-EFOR consistency, $F(1, 17) = 23.47$, $p < .001$, $\eta_p^2 = .94$. We also found a main effect of Matching, $F(1, 18) = 17.10$, $p < .001$, $\eta_p^2 = .50$. With respect to error analysis, we found a two-way interaction (IFOR-EFOR Inconsistency × Matching), $F(1, 19) = 12.39$, $p < .01$, $\eta_p^2 = .11$.

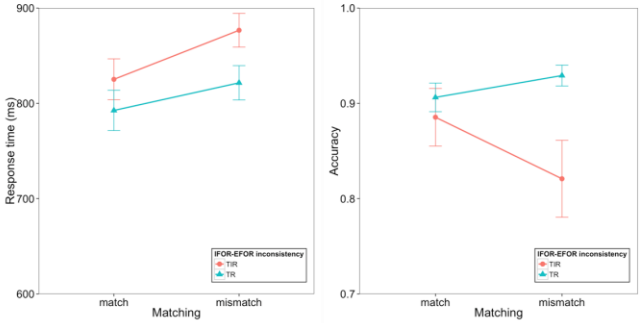

Figure 3: Experiment 1 response times (left) and accuracy (right) by IFOR-EFOR Consistency and Matching. TIR: Task irrelevant IFOR - EFOR consistent. TR: Task relevant IFOR - EFOR consistent. Error bars depict standard errors of the mean.

Discussion

For the present experiment, we examined the role of two IFORs when people encoded spatial representations and judged spatial relationships in a modified ToM task. Both RT and accuracy results revealed that the demanding cognitive processes came from the inconsistency between the two IFORs. More precisely, we found longer RTs as well as a higher percentage of inaccurate responses when the judgments about Michael's perspective differed from judgments about Rachel's perspective (e.g., IFOR-IFOR inconsistent condition). Our results indicate that people were able to process and maintain multiple avatars' perspectives. Note that in order to complete the task, it was not necessary for them to take the perspective of the avatars in the task-irrelevant IFOR. Yet, participants' performance was still influenced by the task irrelevant IFOR, despite the fact that the task relevant IFOR was explicitly prompted by the spoken sentence. This result confirms the previous finding that people process multiple IFORs, but with limited cognitive capabilities in handling conflicting IFORs (Tamborello et al., 2012).

In the second round of ANOVA, we separated two kinds of IFOR-EFOR conflicts from the IFOR-IFOR inconsistent condition. Longer response times were observed when EFOR was in conflict with the task irrelevant IFOR, as compared to when EFOR was in conflict with the task relevant IFOR. This finding suggests that participants might be influenced by their own visual experience (EFOR) even when they were instructed to judge what the avatar saw. Recent findings indicate that people experience difficulty inhibiting their own perspective when judging other's perspective (Samson et al., 2010). It is therefore likely that EFOR plays a somewhat dominant role in processing spatial representations. For example, people may encode external spatial information based on multiple FORs, yet depend on EFOR. One possibility is that EFOR plays a critical role in updating spatial representations based on the transformation between self (EFOR) and object (the most salient IFOR) (Kessler & Rutherford, 2010; Kessler & Thomson, 2010; Mou & McNamara, 2002; Zacks & Michelon, 2005). Regardless, we were unable to directly test this hypothesis because IFOR-EFOR conflicts were confounded with the IFOR-IFOR conflicts in this version of the task.

Interestingly, the behavioral patterns we observed in the second round of ANOVA differed from previous findings in terms of how the task irrelevant IFOR played a role in performance. Since they knew in advance which avatar (Michael or Rachel) was task relevant, presumably participants should have had longer response times when EFOR was in conflict with task-relevant IFOR but not with the task irrelevant IFOR. Hence it would have been more efficient for participants to have identified the task relevant IFOR as soon as visual stimulus appeared. One possibility is that people can detect multiple IFORs during an early stage of visual processing while their limited cognitive resources may only afford to support one or a few of the FORs to be further processed. In this scenario, we would expect to observe distinctive eye movement patterns in different conditions. This hypothesis was tested in Experiment 2.

