Comparing the Mechanisms of First-Order and Second

Authors: Wang Jiayin, Li J

Date: 2026-01-21T12:18:25+00:00

Abstract

Visual perspective taking (VPT) is divided into Level-1 VPT and Level-2 VPT. Existing theories are fundamentally divided regarding the relationship between the two: dual-system theories propose that their internal mechanisms are independent, whereas single-system theories argue that they share the same system. Integrating these two theoretical perspectives, this paper proposes a three-stage processing model, which posits that both Level-1 and Level-2 VPT undergo three stages: information processing, perspective simulation, and information integration and response selection. Behavioral and neural evidence indicates that, across these three stages, the processing mechanisms of Level-1 and Level-2 VPT may exhibit both differences and similarities: in the information processing stage, they share a basic encoding of spatial information, but the representations involved in Level-2 VPT are more fine-grained; in the perspective simulation stage, Level-1 VPT relies on rapid, non-embodied gaze following, whereas Level-2 VPT requires embodied reference frame transformation and alignment via mental self-rotation; in the information integration and response selection stage, both may share the understanding of others' intentions, but Level-2 VPT demands stronger cognitive control. The three-stage model constructed in this paper provides a unified framework for understanding Level-1 and Level-2 VPT. Future research should focus on developing experimental paradigms that can dissociate each processing stage, further testing the temporal dynamics of this model, and exploring in depth the triggering conditions of embodied mechanisms in Level-2 VPT and their cross-modal integration.

Full Text

Comparing the Mechanisms of Level-1 and Level-2 Visual Perspective Taking: Theoretical Controversies, Behavioral and Neuroscientific Evidence

WANG Jiayin, LI Jing
(School of Psychology, Nanjing Normal University, Nanjing 210097, China)

Abstract

Visual perspective taking (VPT) is commonly divided into level-1 and level-2 visual perspective taking. Current theories fundamentally disagree on their relationship: the two-systems account posits distinct mechanisms, while the single-system account suggests shared processing. Integrating these perspectives, this article proposes a three-stage processing model, which posits that both level-1 and level-2 VPT involve three stages: information processing, perspective simulation, and information integration with response selection. Behavioral and neural evidence indicates that the mechanisms of level-1 and level-2 VPT exhibit both similarities and differences across these stages. During information processing, both share basic spatial information encoding, but level-2 VPT requires more fine-grained representations. In the perspective simulation stage, level-1 VPT relies on rapid, non-embodied gaze tracking, whereas level-2 VPT involves embodied self-rotation with reference frame transformation. During information integration and response selection, both may share an understanding of others' intentions, though level-2 VPT demands stronger cognitive control. The proposed three-stage model offers a unified framework for understanding level-1 and level-2 VPT. Future research should focus on developing experimental paradigms to dissociate these stages, employing high-temporal-resolution techniques to examine the temporal dynamics of the model, and further investigating the triggering conditions of embodied mechanisms in level-2 VPT, as well as cross-modal integration.

Keywords: visual perspective taking, two-systems account, single-system account, spatial cognition

The English poet John Donne once wrote, "No man is an island···each is a piece of the continent." As social beings, we constantly need to interact with others, predict their behavior, understand their intentions, and respond appropriately. Visual perspective taking (VPT) refers to the ability to mentally simulate and understand others' visual experiences (Flavell et al., 1981). During social interaction, VPT enables us to infer others' attentional focus and what they see from their viewpoint, thereby facilitating coordinated action (邵雨婷等, 2020). For instance, when asking a friend sitting across from us to pour a glass of water, we must recognize that they can see the cup on the table and understand that the kettle on our right is on their left. VPT not only makes us aware of differences between others' visual information and our own, but also provides a foundation

for inferring more complex mental states such as beliefs and intentions (肖承丽等, 2021; Pesimena & Soranzo, 2023; Samuel, Cole et al., 2023).

