frontiers in
HUMAN NEUROSCIENCE

HYPOTHESIS AND THEORY ARTICLE
published: 14 February 2014
doi: 10.3389/fnhum.2014.00058

Insight into others' minds: spatio-temporal representations by intrinsic frame of reference

***Yanlong Sun*[1] * and *Hongbin Wang*[2] ***

[1] *The University of Texas Health Science Center at Houston, Houston, TX, USA*
[2] *Center for Biomedical Informatics, Texas A&M University Health Science University, Houston, TX, USA*

Edited by:
Klaus Kessler, Aston University, UK

Reviewed by:
Klaus Kessler, Aston University, UK
Christian Sorg, Klinikum rechts der Isar der Technischen Universität München, Germany

***Correspondence:**
Yanlong Sun, The University of Texas Health Science Center at Houston, 7000 Fannin Suite 600, Houston, TX 77030, USA
e-mail: Yanlong.Sun@uth.tmc.edu;
Hongbin Wang, Center for Biomedical Informatics, Texas A&M University Health Science University, 2121 West Holcombe Blvd., Suite 1109, Houston, TX 77030, USA
e-mail: hwang@tamhsc.edu

Recent research has seen a growing interest in connections between domains of spatial and social cognition. Much evidence indicates that processes of representing space in distinct frames of reference (FOR) contribute to basic spatial abilities as well as sophisticated social abilities such as tracking other's intention and belief. Argument remains, however, that belief reasoning in social domain requires an innately dedicated system and cannot be reduced to low-level encoding of spatial relationships. Here we offer an integrated account advocating the critical roles of spatial representations in intrinsic frame of reference. By re-examining the results from a spatial task (Tamborello et al., 2012) and a false-belief task (Onishi and Baillargeon, 2005), we argue that spatial and social abilities share a common origin at the level of spatio-temporal association and predictive learning, where multiple FOR-based representations provide the basic building blocks for efficient and flexible partitioning of the environmental statistics. We also discuss neuroscience evidence supporting these mechanisms. We conclude that FOR-based representations may bridge the conceptual as well as the implementation gaps between the burgeoning fields of social and spatial cognition.

Keywords: theory of mind, false belief, spatial cognition, frame of reference, predictive learning

INTRODUCTION

Recent research has seen a growing interest in the connections between two disparate lines of investigations: spatial cognition that focuses on spatial and bodily representations, and, social cognition that examines the abilities of attributing other's intentions and beliefs, namely, theory of mind (TOM). Although researchers have learned much about the underlying mechanisms in each domain, there are still opposing perspectives and considerable conceptual gaps between the two domains. In particular, much contest revolves around the contribution of domain-specific spatial processing to domain-general TOM abilities.

At the center of the debate, is an apparent contradiction between the findings that human infants can pass false-belief tasks (e.g., holding an agent's belief about the original location of an object, which has been changed in the absence of the agent) and the general claim that children first understand false-beliefs at around 4 years of age (for reviews, see, Apperly and Butterfill, 2009; Perner et al., 2011; Frith and Frith, 2012). Some have suggested that sophisticated TOM inferences, as indicated by successfully performing the false-belief tasks, may evolve from a set of low-level encoding processes, for example, agent-object-location associations (Perner and Ruffman, 2005; Ruffman and Perner, 2005), identification of "external referent" (Perner et al., 2011), and, spatial perspective taking (Kessler and Rutherford, 2010; Kessler and Thomson, 2010). Yet other theorists have posited that beliefs are "invisible abstract entities" (Saxe, 2006), and that making inferences about other's beliefs requires a dedicated or innate system that cannot be accounted for by mere associations (Leslie, 2005; Saxe and Wexler, 2005; Csibra and Southgate, 2006; Baillargeon et al., 2010).

In the present paper, we attempt to bridge the conceptual gaps between different perspectives by advocating an integrated account. We argue that a fundamental spatio-temporal association process, which is fraught in the domain of spatial cognition, is also essential in the domain of social cognition. At the computational level, spatio-temporal association is to extract statistical regularities from the task environment by detecting the correlations between representations of events over space and time. However, spatio-temporal association is not merely about matrices of associative weights that connect different representations in a static manner. Instead, it takes place over space and time through the lens of *predictive learning*. Recent advances in neuroscience suggest that – at both the algorithmic and neural architectural levels – it is not reward that drives learning *per se*, but the temporal discrepancy between actual and expected outcomes (Gerstner et al., 2012; O'Reilly et al., 2012). That is, the task environment constantly changes. At any moment, environmental statistics present themselves as multimodal inputs to the mind. By constantly comparing the observed and expected outcomes, the mind selectively re-encodes the raw environmental statistics and transforms them into a hierarchy of representations at different levels of abstraction, which eventually produce complex behaviors such as thought, language, and, intelligence (Hawkins and Blakeslee, 2004).

Our approach to understanding the process of spatio-temporal association utilizes frames of reference (FOR) as the building blocks of both spatial and social cognition. A growing body of

research has shown that FOR-based representations are not only behaviorally plausible but are also supported by the neurological structures in both human and animal brains. As spatio-temporal association re-encodes the environmental statistics by removing task-irrelevant variances (e.g., instability, noise), FOR-based representations provide a straightforward way of partitioning spatio-temporal variances. In addition, it has been a central contention that theory-of-mind abilities are subject to competing demands for efficient and flexible processing and require two distinct systems, "one that is efficient and inflexible and one that is flexible but cognitively demanding" (Apperly and Butterfill, 2009, p. 957). Instead of focusing on the distinction between different systems, we emphasize the common representations shared by different sets of abilities and mechanisms. We argue that when people perform spatial and social tasks, both efficiency and flexibility can emerge from the expectation-driven competition among multiple FOR-based representations.

INTRINSIC FRAME OF REFERENCE (IFOR) IN SPATIAL COGNITION

The notion of "FOR" has been crucial to all the disciplines that study spatial relationships and relies on a diverse terminology for its classification (Levinson, 2004). For example, a conventional approach is to classify a reference system by its origin: whether it is anchored to the observer self (e.g., "egocentric") or the environment (e.g., "allocentric"; Andersen et al., 1997; Wang and Spelke, 2002; Burgess, 2008). However, we adopt a classification system that – besides the self-centered egocentric frame of reference (EFOR) – further differentiates the environment-centric frames into two categories: *allocentric* (AFOR, with an absolute and fixed anchor), and, *intrinsic* (IFOR, with a relative and flexible anchor). With roots in psycholinguistic research, the advantage of this classification scheme is that it reduces ambiguity in spatial descriptions of the world (Miller and Johnson-Laird, 1976; Carlson-Radvansky and Irwin, 1993; Levelt, 1996; Levinson, 2004; Carlson and Van Deman, 2008). For example, when describing the location of a coffee cup, one may say, "the cup is in front of me (observer self)" (in EFOR); "the cup is on the desk" (in AFOR); or "the cup is in front of John" (in IFOR). Note that, while both AFOR and IFOR use an external anchor, the anchor in AFOR (the desk in this case) is more stable than IFOR (John in this case, who can freely change his location or orientation). Our interest in IFOR is motivated by vision and spatial memory research that emphasizes the dynamic updating of object-centered representations (Marr, 1982; Wang et al., 2005a; Mou et al., 2008; Sun and Wang, 2010; Chen and McNamara, 2011). In this respect, the interactions between EFOR and IFOR (e.g., the intertwined representations of self-other-object relationship) are ubiquitous in everyday tasks, where the "other" can be either an anchoring object (Wang et al., 2005b; Tamborello et al., 2012), or another agent or human being as in social situations (Mitchell, 2006; Kessler and Rutherford, 2010; Kessler and Thomson, 2010; Perner et al., 2011).