Experiment 2

Methods

Fourteen participants (aged from 18 to 50, mean age: 31.6 years), composed of graduate students and staff from the University of Texas Health Science Center at Houston, were recruited to participate in the experiment. In order to ensure good eye-tracking results, participants were required to have either normal vision or corrected normal vision with contact lenses. Each participant received gift cards in return for their participation.

The design of Experiment 2 was the same as Experiment 1, except that participants' eye movements were recorded with a SmartEye 5.2 eye tracker (SmartEye AB, Gothenburg, Sweden). Each participant was seated approximately 50 cm in front of the computer screen, leading to a 22.6° visual angle for the entire visual stimulus image. The SmartEye program was run on a different computer than the computer running the experimental task. An in-house E-Prime package was used to synchronize the stimulus presentation and the eye tracker, which automatically collected and recorded eye fixations in real-time.

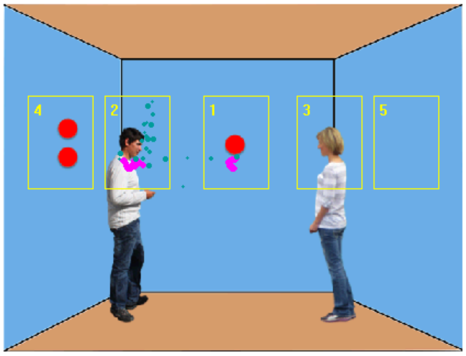

Figure 4: Areas of interest (AOI, marked by yellow rectangles) and an example of eye movements (green dots for saccades and pink dots for fixations). AOI 2 and AOI 3 cover the task relevant and task irrelevant avatars, respectively. AOI 1 covers the area that is visible to both avatars. AOI 4 and AOI 5 cover the areas that are visible to only one of the avatars (e.g., Michael cannot see objects displayed in AOI 4).

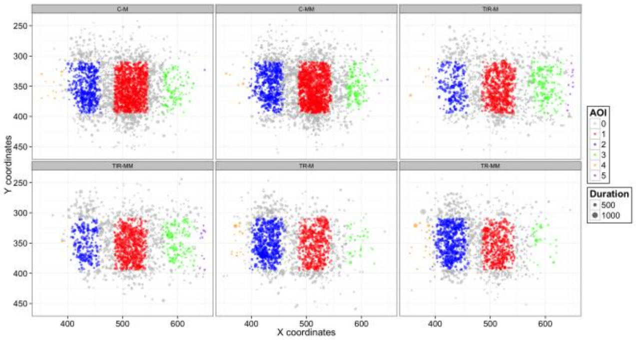

Figure 5: Scatter plots for the distribution and durations (in milliseconds) of eye fixations across six conditions: IFOR-IFOR Consistent - Match (C-M), IFOR-IFOR Consistent - Mismatch (C-MM), Task relevant IFOR-EFOR Inconsistent - Match (TR-M), Task relevant IFOR-EFOR Inconsistent - Mismatch (TR-MM), Task irrelevant IFOR-EFOR Inconsistent - Match (TIR-M), and Task irrelevant IFOR-EFOR Inconsistent - Mismatch (TIR-MM). AOI 0 indicates those fixations falling outsides AOI 1~5.

Results

The percentages of RT and accuracy outliers found in Experiment 2 data were in a similar range as those found in Experiment 1, and were therefore omitted here. In the following, we focus on the analyses of eye movement data. Figure 4 shows the five areas of interest (AOI) for fixation analyses, and Figure 5 shows the scatter plots of actual eye fixations aggregated over all trials in each condition.