Flavell et al. (1981) proposed that VPT can be divided into two levels: level-1 VPT and level-2 VPT. Level-1 VPT focuses on simple binary judgments about whether an object falls within another' s visual field—for example, realizing that a person facing away from a bookshelf cannot see the books on it. Level-2 VPT, by contrast, involves more complex inferential processes. It requires not only judging whether an object is within another's visual field, but also understanding that even when both self and other can see the same object, differences in viewing distance, angle, and other environmental factors will result in different visual appearances (Flavell et al., 1981). For instance, when two people facing each other simultaneously see a number on a table, we can understand that the "6" in our view appears as "9" in the other' s view (Surtees et al., 2012). This distinction between level-1 and level-2 VPT reflects the complexity of simulating others' visual experiences.

Although Flavell et al. (1981) conceptually distinguished level-1 from level-2 VPT, the relationship between their internal mechanisms remains unclear—whether they involve fundamentally different processing mechanisms or share similar internal processes. Cole et al. (2020) noted that the VPT field currently lacks a unified theoretical framework, and that clarifying the unique and shared processing mechanisms of level-1 and level-2 VPT is a prerequisite for filling this gap. Only by deeply understanding the internal processes of each type can we construct a systematic theory covering different VPT types. Several leading VPT researchers have jointly identified examining whether different forms of VPT share processing mechanisms as a key issue for future perspective-taking research (Samuel et al., 2024). Therefore, systematically elucidating the similarities and differences between level-1 and level-2 VPT processing mechanisms holds significant research value. Existing domestic reviews of perspective taking have primarily focused on developmental research (赵婧等, 2010), behavioral paradigms (张越等, 2018), mechanisms of egocentric bias (吴梦慧等, 2022), debates on implicit mentalizing versus submentalizing in automatic perspective taking (李艺, 肖风, 2021), and the relationship between perspective taking and spatial language communication (肖承丽等, 2021). No study has specifically examined the relationship between level-1 and level-2 VPT.

Based on this, the present article aims to systematically clarify the similarities and differences in processing mechanisms between level-1 and level-2 VPT. We first examine the core controversy between the two-systems and single-system accounts, identify their respective limitations, and propose an integrative "three-stage processing model" as a novel analytical framework. Following this framework, we systematically review behavioral and cognitive neuroscience evidence, elaborating on the unique and shared mechanisms of level-1 and level-2 VPT across the three stages of information processing, perspective simulation, and information integration with response selection. Finally, based on this evidence integration, we formally construct and elaborate the three-stage model of level-1

and level-2 VPT, and propose future research directions to provide new perspectives for the field.

2 The Relationship Between Level-1 and Level-2 Visual Perspective Taking from Multiple Theoretical Perspectives

From a theoretical standpoint, no unified framework specifically for VPT currently exists. However, some "mindreading" theories have offered different interpretations of the internal processes of level-1 and level-2 VPT in previous research. Mindreading theories explore how humans understand others' mental states, and some have discussed the relationship between level-1 and level-2 VPT as supporting evidence. The most representative theories are the two-systems account and the single-system account.

2.1 Basic Tenets of the Two Theories

The two-systems account posits two independent yet complementary systems responsible for understanding others' mental states: a minimal mindreading system and a full-blown mindreading system (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013; Low et al., 2016). The minimal mindreading system develops early, with its core function being to represent "registration"—recording spatial relationships between individuals, others, and objects (Apperly & Butterfill, 2009). It operates rapidly, automatically, and is not easily disrupted by background knowledge, but struggles with tasks requiring high flexibility such as those involving beliefs (Apperly et al., 2006; Butterfill & Apperly, 2013). When task demands become more complex, individuals must employ the full-blown mindreading system (Surtees et al., 2012; Thompson, 2014). The full-blown mindreading system, also called the flexible system, matures later in development and is supported by language and executive function (Apperly & Butterfill, 2009; Samson & Apperly, 2010). It can flexibly handle more complex mental state reasoning tasks, but operates slowly and consumes substantial cognitive resources (Low et al., 2016).