One fundamental distinction among different FOR-based representations is the manner in which each representation handles *temporal instability* during the interactions between the mind and the environment. Temporal instability manifests itself as both spatial and temporal variances during the encoding of spatio-temporal relationships between various entities in the environment (e.g., self, agents, objects, locations, and events). Different reference systems partition these variances in different manners and therefore afford structures at different levels of instability. In the "coffee cup" example, the spatial relations among relevant entities can change over time. To locate the coffee cup, an EFOR representation from the observer's perspective is relatively stable, to the extent that the anchor is always the "observer self." In contrast, an IFOR representation of the coffee cup anchored to John is unstable because John can freely move around and the observer is therefore required to track both the coffee cup and John in order to maintain an IFOR representation.

Critically, temporal instability evokes *predictive learning*. Simply put, whereas temporal instability means that the current input is expected to change at the next time point, predictive learning is a process of spatio-temporal integration in which the internal representation is constructed by remapping attention toward the expected outcomes (Hawkins and Blakeslee, 2004; O'Reilly et al., 2012). It has been suggested that predictive learning is a driving force in learning structured abstractions of the environment (Hawkins and Blakeslee, 2004; Krauzlis and Nummela, 2011; Rolfs et al., 2011; Gerstner et al., 2012; O'Reilly et al., 2012). Consider the coffee cup example again: predictive learning takes the anticipated movements into consideration and produces a dynamic representation of the relevant spatial relations. When an observer is reaching for a coffee cup, predictive learning occurs within EFORs, such that the coffee cup's location is updated relative to the observer's hand or body. By making constant predictions, the observer would know when to grab even before her hand touches the cup. When the observer watches John reaching for the coffee cup, predictive learning involves IFORs, such that the coffee cup's location is updated relative to John. Yet, should John suddenly change his course and pick up another object (e.g., a stapler), the observer would be surprised as John's initial movements led to an expectation that he would pick up the coffee cup instead of the stapler.

That the mind uses different FOR to manage temporal instability and drive spatio-temporal association is consistent with an accumulating body of neurological and behavioral studies (Marr, 1982; Krauzlis and Nummela, 2011; Pertzov et al., 2011; Rolfs et al., 2011; Van Der Werf et al., 2013). To further illustrate this notion, consider an example from the two-cannon experiment reported by Tamborello et al. (2012). In their experiment (**Figure 1**), participants were instructed to use the arrow keys to rotate the cannon in the same color of a to-be-revealed target as quickly as possible, so that the cannon could point to (and shoot at) the target. Three different types of reference systems can be used to describe the target location (**Figure 1A**). In an EFOR representation (relative to the observer), the target is at the front-top of the observer's visual field (the observer's line of sight was perpendicular to the plane of the computer screen). In an AFOR representation, the target can be described in reference to the computer screen frames. In an IFOR representation (relative to a cannon), the target has a counterclockwise bearing relative to the orientation of the blue cannon (or a clockwise bearing relative to the red cannon). Mathematically, all of these representations are equivalent, to the extent that one representation can be transformed into another without

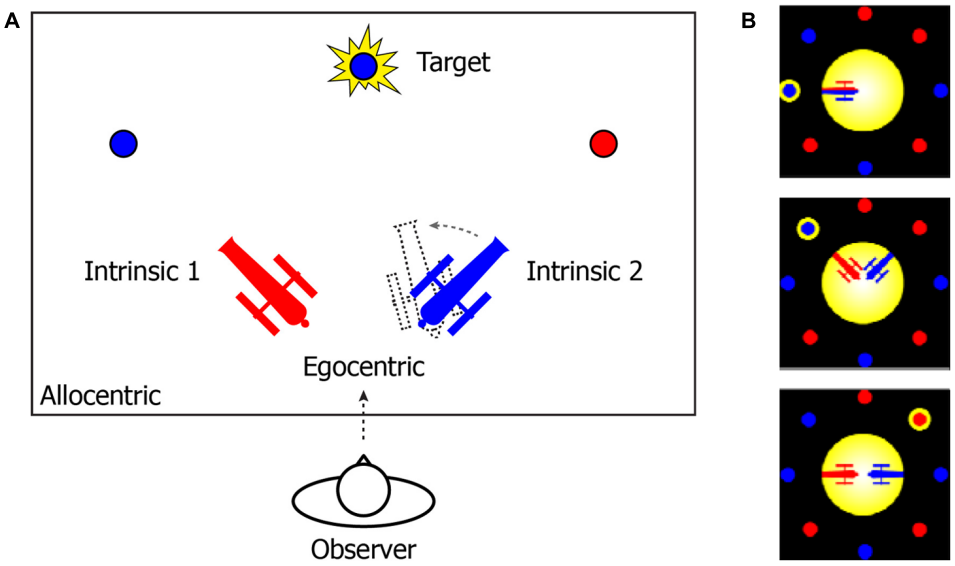

FIGURE 1 | (A) A schematic illustration of the two-cannon task reported by Tamborello et al. (2012). In the experiment, participants ("observers") were sitting in front of a computer (the observer's line of sight was perpendicular to the plane of the computer screen). **(B)** Actual task displays in the experiment. At the beginning of each trial, two cannons (one red and one blue) and multiple pellets (in either red or blue) were presented together on the computer screen. After a one-second pause, a randomly selected pellet would flash as the target. The participants' task was to use the arrow keys to rotate the cannon in the same color of the target toward the target as quickly as possible. The cannons' orientations and the ratio between the number of red and blue pellets were varied across trials. The angles between two cannons were either "conflict-absent" (zero degree) or "conflict-present" (90 or 180°).

losing any information. However, in terms of efficient and flexible removal of task-irrelevant variance, different representations are unique in the way they are updated and maintained.

Let us first examine temporal instability. It is clear that both EFOR and AFOR representations have relatively fixed anchors (e.g., the observer and the computer monitor frames, respectively). In contrast, IFOR is only *tentatively* anchored to one of the two cannons: the color and location of the target is initially unknown, thus, which cannon is task-relevant depends on the visual input at the next time point. Recall that temporal instability evokes predictive learning, in which internal representations of the environment are constructed based on the current observations toward the expected future outcomes. In this case, the color ratio of the pellets provides a reliable cue for predicting the relevancy between two competing cannons. **Figure 2A** shows that reaction times in the conflict-present condition (cannons pointing to different directions) were significantly slower than those in the conflict-absent condition (cannons pointing to the same direction). Within the conflict-present conditions, the cannon in the same color of the majority pellets resulted in faster reaction times. These results indicate that in resolving the conflict between different IFOR representations, participants planned their responses by predicting the task-relevant cannon based on the pellet color ratio. That is, prediction occurs before the appearance of an actual target, leading to a stronger IFOR representation anchored to the task-relevant cannon, thus resulting in faster reaction times.