To compare the eye fixations across conditions, we computed the mean fixation duration in each condition as the total time of fixations divided by the number of trials with correct responses in that condition (see Figure 6). We conducted a repeated measures ANOVA with the mean fixation duration as the dependent variable and AOI (AOI 1 to 5) and conditions (C-M, C-MM, TR-M, TR-MM, TIR-M, TIR-MM) as the within subject variables. The interaction between AOI and Condition reached significance, $F(17, 142) = 10.78$, $p < .001$, $\eta_p^2 = .49$. We found a main effect of AOI, $F(4, 29) = 7.37$, $p < .001$, $\eta_p^2 = .35$, and a main effect of Condition, $F(5, 49) = 79.34$, $p < .001$, $\eta_p^2 = .86$. Follow-up t-test indicated a significant difference between C-M and TIR-M on AOI 3, $t(13) = 9.39$, $p < .001$. Significant differences were also found for AOI 2 between C-M and TR-M, $t(13) = 19.72$, $p < .001$, and between C-MM and TR-MM, $t(13) = 20.27$, $p < .001$. Similarly, comparisons between TR-M and TIR-M on AOI 2 showed significant differences, $t(13) = 6.91$, $p < .001$, and comparisons between TR-MM and TIR-MM on AOI 2 also showed a significant difference, $t(13) = 4.30$, $p < .001$.

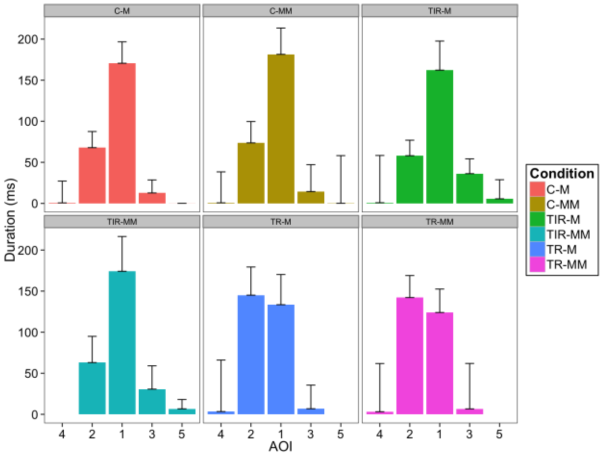

Figure 6: Duration of fixations attributed to individual AOIs. Error bars depict standard errors of the mean.

Figure 7 shows the mean duration for each AOI over the ordinal number of eye fixations. Note that the duration of fixations for AOI 3 for both the TIR-M and TIR-MM conditions projects an upward momentum on the second fixations, $t(13) = 2.06$, $p < .05$. Note also that for the second fixations, although the duration of fixations on AOI 2 reveals a clear upward momentum starting from the second fixations, $t(13)= 2.47$, $p < .05$, in TR-M and TR-MM conditions it shows an even more remarkable increment, $t(13) = 2.57$, $p < .05$.

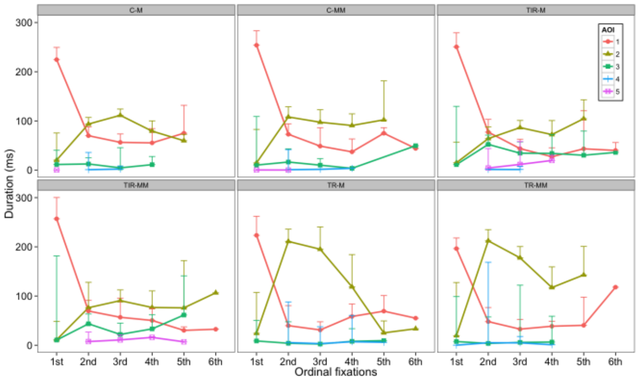

Figure 7: Mean duration of fixations over ordinal fixations. There were up to 6 fixations in each trial, listed in the chronological order on x-axis. Error bars depict standard errors of the mean.