In contrast, the single-system account argues that from infancy to adulthood, only a single, continuously developing mental reasoning system exists (Carruthers, 2016; Gómez-Tabares, 2023; Kloo et al., 2020). This system has a rudimentary form in infancy and its functions gradually enrich during development, evolving into the mature adult system. This theory denies the existence of two independent parallel systems, attributing performance differences across tasks not to different systems but to varying cognitive resource demands.

In summary, the two-systems and single-system accounts propose fundamentally different assumptions about the basic architecture of mindreading: the former advocates two independent, functionally complementary cognitive systems responsible for different categories of mental reasoning, while the latter emphasizes that a unified system can explain performance differences across mental reasoning tasks (李鸿锴, 2025). This theoretical disagreement also gen-

erates conflicting views on the internal mechanisms of level-1 and level-2 VPT. Next, we further clarify the specific viewpoints and conflicts of these two theories in explaining VPT mechanisms.

2.2 Theoretical Conflicts Regarding VPT Mechanisms

The core conflict between the two-systems and single-system accounts concerning VPT internal mechanisms centers on whether level-1 and level-2 VPT represent two completely independent mechanisms or share the same internal mechanism.

The two-systems account holds that level-1 and level-2 VPT depend on two distinct processing mechanisms: level-1 VPT relies on the minimal mindreading system, requiring only simple tracking of location visibility information, whereas level-2 VPT involves the full-blown mindreading system and thus can understand others' specific representations of objects (Low et al., 2016). The two VPT mechanisms are parallel and independent, unable to directly exchange information. This implies that under the two-systems framework, level-1 VPT tends toward efficient automatic processing, while level-2 VPT depends more on flexible but cognitively demanding processes. Key evidence for the two-systems account comes from classic paradigms for level-1 and level-2 VPT tasks. In the classic level-1 VPT paradigm—the dot-perspective task (Samson et al., 2010)—participants judge the number of dots they or an avatar sees. Results show that when perspectives conflict, even when explicitly told to ignore the avatar' s perspective, participants'responses slow down and errors increase. The two-systems account interprets this as strong evidence that level-1 VPT can process others' visual information rapidly and nearly automatically without explicit prompting. In contrast, in the number recognition task measuring level-2 VPT (Surtees et al., 2012), participants judging their own perspective were not slowed by the discrepancy with the avatar' s perspective, suggesting that level-2 VPT content cannot be processed automatically and requires effortful mobilization of cognitive resources.

However, single-system theorists have challenged these claims (Bohl & van den Bos, 2012; Carruthers, 2016; Cole et al., 2020; Cole & Millett, 2019; Jacob, 2019; Tomasello, 2018). Research shows that the automatic processing in level-1 VPT is not inevitable and can be modulated by beliefs about others' perceptual abilities (Cole & Millett, 2019; Furlanetto et al., 2016). For example, in Furlanetto et al.' s (2016) adapted dot-perspective task, when participants believed the avatar wore opaque goggles, their response times did not increase even when the avatar' s perspective conflicted with their own. This indicates that level-1 VPT judgments are not fully automatic; they only occur when participants know the other can see the object. Similarly, Elekes et al. (2016) adapted the level-2 VPT number recognition task with three conditions: individual, joint perspective-dependent, and joint perspective-independent. In the individual condition, participants performed the number task alone. In the joint perspective-dependent condition, a real partner sat before them, suppos-

edly performing the same task. In the joint perspective-independent condition, participants were told the partner was performing a different task. Results showed that when participants knew and believed their partner was performing the same task, their responses were significantly faster than in the individual condition.