Second, in order to achieve computational efficiency and flexibility, multiple IFOR representations may coexist and interact with each other. **Figure 2A** shows that even when participants made correct predictions on the task-relevant cannon in the conflict-present condition, their reaction times were still significantly slower than that in the conflict-absent condition. This indicates that, while anticipating the upcoming target, the competition between two conflicting IFOR representations resulted in a partial dissociation. That is, as the IFOR representation anchored to the predicted task-relevant cannon was the focus of attention, the other one was only partially disengaged – a strategy of prioritizing but still preparing for the unexpected. As a result, even when the prediction was correct, the partially disengaged IFOR representation interfered with performance and produce longer reaction times.

Third, an interaction may also occur between EFOR and IFOR representations. **Figure 2B** shows that reaction times were significantly dependent on the angular disparity between the self and cannon orientations, indicating a strategy of combining EFOR and IFOR representations, or *perspective taking*. Perspective taking has been considered as an important stepping stone from automatic and unaware perception toward a conscious and deliberate process in which people mentally perform a movement simulation of other people or objects (Kessler and Rutherford, 2010; Kessler and Thomson, 2010; Zwickel et al., 2011). Here, we consider perspective taking in terms of partitioning the statistical variances in the task environment.

Specifically, for a given cannon, we consider three parts of the spatial variances (angular disparities) that could be mentally encoded: self-cannon, self-target, and cannon-target. Since the correct response is determined by the cannon-target variance, it requires either a complete or a partial disengagement of the EFOR representation. If the EFOR representation is to be completely disengaged (i.e., removing self-target and self-cannon variances), the task could be accomplished by *object rotation* based

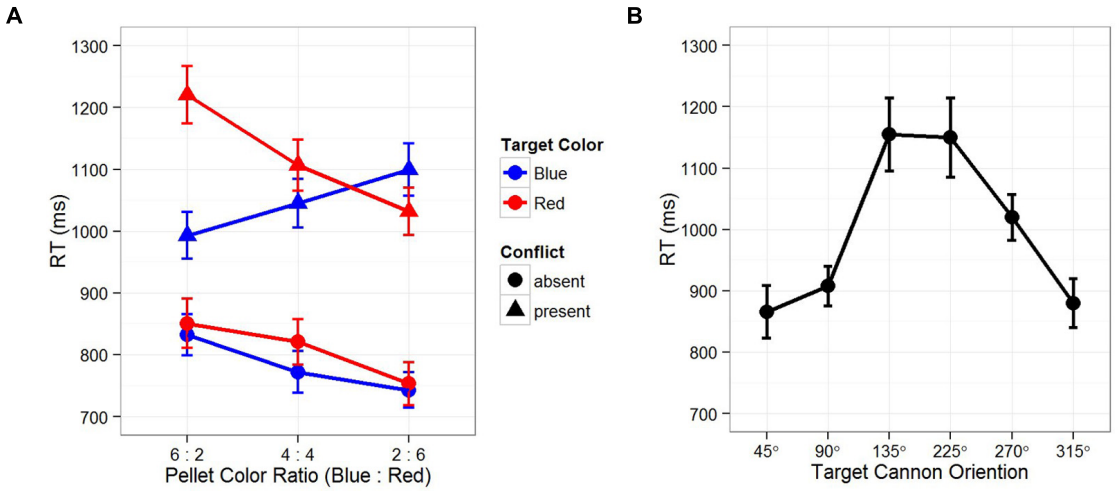

FIGURE 2 | Results from Tamborello et al. (2012). (A) Reaction times as function of target color, conflict between two cannons, and the color ratio in surrounding pellets. Across three sequentially presented blocks of trials, the surrounding pellets varied from trials of more blue pellets (B:R = 6:2) to trials of more red pellets (B:R = 2:6). Reaction times in the "conflict-absent" conditions (two cannons pointing to the same direction) were significantly faster than that those in the "conflict-present" conditions (90 or 180° between two cannons). Within the "conflict-present" conditions, reaction times were significantly faster for the cannon in the same color of the majority pellets, indicating the effect of expectation, where participants had made predictions on the task-relevant cannon before the appearance of the actual target. **(B)** Reaction time was dependent on the angular disparity between the participants' "up" and the target cannon orientations (self-cannon variance), indicating an interaction between EFOR and IFOR representations, namely, the effect of perspective taking. Note that since participants were always facing the computer screen, their "up" was congruent with the "up" on the computer screen. In both figures, error bars depict standard error of the mean.

only on an IFOR representation. However, the reaction time pattern in **Figure 1B** suggests a case of partial EFOR disengagement: the task was accomplished by *self rotation with perspective taking*, in which the self-cannon variance was first removed so that the self-target variance became exactly the same as the cannon-target variance. Similar to the interaction between multiple IFOR representations, the interaction between EFOR and IFOR representations also serves the purpose of both computational efficiency and flexibility. On the one hand, an IFOR representation is parsimonious in encoding only task-specific variances (e.g., encoding only the target-cannon but not the self-cannon, the self-target relations). On the other hand, an EFOR representation tend to be automatic and effortless (Wang and Spelke, 2002; Frith and Frith, 2007; Kessler and Thomson, 2010). Therefore, an efficient and flexible solution would be to combine EFOR and IFOR representations into one representation. That is, instead of utilizing a purely IFOR-based strategy in which the cannon is mentally rotated toward the target (i.e., object rotation), participants might superimpose their egocentric perspective onto the cannon – that is, take the perspective of the cannon – then mentally self-rotate toward the target.

Overall, this new interpretation of the two-cannon experiment results suggests that expectation-driven competitions can take place not only between different IFOR representations (**Figure 2A**), but also between EFOR and IFOR representations (**Figure 2B**). By this account, the internal spatial representation of the environment is always dynamically constructed and updated toward the anticipated outcomes, rather than static associations of the current spatial configuration. Depending on whether there are conflicts between representations and whether the actual outcome meets the expectation, competition takes place at different levels and results in the engagement and disengagement of different FOR-based representations. In the following section, we demonstrate that the same mechanisms may well lay the foundation for more complex representations in the domain of social cognition.

INTRINSIC FRAME OF REFERENCE IN BELIEF ATTRIBUTION

A landmark finding in belief attribution is that fifteen-month-old infants appear to be able to appeal to other's beliefs, that is, they were able to keep track of an actor's perception about the location of a toy, and, using this perception rather their own, to predict the actor's searching behavior (Onishi and Baillargeon, 2005). This finding has triggered a substantial debate over the question whether the theory-of-mind abilities evolved from "actor-object-location associations" (Perner and Ruffman, 2005, p. 215), or are due to an innate mechanism specialized for belief attribution (Leslie, 2005; Baillargeon et al., 2010). Here we offer a reinterpretation of the original findings based on the same spatio-temporal association account outlined above.