Discussion

The results of Experiment 2 revealed that task relevant IFORs received significantly more eye fixations as compared with task irrelevant IFORs. Since participants were aware of the task relevant avatar prior to the appearance of the visual stimuli, searching for the task relevant avatar was necessary in order to optimize the performance. However, comparisons between conditions showed that when EFOR conflicted with task irrelevant IFOR, the task irrelevant avatar received more eye fixations than in the other conditions. When there were conflicts between EFOR and task relevant IFOR, task relevant avatars received more eye fixations. However, the considerable amount of eye fixations found for task irrelevant avatars suggests that participants did attend to those. Therefore, it is possible that participants detected the conflict between task irrelevant IFOR and EFOR.

By separating the duration of fixations in chronological order of fixations, it appears that the shift in eye fixations conformed to the same pattern as depicted in Figure 6. Interestingly, the observed pattern appears for the second fixations, which seem to follow the onset of the visual stimuli very closely, suggesting the possibility that participants were able to immediately detect the IFOR-EFOR conflicts and recognize the identity of the avatar. Note that in both Figure 6 and Figure 7, AOI 1 received significantly longer durations of eye fixations, most of which occurred at participants' first eye fixations. This finding was likely an artifact of the instruction to look at the center of the screen, which also happened to be the center of AOI 1, before the onset of the visual stimuli. Hence the second fixations permitted observation of participants' visual attention at an early stage of the task.

General Discussion

The results from the present investigation showed rapid encoding of segmented internal representations based on multiple FORs. These results strongly suggest that FOR-based spatial processing serves as a basis for ToM processing. It was also found that participants spent more time judging trials during which EFOR presented inconsistent information with the task irrelevant IFOR. This result is consistent with previous studies showing that people are unable to inhibit a third person's perspective when instructed to judge a ToM task from their own perspective (Samson et al., 2010). In our view, the conflicts between an individual’s perspective and a third person's perspective reflect the incompatibility between EFOR representations and IFOR representations. Moreover, when the task involved multiple IFORs, resolving the incompatibility between IFORs and that between IFOR and EFOR contributed significantly to the overall performance.

The intricate interactions between multiple IFORs and EFOR representations found in the present study suggest possible functional processes by which people abstract complex spatial information from the task environment. In particular, we found that perspective taking could be triggered by both task relevant and irrelevant IFORs (e.g., slower response times in the task irrelevant IFOR-EFOR inconsistent condition). Consider that participants had been given an audio prompt of the task relevant avatar before the visual stimuli, and the eye fixation results showed that participants’ attention was engaged onto the task relevant avatar fairly early in the task. Together, these observations indicated that participants could have spontaneously established multiple IFOR representations even before their visual attention was completely shifted towards the task relevant avatar. As a result, perspective taking could take place almost simultaneously during the process of identity recognition.

In closing, we claim that the FOR-based representations in both spatial and social cognitive processing may offer a viable alternative explanation concerning whether the domain-general ToM abilities are supported by a separate and innate system or intertwined with domain-specific spatial abilities. In our view, complex and abstract cognitive achievements such as number cognition and theory of mind may nevertheless rest on a set of fundamental processes by spatial processing and spatio-temporal association (Sun & Wang, 2014). That is, different sets of cognitive abilities may not be domain specific per se. Rather, given their common low-level substrates, they are constrained by the statistical structures of the task environment and subject to the competing demands of computational efficiency and flexibility. In the effort of partitioning the variances in the environmental statistics, the internal representations evolve by first developing FOR-based representations, and then, encoding the achieved invariance at different levels of abstraction. Since the statistical structures include not only the spatial relations between static configurations but also the temporal relations between sequential events, predictive encoding serves the key to integrating and selecting various representations. Together, abstract representations of the task environment would eventually emerge from the competitions among multiple FOR-based spatial representations.

Acknowledgments

This work was supported by the Office of Naval Research (ONR) grant number N00014-08-1-0042, and Intelligence Advanced Research Projects Activity (IARPA) via Department of the Interior (DOI) contract number D10PC20021. The US Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI, or the US Government. We thank for Dr. Paul J. Schroeder and Ping Chen for their helpful comments.