Westra (2017) noted that level-2 VPT tasks are not inherently "slow system" tasks; with correct background knowledge and sufficient motivation, level-2 VPT can also be rapid and efficient. Therefore, the single-system account argues that differences between level-1 and level-2 VPT tasks stem from varying task resource demands, not operation of two independent systems. Level-1 VPT appears rapid and automatic because it only involves simple spatial relational reasoning, whereas level-2 VPT appears slow and effortful because it requires more complex operations (e.g., mental rotation) and thus mobilizes more cognitive resources. This difference is quantitative rather than qualitative, with both sharing the same underlying processing system. However, the single-system explanation of a shared mechanism faces challenges. If level-1 and level-2 VPT truly depend on a single system, their processing mechanisms should be highly consistent, yet existing research reveals differences: level-1 VPT relies on rapid gaze tracking (Michelon & Zacks, 2006), while level-2 VPT more heavily involves bodily representation and simulation (Kessler & Thomson, 2010). These differences are difficult to explain solely by differences in cognitive resource allocation.

2.3 A New Theoretical Direction: The Multi-Stage Processing Model

Although the two-systems and single-system accounts provide insightful frameworks for understanding the relationship between level-1 and level-2 VPT, they essentially treat VPT as a holistic process dependent on specific system modules—whether two independent systems or one unified system. However, this system-level division struggles to fully explain all empirical findings: the two-systems account cannot explain why level-1 VPT' s "automatic" processing is modulated by beliefs, nor why level-2 VPT can sometimes be rapid and efficient; while the single-system account cannot clarify why level-1 and level-2 VPT stably depend on different core cognitive mechanisms. These limitations suggest that a theoretical perspective at the holistic "system" level may fail to capture the complexity of internal processing in level-1 and level-2 VPT, necessitating a more inclusive and integrative new theoretical framework. VPT tasks are not single cognitive acts but composite tasks comprising multiple subprocesses. Whether completing level-1 or level-2 VPT, individuals must process perceptual information, simulate others' perspective content, distinguish self from others' perspectives, and inhibit interference from irrelevant perspectives. Therefore, VPT cannot be simply summarized as "one system" or "two systems." A more reasonable speculation is that the relationship between level-1 and level-2 VPT is dynamic across different processing subprocesses: at some stages, their cognitive mechanisms may highly overlap, while at others they may call upon completely

different processing mechanisms.

Based on this, we propose shifting the analytical focus from "system" to "stage," decomposing VPT into three main processing stages according to task demands: (1) an information processing stage, involving preliminary perception and representation of spatial relations and properties of self, others, and objects; (2) a perspective simulation stage, involving specific mental operations to simulate information content from others' viewpoints; and (3) an information integration and response selection stage, involving integration of multiple information sources, inhibition of irrelevant perspective interference, and final decision-making. This "three-stage processing model" does not presuppose that level-1 and level-2 VPT are entirely independent or entirely shared, but instead seeks to examine at a more fine-grained level which stages involve overlapping mechanisms and which involve fundamental differences. This provides a more integrative theoretical framework for understanding the complex relationship between level-1 and level-2 VPT, which exhibits both commonalities and differences. Next, we systematically examine behavioral and neuroscientific evidence to lay a solid foundation for constructing such an integrative multi-stage VPT processing model.

3 Behavioral Studies on Level-1 and Level-2 Visual Perspective Taking

Theoretical controversies in mindreading suggest that level-1 and level-2 VPT may involve both distinct and similar processes. As previously discussed, we propose a three-stage processing framework comprising information processing, perspective simulation, and information integration with response selection. Following this framework, we systematically elaborate on the behavioral evidence for mechanistic similarities and differences between level-1 and level-2 VPT across these stages.

3.1 Information Processing Stage: Differences in Representational Depth and Scope

At the initial processing stage of level-1 and level-2 VPT tasks, individuals must encode basic elements of the visual scene, constructing a spatial situation model encompassing self, others, and objects. Behavioral research indicates that the core difference between level-1 and level-2 VPT at this stage lies in the depth and scope of information representation.