Figure 3 re-produces the experimental setup and results from Onishi and Baillargeon (2005). Note that we have re-labeled the experimental conditions by replacing the original object labels with location labels from the actor's perspectives: "green box" replaced by "L" (actor's left-hand side), and, "yellow box" replaced by "R" (actor's right-hand side). Hence, our new labels are essentially placeholders for representing different locations. However, the new labels also highlight the spatial component of the task environment and potential interference between the different FOR. Similar to the two-cannon experiment, this task involves the

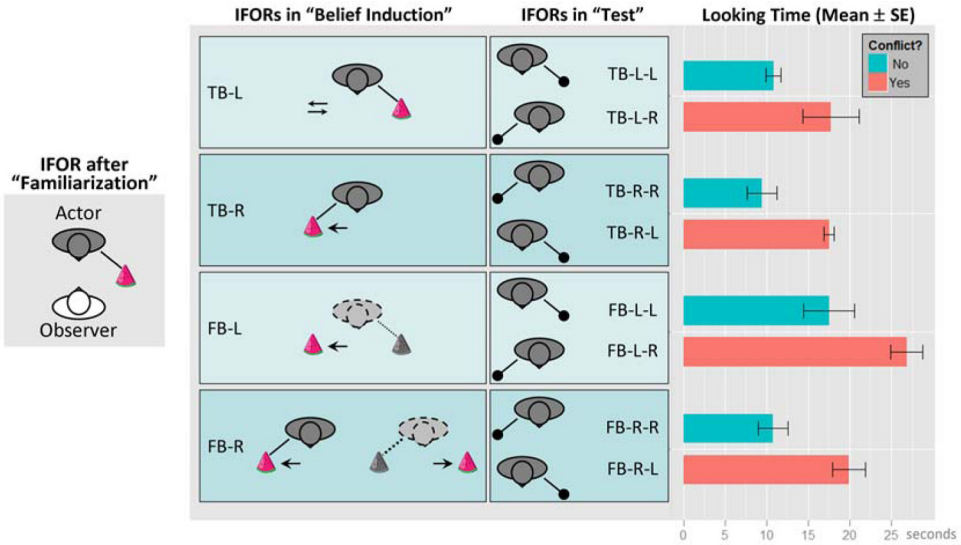

FIGURE 3 | The experimental setup and results, re-produced based on Onishi and Baillargeon (2005). Conditions have been renamed by replacing the original labels "green box" and "yellow box" with location labels "L" and "R", respectively ("L" and "R" indicate the toy's location from the actor's perspective). The experiment consisted of three phases: (1) "familiarization", (2) "belief induction", and (3) "test". During (1), infants ("observer") watched the actor reaching toward a box for a toy at one of two locations (boxes are not shown here). At the end of this phase, the toy was located on the actor's left-hand side. In (2), infants were assigned to one of four conditions, in which they watched some movements of the boxes or the toy in the actor's presence or absence. Here we used dyadic labels to represent the validity of the actor's belief ("TB" for true belief and "FB" for false-belief) and the location of the toy last known to the actor from the actor's perspective ("L" for the left-hand side and "R" for the right-hand side). In addition, arrows represent movements of the box or the toy; colored toy and solid lines indicate the actor's true belief about the toy's location; grayed toy and dotted lines represent the actor's false-belief as the location of the toy was changed in her absence. In the "TB-L" condition, the toy remained at the actor's left-hand side and only the box at the actor's right-hand side was moved toward the toy then back to its original location. In the "TB-R" condition, the toy was moved from the actor's left to her right in her presence. In the "FB-L" condition, the toy was last seen by the actor at her left but was moved to her right in her absence. In the "FB-R" condition, the toy was moved from the actor's left to her right in her presence but moved back to her left in her absence. In test phase (3), infants watched the actor reaching one of the locations for the toy and their average looking times were recorded and analyzed. Here we use triadic labels to represent each test condition, with the first two parts repeating the label for the corresponding belief induction condition, and the last part representing the direction where the actor reached for the toy. The equality between the last two parts represents whether there is a conflict between the IFOR representation at the end of the belief induction phase and the one in the test phase. For example, "TB-LL" represents the condition in which the actor held a true belief that the toy was at her left-hand side and she actually reached the same location for the toy ("no conflict"); In comparison, "TB-LR" represents the condition in which the actor held a true belief that the toy was at her left-hand side but she actually reached her right-hand side for the toy ("conflict"). Infants' looking times (mean and standard errors in seconds) in each test condition are shown on the rightmost panel.

interplay of multiple representations. For example, the toy's location can be described in EFOR (relative to the observer, which is the infant in the experiment), AFOR (relative to the table or the room), or IFOR (relative to the actor). According to the original object labels, the toy's location was described by the color of the box, which was the same to both the infant and the actor. In contrast, as the infant was facing the actor, the "left" and "right" labels were completely opposite, depending on whether they were from the infant's perspective (EFOR) or from the actor's perspective (IFOR). Therefore, the new labels were more effective in distinguishing EFOR and IFOR representations.

COMPARISON WITHIN BELIEF INDUCTION CONDITIONS

The main finding by Onishi and Baillargeon (2005) involved comparing the infants' looking times between the two "test" conditions within each of the four "belief induction" conditions. They reported that looking times were shorter when the actor reached for the toy where she believed it was located ("no conflict" conditions in **Figure 3**) and longer when the actor reached the opposite location ("conflict" conditions). Based on this comparison, the authors concluded that infants were able to use the actor's belief state instead of the actual toy location from infants' own perspective to predict the actor's reaching behavior.

Rather than resorting to an innately dedicated belief attribution mechanism, we would like to offer a different explanation based on fundamental spatial information processing mechanisms. Our interpretations is that belief attribution derives from the proper maintenance of and dissociation between multiple representations based on EFOR (for encoding self-toy or self-actor relations) and IFOR (for encoding actor-toy relations). In particular, it has been suggested that infants' looking time provides a measurement of surprise, such that longer looking times indicate greater violation of infants' expectations relative to their prior knowledge or greater novelty relative to their interpretation of habituation stimuli (Baillargeon, 1986; Onishi and Baillargeon, 2005; Téglás et al., 2011). Here we argue that for the false-belief task by Onishi and Baillargeon (2005) surprise might have resulted from the violation of infant's expected spatial configuration relative to the actual one.

Our earlier argument suggests that, among all possible FOR-based representations, those leading to task-relevant predictions tend to be actively updated and maintained. Since the looking times were about the actor's reaching for the toy, both the expected and actual spatial configurations would be encoded in the form of IFOR representations (actor-toy), rather than irrelevant EFOR representations (infant-toy). In other words, the IFOR-based expectation reflects a simple behavioral rule by means of spatial association – people (the actor) look for objects at their last known location (Ruffman and Perner, 2005). Consequently, the difference in looking times between "conflict" and "no conflict" conditions may be explained by the effort of resolving the discrepancy between the IFOR representation at the end of the belief induction phase, relative to the actual IFOR representation in the test phase. Results in **Figure 3** support this explanation by showing that, in each of the four belief conditions, looking times were reliably longer (with a mean difference always around 7∼9 s) when there was a conflict between the IFOR representations at the end of the induction phase (the same as the expectation) and in the test phase (the actual outcome). For example, looking times for "*x*-L-R" conditions were consistently longer than those for "*x*-L-L" conditions ("*x*" stands for either "TB" or "FB", and, a conflict is present if the last two alphabets are different).

COMPARISON BETWEEN BELIEF INDUCTION CONDITIONS

It is apparent from **Figure 3** that there were differences in looking times among the four belief induction conditions. For example, whereas the FB-L condition had the longest looking times, the FB-R condition had similar looking times as those in TB conditions. It is surprising that these differences were not mentioned nor accounted for by Onishi and Baillargeon (2005). Using the same argument in the two-cannon task, we speculate that the looking time difference between belief induction conditions might also be due to the interference from a partially disengaged representation. In this case, there could be different levels of the dissociation between EFOR and IFOR representations due to the different sequences of temporal events during the belief induction phase. Based on the comparison between "test" conditions above, it appears that the surprise effect (i.e., "conflict" versus "no conflict") in all belief induction conditions remained approximately constant (7∼9 s). This implies that the variance in looking times, less the surprise effect, would be independent of the predictions by the actor-toy IFOR representation. Accordingly, the remaining variance in looking times could be due solely to the interference from the infant-toy EFOR representation.