References

Chen, X., & McNamara, T. (2011). Object-centered reference systems and human spatial memory. *Psychonomic Bulletin & Review, 18*(5), 985-991. doi: 10.3758/s13423-011-0134-5

Colby, C. L., Duhamel, J. R., & Goldberg, M. E. (1995). Oculocentric spatial representation in parietal cortex. *Cerebral Cortex, 5*(5), 470-481.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron, 58*(3), 306-324. doi: 10.1016/j.neuron.2008.04.017

Duhamel, J., Colby, C. L., & Goldberg, M. E. (1991). Congruent representations of visual and somatosensory space in single neurons of monkey ventral intra-parietal cortex (area VIP) *Brain and space*. NY: Oxford University Press.

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology, 63*(1), 287-313. doi: 10.1146/annurev-psych-120710-100449

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences, 7*(2), 77-83.

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

Kessler, K., & Thomson, L. A. (2010). The embodied nature of spatial perspective taking: Embodied transformation versus sensorimotor interference. *Cognition, 114*(1), 72-88.

May, M., & Klatzky, R. L. (2000). Path integration while ignoring irrelevant movement. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 169.

McCleery, J. P., Surtees, A. D., Graham, K. A., Richards, J. E., & Apperly, I. A. (2011). The neural and cognitive time course of theory of mind. *The Journal of Neuroscience, 31*(36), 12849-12854. doi: 10.1523/JNEUROSCI.1392-11.2011

Mitchell, J. P. (2006). Mentalizing and Marr: An information processing approach to the study of social cognition. *Brain Research, 1079*(1), 66-75. doi: 10.1016/j.brainres.2005.12.113

Morton, J. B., & Munakata, Y. (2002). Active versus latent representations: A neural network model of perseveration, dissociation, and decalage. *Developmental Psychobiology, 40*(3), 255-265. doi: 10.1002/dev.10033

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 162-170. doi: 10.1037//0278-7393.28.1.162

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255-258. doi: 10.1126/science.1107621

Perner, J., Mauer, M. C., & Hildenbrand, M. (2011). Identity: Key to children's understanding of belief. *Science, 333*(6041), 474-477. doi: 10.1126/science.1201216

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science, 308*(5719), 214-216. doi: 10.1126/science.1111656

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5), 1255.

Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain, 128*(Pt 5), 1102-1111. doi: 10.1093/brain/awh464

Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. *Cognitive Psychology, 43*(4), 274-310. doi: 10.1006/cogp.2001.0758

Sun, Y., & Wang, H. (2010). Perception of space by multiple intrinsic frames of reference. *PLoS ONE, 5*(5), e10442. doi: 10.1371/journal.pone.0010442

Sun, Y., & Wang, H. (2014). Insight into others' minds: Spatio-temporal representations by intrinsic frame of reference. *Frontiers in Human Neuroscience, 8:58*. doi: 10.3389/fnhum.2014.00058

Tamborello, F. P., Sun, Y., & Wang, H. (2012). Spatial reasoning with multiple intrinsic frames of reference. *Experimental psychology, 59*(1), 3-10. doi: 10.1027/1618-3169/a000119

Wang, H., Johnson, T. R., Sun, Y., & Zhang, J. (2005). Object location memory: The interplay of multiple representations. *Memory & cognition, 33*(7), 1147-1159.

Wang, H., Johnson, T. R., & Zhang, J. (2001). *The mind's views of space.* Paper presented at the Proceedings of the Third International Conference of Cognitive Science.

Wang, R., & Spelke, E. (2002). Human spatial representation: insights from animals. *Trends in Cognitive Sciences, 6*(9), 376.

Zacks, J. M., & Michelon, P. (2005). Transformations of visuospatial images. *Behavioral and Cognitive Neuroscience Reviews, 4*(2), 96-118. doi: 10.1177/1534582305281085.