In the following, we use the conditional means and standard errors reported in the original study to make three sets of *post hoc* comparisons across different belief induction conditions but within the same "conflict" or "no conflict" test conditions (e.g., *x*-L-L compared with *x*-R-R, *x*-L-R compared with *x*-R-L, and etc.).

First, the mean looking times were about the same in the TB-L and TB-R conditions (i.e., TB-L-L $\approx$ TB-R-R, and, TB-L-R $\approx$ TB-R-L), despite different manipulation sequences in the belief induction phase – the former (TB-L) only involved the movement of an empty container (the "yellow box" on the actor's left hand side) and the latter (TB-R) involved the change of the toy's location (see **Figure 3**). This indicates that the looking times were primarily determined by the active maintenance of the IFOR representation of the actor-toy relationship. If there was any interference from the EFOR representation of the infant-toy relationship, the effect remained constant between these two conditions.

Second, the mean looking times were significantly longer in the FB-L condition than in the TB-R condition (i.e., FB-L-L > TB-R-R, mean difference $\approx$ 8 s; FB-L-R > TB-R-L, mean difference $\approx$ 9 s; two tailed $p < 0.05$ in both comparisons). Such differences could be accounted for by stronger interference from the EFOR representation in the FB-L condition than in the TB-R condition. Specifically, the change of the toy's location was visible only to the infant in the FB-L condition but visible to both the infant and the actor in the TB-R condition. Thus, the infant-toy EFOR representation in the FB-L condition would be relatively stronger (more engaged). Being task-irrelevant (e.g., irrelevant to the actor's fetching action), the stronger EFOR representation in the FB-L condition would lead to greater interference, resulting in longer looking times during the test phase.

Third, the mean looking times were significantly shorter in the FB-R condition than in the FB-L condition (i.e., FB-L-L > FB-R-R, mean difference $\approx$ 7 s; FB-L-R > FB-R-L, mean difference $\approx$ 7 s; one tailed $p < 0.05$ in both comparisons). Interestingly, despite the more complicated manipulation sequences in the FB-R condition, looking times were about the same as those in the true belief conditions (TB-L and TB-R). Consistent with the aforementioned explanation, it is likely that the IFOR representation in the FB-R condition became stronger when it was reinforced in the presence of the actor (the actor last saw the toy moving to her right-hand side). By competition, a stronger IFOR representation led to a weaker EFOR representation. Although both were false-belief conditions, the weaker EFOR representation in the FB-R condition resulted in less interference and, therefore, shorter looking times than the FB-L condition.

In summary, it appears that FOR-based representations may provide a more transparent and detailed explanation to the findings reported by Onishi and Baillargeon (2005). In contrast to the two-cannon experiment by Tamborello et al. (2012), this false-belief task was not explicitly designed to detect the EFOR–IFOR interaction (e.g., infants were always facing the actor with the same bearing). Therefore, the interpretation of our *post hoc* comparisons between belief induction conditions could be limited. Nevertheless, our interpretation remained consistent across all comparisons and across both tasks. That is, in order to track and predict other agent's behavior, the internal process would involve at least a partial disengagement of EFOR representations, an active engagement of IFOR representations, and, potential interference between EFOR and IFOR representations.

Note that our interpretation is in the same vein as the "actor-object-location association" account (Perner and Ruffman, 2005). In addition, we identify the role of EFOR–IFOR dissociation. This interpretation is along the same line as the proposals that belief attribution may evolve from low-level spatial encoding processes, including the identification of "external referent" (Perner et al., 2011) and perspective taking (Kessler and Rutherford, 2010; Kessler and Thomson, 2010). Similar to the original interpretation by Onishi and Baillargeon (2005), here we also emphasize the role of expectation. However, expectation in our account is not

the end product of belief attribution. Rather, it starts early at the level of FOR-based spatial representations. In this respect, belief representation emerges as the mind integrates different spatial representations at different time points by reducing the discrepancy between the actual and the expected outcomes.

FROM SPATIAL TO SOCIAL: THE COMMON NON-COGNITIVE ORIGINS

Although we have demonstrated that the same language from spatial cognition may be used to interpret infants' performance in the false-belief task, we do not claim that social cognitive abilities can be completely accounted for by those in spatial cognition. Moreover, we do not claim a parallel between an explicit spatial orientation task and 15-month-old infants' preferential looking task. Rather, we focus on the common representations underlying these two seemingly different tasks. We argue that abilities from both spatial and social domains share common non-cognitive origins at the level of spatio-temporal association in extracting the environmental statistics. Ergo, these abilities, even if they appear different from each other, may not be domain-specific *per se*, but reflect the different requirements in computational efficiency and flexibility.

In bridging the conceptual gaps between spatial and social cognitive abilities, it is critical to understand the common dynamic nature of spatio-temporal association in both domains. In the present paper, we have shown that, in terms of FOR-based representations, the two-cannon task and the false-belief task share at least three computational properties. First, both tasks require encoding multiple spatial relations with different reference points (spatial association); Second, both involve comparisons of representations at different time points (temporal association); Third, the internal representations for both tasks are not static spatial encodings at isolated time points, rather, they are constructed and maintained through competitions toward the expected outcomes (predictive learning). We argue that all these three properties are governed by the same principle, whether one's goal is to learn a spatial configuration or infer other's intentions and beliefs. That is, the internal representations are developed in the direction of reducing spatio-temporal instability (variances) in order to extract statistical regularities at different levels of abstraction from the task environment.

Commonly shared computational processes could well be supported by commonly shared neural implementations. A growing body of research suggests that brain mechanisms supporting sophisticated social abilities may derive from low-level processes such as spatial tracking, predictive encoding, and attention shifting (for reviews, see, Mitchell, 2006; Corbetta et al., 2008; Frith and Frith, 2012). In the same vein, we argue that the key ingredient in both spatial and social cognition is the expectation-driven competition between multiple FOR-based representations, that are supported by a set of intrinsically distributed neural networks, rather than separately dedicated brain mechanisms. In the following, we discuss the neural evidence that supports this view.

Even a simple task could demand multiple representations of the task environment at different temporal points. Then, the need for selection arises at different levels of processing due to the limitation of resources. On the basis of functional and anatomical distinctions, a model of attention selection has been proposed, suggesting that the attentional operations are carried out by the interactions between two fronto-parietal systems – a dorsal attention system (also referred to as top-down attention network, or, canonical sensory-motor pathway) and a ventral attention system (or, bottom-up attention network; Corbetta and Shulman, 2002; Corbetta et al., 2008; Yeo et al., 2011). The dorsal system is bilateral and mainly composed of the frontal eye field (FEF) and the intraparietal sulcus (IPS). It is specialized for selecting and linking stimuli and responses by sending top-down "filtering" signals to visual areas and via the middle frontal gyrus (MFG) to the ventral network. The ventral system is right-lateralized and includes the right temporal-parietal junction (TPJ), the right ventral frontal cortex (VFC), parts of the MFG, and the inferior frontal gyrus (IFG). Coordinated by the dorsal system, the ventral system sends bottom-up "reorienting" signals that interrupt and reset ongoing activity upon detection of salient targets, especially when there is a violation of expectation (for reviews, see, Corbetta et al., 2008).

The filtering and reorienting functionality in the dorsal–ventral attention networks is particularly useful for implementing the computation of multiple FOR-based representations, particularly when multiple FORs compete. We consider two levels of competition: (1) competition within the dorsal pathway (filtering), and (2), competition carried out by the interaction between the dorsal and ventral pathway (reorienting). Some evidence suggest that, along the dorsal pathway, multiple representations in different FOR can coexist – from lower-level retinotopic representations to higher-level self-centered (EFOR) and world-centered representations (IFOR and AFOR), and that the parietal cortex, particularly the IPS, is central to the construction of these representations (Marr, 1982; Andersen et al., 1997; Colby and Goldberg, 1999; Burgess, 2008; Pertzov et al., 2011; Van Der Werf et al., 2013). Recent rest-state data indicate that the dorsal attention network follows a serial and hierarchical organization, whereas the functional connectivity of parietal and prefrontal association cortices appears to be embedded with largely parallel and interdigitated circuits (Yeo et al., 2011). We argue that such an organization would allow a hierarchical abstraction of the task environment based on flexible selections among multiple representations. That is, in terms of FOR-based representations, it is possible that the invariance extracted at early cortical stages (e.g., visual areas and the parietal cortex) is incomplete, causing different representations to overlap with one another. In order to support higher-level abstractions, a more complete dissociation is required at the level of the prefrontal areas. For instance, it has been suggested that the FEF region plays a crucial role in the construction of intrinsic reference frames among multiple objects in spatial tasks (Wallentin, 2012). Likewise, studies with neural network simulations have shown that, although partial dissociation between different types of spatial information can occur by re-encoding visual information in the parietal cortex, dorsal control from the prefrontal cortex is necessary to achieve a more explicit dissociation (Sun and Wang, 2013); Moreover, efficient and flexible representations of the changing environment requires the maintenance of both latent representations (through altered firing thresholds in non-frontal regions) and active representations (through sustained firing in the prefrontal cortex) (Morton and Munakata, 2002). It is suggested

that such a maintenance mechanism is involved when the infants created actor-object-location associations in the false-belief task (Perner and Ruffman, 2005).

More dramatic competition between multiple representations would likely occur when expectations derived from actual sensory input have been violated. In such instances, the ventral attention network sends out reorienting signals and the dorsal attention network is reconfigured (Corbetta et al., 2008). Evidence for dorsal–ventral interaction comes from studies that use perspective taking tasks, which typically involve conflicting perspectives in EFOR and IFOR representations. For example, it has been reported that the transformation from participants' own perspective to another agent's body axis was associated with activations in posterior parietal cortical regions, such as the left inferior parietal lobe (IPL) and parietal–temporal–occipital junction as well as the right superior parietal lobe (Vogeley et al., 2004; David et al., 2006). Additionally, it has been found that TPJ shows enhanced activities in voluntary orienting of attention when participants are cued about the future location of a target stimulus (Corbetta et al., 2000), and when they need to distinguish between self-produced actions and actions generated by others (Blakemore and Frith, 2003; Jackson and Decety, 2004). Recently, Mazzarella et al. (2013) reported that responses in right IFG are sensitive to another person's orientation when participants perform the task from their own egocentric perspective. Thus, these studies are consistent with the suggestion that taking another person's perspective requires extra effort as compared with using one's own perspective (Kessler and Thomson, 2010).

It should be pointed out that among different brain areas, the TPJ region has been a major topic of debate regarding the neural mechanisms of belief attribution abilities in social interactions. Some researchers argue that this region is specifically involved in the theory-of-mind functions (Saxe and Kanwisher, 2003; Apperly et al., 2004; Saxe and Wexler, 2005; Saxe and Powell, 2006; Saxe et al., 2009; Young et al., 2010). However, the studies mentioned above suggest that the TPJ's function is not unique in the social context. In fact, many theorists consider the TPJ the key hub of the ventral attention network, which essentially supports attention reorienting for resolving conflicts between different visual perspectives, especially when there is a violation of the expected outcomes (Posner et al., 2006; Decety and Lamm, 2007; Mitchell, 2008; Perner and Aichhorn, 2008). Similarly, it has been suggested that the dorsal part of the TPJ region is involved in representing different perspectives and making behavioral predictions, whereas the more ventral part of TPJ and the medial prefrontal cortex region (MPFC) are responsible for predicting behavioral consequences (Aichhorn et al., 2006). Along the same line, Corbetta et al. (2008, p. 317) posited that, "Similar environmental and bodily representations and their comparison may be co-opted for ToM interactions and that attention signals in TPJ may be important to switch between internal, bodily, or self-perspective and external, environmental, or other's viewpoint, a key ingredient of ToM."

In sum, we argue that by supporting different levels of competition between multiple representations, the functions of dorsal–ventral attention networks play a major role in both spatial and social cognitive abilities. Whereas the filtering function manages competition among representations required for the ongoing activity, the reorienting function facilitates competition and reconfiguration when the new sensory input violates the expectation from the current representations. Crucially, different levels of competition allow partial engagement (or disengagement) of certain representations, which facilitate the integration of potentially conflicting representations. As mentioned earlier, maintaining multiple IFOR representations is essential for prioritizing while being prepared for the unexpected. Combining EFOR and IFOR representations (perspective taking) takes advantage of both the efficient removal of task-irrelevant variance and fast mental simulation. When infants start to learn by copying others' actions (Meltzoff, 1995; Tomasello et al., 2005; Nielsen, 2006), it is important for them to hold both EFOR and IFOR representations so that imitation and emulation are possible.

SUMMARY

The central theme in our proposal is that the complex achievements in either spatial cognition or social cognition may rely on the fundamental processes of spatio-temporal integration and, moreover, that there is a set of distributed brain regions shared by both types of cognition. In our framework, both spatial and social abilities arise in the form of spatio-temporal association in which the mind constantly deals with the temporal instability in the environment by predictive learning. In the effort of extracting statistical regularities, the internal representations evolve by first partitioning the environmental variances – namely, developing FOR-based representations – then, encoding statistical invariance at different levels of abstractions. Since the statistical regularities include not only the spatial relations of static configurations but also the temporal relations between sequential events, predictive learning links various representations with different anchors (spatial integration) at different time points (temporal integration). Together, abstract knowledge of the environment (including those about other's beliefs and intentions) emerges from the expectation-driven competitions among multiple FOR-based representations.

In our view, different abilities are not domain-specific *per se*, rather, they are subject to the competing demands of computational efficiency and flexibility, yet are bounded by the statistical structures in the environment. By reinterpreting the results from the two-cannon experiment (Tamborello et al., 2012) and the false-belief task (Onishi and Baillargeon, 2005) and reviewing recent neurocognitive findings, we advocate an integrated approach that connects low-level perceptual processes, such as spatial representations, with high-level functions such as belief reasoning. The advantage of this approach is that, rather than singling out a certain brain system for a certain set of cognitive abilities (e.g., the TPJ for belief reasoning), we can pursue a better understanding of the mind–environment interaction over a developmental continuum. For example, the FOR-based account proposed here largely relies on the mechanisms of attentional network in spatial cognition, which have been extensively studied on from non-human animals to human infants and adults (for reviews, see, Corbetta and Shulman, 2002; Posner et al., 2006; Corbetta et al., 2008; Kavšek, 2013). Thus, this account may provide not only a transparent partitioning of the environmental statistics, but also potential explanations for the relationship between different abilities and

the development of specific attentional networks. For instance, it has been suggested that "rudimentary executive attention capacities may emerge during the first year of life but that more advanced conflict resolution capacities are not present until 2 years of age" (Posner et al., 2006, p. 1425). This line of reasoning could explain why young infants suddenly appear to comprehend the complex world and pass various spatial tasks (McCrink and Wynn, 2007; Surian et al., 2007; Kovács et al., 2010; Gweon and Schulz, 2011; Téglás et al., 2011).

Legend has it that in formulating his theory of gravitation, Newton was inspired by observing the acceleration of an apple falling from a tree. Subsequently, he inferred the existence of gravity and extended the effect from to the top of the tree to the Moon (White, 1991). Perhaps more interestingly, Newton also first stated the principle of relativity (later modified by Einstein), which essentially claims that observations of the physical world depend on the particular "frame of reference" (Feynman et al., 1963, p. 162). Although we may never know the exact details of his revelation, the "apple incident" exemplifies how early perceptual analyses are triggered by temporal instability in the environment and the resulting extraction of statistical regularities with various reference points. In addition, it illuminates recent proposals that complex achievements such as mathematics and geometry, which are uniquely human in their full linguistic and symbolic realization, rest nevertheless on a set of core knowledge systems that are driven by the representations of object, space, time and number (Spelke and Kinzler, 2007; Spelke et al., 2010), and, knowledge structures emerge from non-cognitive processes by dynamic associations (McClelland et al., 2010). While controversies still exist between seemingly diverging perspectives, we take the primary theme of the debates to be the converging efforts of seeking for the cognitive or non-cognitive origins of human thinking and reasoning abilities. If we subscribe to the notion of "bounded rationality" (Simon, 1982), both spatial and social abilities are bounded by the learning agent's computation capacity and the structure of the environment. In order to bridge the conceptual gaps between spatial and social cognition, the key is to understand the interactions between "genetic endowment and the environment" (Ruffman and Perner, 2005, p. 462).

ACKNOWLEDGMENTS

This work was supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of the Interior (DOI) Contract no. D10PC20021 and the Office of Naval Research (ONR) Grant no. N00014-08-1-0042. 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 would like to thank Dr. Paul J. Schroeder for helpful comments.

REFERENCES

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

Andersen, R. A., Snyder, L. H., Bradley, D. C., and Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annu. Rev. Neurosci.* 20, 303–330. doi: 10.1146/annurev.neuro.20.1.303

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

Apperly, I. A., Samson, D., Chiavarino, C., and Humphreys, G. W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. *J. Cogn. Neurosci.* 16, 1773–1784. doi: 10.1162/0898929042947928

Baillargeon, R. (1986). Representing the existence and the location of hidden objects: object permanence in 6- and 8-month-old infants. *Cognition* 23, 21–41. doi: 10.1016/0010-0277(86)90052-1

Baillargeon, R., Scott, R. M., and He, Z. (2010). False-belief understanding in infants. *Trends Cogn. Sci.* 14, 110–118. doi: 10.1016/j.tics.2009.12.006

Blakemore, S.-J., and Frith, C. (2003). Self-awareness and action. *Curr. Opin. Neurobiol.* 13, 219–224. doi: 10.1016/S0959-4388(03)00043-6

Burgess, N. (2008). Spatial cognition and the brain. *Ann. N. Y. Acad. Sci.* 1124, 77–97. doi: 10.1196/annals.1440.002

Carlson, L. A., and Van Deman, S. R. (2008). Inhibition within a reference frame during the interpretation of spatial language. *Cognition* 106, 384–407. doi: 10.1016/j.cognition.2007.03.009

Carlson-Radvansky, L. A., and Irwin, D. E. (1993). Frames of reference in vision and language: where is above? *Cognition* 46, 223–244. doi: 10.1016/0010-0277(93)90011-J

Chen, X., and McNamara, T. (2011). Object-centered reference systems and human spatial memory. *Psychon. Bull. Rev.* 18, 985–991. doi: 10.3758/s13423-011-0134-5

Colby, C. L., and Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annu. Rev. Neurosci.* 22, 319–349. doi: 10.1146/annurev.neuro.22.1.319

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., and Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.* 3, 292–297. doi: 10.1038/73009

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

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Csibra, G., and Southgate, V. (2006). Evidence for infants' understanding of false beliefs should not be dismissed. *Trends Cogn. Sci.* 10, 4–5. doi: 10.1016/j.tics.2005.11.011

David, N., Bewernick, B. H., Cohen, M. X., Newen, A., Lux, S., Fink, G. R., et al. (2006). Neural representations of self versus other: visual-spatial perspective taking and agency in a virtual ball-tossing game. *J. Cogn. Neurosci.* 18, 898–910. doi: 10.1162/jocn.2006.18.6.898

Decety, J., and Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neuroscientist* 13, 580–593. doi: 10.1177/1073858407304654

Feynman, R. P., Leighton, R. B., and Sands, M. (1963). *The Feynman Lectures on Physics*, Vol. 1. Reading, MA: Addison-Wesley.

Frith, C. D., and Frith, U. (2007). Social cognition in humans. *Curr. Biol.* 17, R724–R732. doi: 10.1016/j.cub.2007.05.068

Frith, C. D., and Frith, U. (2012). Mechanisms of social cognition. *Annu. Rev. Psychol.* 63, 287–313. doi: 10.1146/annurev-psych-120710-100449

Gerstner, W., Sprekeler, H., and Deco, G. (2012). Theory and simulation in neuroscience. *Science* 338, 60–65. doi: 10.1126/science.1227356

Gweon, H., and Schulz, L. (2011). 16-month-olds rationally infer causes of failed actions. *Science* 332, 1524. doi: 10.1126/science.1204493

Hawkins, J., and Blakeslee, S. (2004). *On Intelligence*. New York: Henry Holt.

Jackson, P. L., and Decety, J. (2004). Motor cognition: a new paradigm to study self-other interactions. *Curr. Opin. Neurobiol.* 14, 259–263. doi: 10.1016/j.conb.2004.01.020

Kavšek, M. (2013). The comparator model of infant visual habituation and dishabituation: recent insights. *Dev. Psychobiol.* 55, 793–808. doi: 10.1002/dev.21081

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

Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88. doi: 10.1016/j.cognition.2009.08.015

Kovács, Á. M., Téglás, E., and Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science* 330, 1830–1834. doi: 10.1126/science.1190792

Krauzlis, R. J., and Nummela, S. U. (2011). Attention points to the future. *Nat. Neurosci.* 14, 130–131. doi: 10.1038/nn0211-130

Leslie, A. M. (2005). Developmental parallels in understanding minds and bodies. *Trends Cogn. Sci.* 9, 459–462. doi: 10.1016/j.tics.2005.08.002

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. Garrett (Cambridge, MA: MIT Press), 77–108.

Levinson, S. C. (2004). *Space in Language and Cognition: Explorations in Cognitive Diversity.* New York: Cambridge University Press.

Marr, D. (1982). *Vision.* San Francisco, CA: W. H. Freeman.

Mazzarella, E., Ramsey, R., Conson, M., and Hamilton, A. (2013). Brain systems for visual perspective taking and action perception. *Soc. Neurosci.* 8, 248–267. doi: 10.1080/17470919.2012.761160

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., et al. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends Cogn. Sci.* 14, 348–356. doi: 10.1016/j.tics.2010.06.002

McCrink, K., and Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychol. Sci.* 18, 740–745. doi: 10.1111/j.1467-9280.2007.01969.x

Meltzoff, A. N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Dev. Psychol.* 31, 838–850. doi: 10.1037/0012-1649.31.5.838

Miller, G. A., and Johnson-Laird, P. N. (1976). *Language and Perception.* Cambridge, MA: Harvard University Press.

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

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cereb. Cortex* 18, 262–271. doi: 10.1093/cercor/bhm051

Morton, J. B., and Munakata, Y. (2002). Active versus latent representations: a neural network model of perseveration, dissociation, and decalage. *Dev. Psychobiol.* 40, 255–265. doi: 10.1002/dev.10033

Mou, W., Fan, Y., McNamara, T. P., and Owen, C. B. (2008). Intrinsic frames of reference and egocentric viewpoints in scene recognition. *Cognition* 106, 750–769. doi: 10.1016/j.cognition.2007.04.009

Nielsen, M. (2006). Copying actions and copying outcomes: social learning through the second year. *Dev. Psychol.* 42, 555–565. doi: 10.1037/0012-1649.42.3.555

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

O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., and Contributors. (2012). *Computational Cognitive Neuroscience*, Wiki Book, 1st Edn, Available at URL: http://ccnbook.colorado.edu (accessed January 31, 2014).

Perner, J., and Aichhorn, M. (2008). Theory of mind, language and the temporoparietal junction mystery. *Trends Cogn. Sci.* 12, 123–126. doi: 10.1016/j.tics.2008.02.001

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

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

Pertzov, Y., Avidan, G., and Zohary, E. (2011). Multiple reference frames for saccadic planning in the human parietal cortex. *J. Neurosci.* 31, 1059–1068. doi: 10.1523/JNEUROSCI.3721-10.2011

Posner, M. I., Sheese, B. E., Odludaş, Y., and Tang, Y.-Y. (2006). Analyzing and shaping human attentional networks. *Neural Netw.* 19, 1422–1429. doi: 10.1016/j.neunet.2006.08.004

Rolfs, M., Jonikaitis, D., Deubel, H., and Cavanagh, P. (2011). Predictive remapping of attention across eye movements. *Nat. Neurosci.* 14, 252–256. doi: 10.1038/nn.2711

Ruffman, T., and Perner, J. (2005). Do infants really understand false belief? Response to Leslie. *Trends Cogn. Sci.* 9, 462–463. doi: 10.1016/j.tics.2005.08.001

Saxe, R. R. (2006). Why and how to study Theory of Mind with fMRI. *Brain Res.* 1079, 57–65. doi: 10.1016/j.brainres.2006.01.001

Saxe, R. R., and Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind." *Neuroimage* 19, 1835–1842. doi: 10.1016/S1053-8119(03)00230-1

Saxe, R. R., and Powell, L. J. (2006). It's the thought that counts: specific brain regions for one component of theory of mind. *Psychol. Sci.* 17, 692–699. doi: 10.1111/j.1467-9280.2006.01768.x

Saxe, R. R., and Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia* 43, 1391–1399. doi: 10.1016/j.neuropsychologia.2005.02.013

Saxe, R. R., Whitfield-Gabrieli, S., Scholz, J., and Pelphrey, K. A. (2009). Brain regions for perceiving and reasoning about other people in school-aged children. *Child Dev.* 80, 1197–1209. doi: 10.1111/j.1467-8624.2009.01325.x

Simon, H. A. (1982). *Models of Bounded Rationality*. Cambridge, MA: MIT Press.

Spelke, E. S., and Kinzler, K. D. (2007). Core knowledge. *Dev. Sci.* 10, 89–96. doi: 10.1111/j.1467-7687.2007.00569.x

Spelke, E. S., Lee, S. A., and Izard, V. (2010). Beyond core knowledge: natural geometry. *Cogn. Sci.* 34, 863–884. doi: 10.1111/j.1551-6709.2010.01110.x

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

Sun, Y., and Wang, H. (2013). The parietal cortex in sensemaking: the dissociation of multiple types of spatial information. *Comput. Intell. Neurosci.* 2013:152073. doi: 10.1155/2013/152073

Surian, L., Caldi, S., and Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychol. Sci.* 18, 580–586. doi: 10.1111/j.1467-9280.2007.01943.x

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

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., and Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science* 332, 1054–1059. doi: 10.1126/science.1196404

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. doi: 10.1017/S0140525X05000129

Van Der Werf, J., Buchholz, V. N., Jensen, O., and Medendorp, W. P. (2013). Reorganization of oscillatory activity in human parietal cortex during spatial updating. *Cereb. Cortex* 23, 508–519. doi: 10.1093/cercor/bhr387

Vogeley, K., May, M., Ritzl, A., Falkai, P., Zilles, K., and Fink, G. R. (2004). Neural correlates of first-person perspective as one constituent of human self-consciousness. *J. Cogn. Neurosci.* 16, 817–827. doi: 10.1162/089892904970799

Wallentin, M. (2012). The role of the brain's frontal eye fields in constructing frame of reference. *Cogn. Process.* 13, 359–363. doi: 10.1007/s10339-012-0461-0

Wang, H., Johnson, T. R., Sun, Y., and Zhang, J. (2005a). Object location memory: the interplay of multiple representations. *Mem. Cogn.* 33, 1147–1159. doi: 10.3758/BF03193219

Wang, H., Sun, Y., Johnson, T. R., and Yuan, Y. (2005b). Prioritized spatial updating in the intrinsic frame of reference. *Spat. Cogn. Comput.* 5, 89–113. doi: 10.1207/s15427633scc0501_4

Wang, R. F., and Spelke, E. S. (2002). Human spatial representation: insights from animals. *Trends Cogn. Sci.* 6, 376–382. doi: 10.1016/S1364-6613(02)01961-7

White, M. (1991). *Isaac Newton: The Story of a Great Mathematician Who Changed Our Perception of the Universe.* Watford, UK: Exley.

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* 106, 1125–1165. doi: 10.1152/jn.00338.2011

Young, L., Dodell-Feder, D., and Saxe, R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia* 48, 2658–2664. doi: 10.1016/j.neuropsychologia.2010.05.012

Zwickel, J., White, S. J., Coniston, D., Senju, A., and Frith, U. (2011). Exploring the building blocks of social cognition: spontaneous agency perception and visual perspective taking in autism. *Soc. Cogn. Affect. Neurosci.* 6, 564–571. doi: 10.1093/scan/nsq088

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: 28 June 2013; accepted: 24 January 2014; published online: 14 February 2014.
Citation: Sun Y and Wang H (2014) Insight into others' minds: spatio-temporal representations by intrinsic frame of reference. Front. Hum. Neurosci. ***8****:60. doi: 10.3389/fnhum.2014.00058*
This article was submitted to the journal Frontiers in Human Neuroscience.