Numerous studies show that level-1 VPT judgments depend on rapid assessment of whether sightlines are unobstructed (Kelly et al., 2004; Michelon & Zacks, 2006). Michelon and Zacks (2006) conducted a classic study demonstrating mechanistic differences between level-1 and level-2 VPT. In their experiment, participants completed two tasks: judging whether objects were to the left or right of an avatar (level-2 VPT task) and judging whether the avatar could see target objects (level-1 VPT task). Results showed that level-1 VPT performance was only affected by the distance between avatar and object, with

longer distances yielding slower responses, but was unaffected by angular differences between participant and avatar. This indicates that sightlines are the key factor in level-1 VPT. Individuals performing level-1 VPT likely engage in a gaze-tracking mechanism, imagining a "virtual line" from the avatar' s eyes to the target object; if this line is not occluded, the object is visible (Michelon & Zacks, 2006). Therefore, during the information processing stage of level-1 VPT, participants only need to encode the physical property of "whether sightlines are blocked," resulting in shallow representational depth limited to the sightline path itself. Subsequent studies manipulating sightline validity have corroborated this view. Goggles tasks by Furlanetto et al. (2016) and Marshall et al. (2018) found that when participants believed the avatar wore opaque glasses, the avatar' s perspective no longer automatically interfered with participants' self-judgments. This shows that level-1 VPT triggering is not an unconditional reflex but depends on beliefs about others' visual perceptual abilities, with sightline unobstruction being the physical prerequisite for such beliefs. Barrier tasks by Baker et al. (2016) and O' Grady et al. (2020) yielded similar results. These barrier tasks, built upon the traditional dot-perspective task, placed barriers blocking sightlines between avatar and objects. In visible conditions, barriers had "windows" allowing wall visibility; in invisible conditions, sightlines were completely blocked. When barriers occluded the avatar' s view, the interference effect of others' perspectives on self-task judgments disappeared. Collectively, these findings suggest that during the information processing stage, level-1 VPT primarily identifies environment information related to sightline paths, such as encoding basic spatial relations like physical occlusion, distance, and head orientation, with relatively limited scope.

In contrast, level-2 VPT requires deep and refined spatial representation. Unlike level-1 VPT, level-2 VPT demands understanding objects' specific forms (e.g., "6" vs. "9") or precise locations (left or right) from others' perspectives. This necessitates deeper processing of spatial information. Most research indicates that sightlines have minimal impact on level-2 VPT (Quesque et al., 2018; Ward et al., 2019; Ward et al., 2020). Ward et al. (2020) adapted a classic mental rotation task, presenting rotated characters on a table with an irrelevant person at the table' s edge. Results showed that even when character rotation differed substantially from participants' own perspective, responses were faster when character orientation aligned with the table-edge person' s upright orientation. Importantly, this process remained unaffected when the individual turned their head away, indicating that level-2 VPT processing involves not only confirming sightline unobstruction but also more complex spatial transformations, requiring precise representation of how objects might appear differently due to perspective, as well as information about self and avatar body orientation. In summary, during the information processing stage, level-2 VPT processes more fine-grained and complex information, with its representational scope extending from simple sightline paths to specific spatial relations and visual forms.

3.2 Perspective Simulation Stage: Rapid Gaze Tracking vs. Embodied Reference Frame Transformation

After completing basic spatial information representation, individuals must execute core mental operations to simulate what others see. At this stage, level-1 and level-2 VPT exhibit distinct processing mechanisms.

Level-1 VPT simulation is relatively direct. As previously discussed, research shows level-1 VPT is affected by distance between avatar and target but not by angular differences between observer and target. Therefore, the perspective simulation stage in level-1 VPT primarily involves rapid "gaze tracking" (Michelon & Zacks, 2006). Creem-Regehr et al. (2013) noted that this simulation mechanism is non-egocentric transformation—individuals need not leave their inherent spatial reference frame but remain anchored in their own perspective, merely performing a virtual "line-drawing" operation within this framework to mentally judge whether sightlines from others' eyes to target objects are interrupted.

Level-2 VPT perspective simulation, however, is more complex. The same study by Michelon and Zacks (2006) showed that in level-2 VPT tasks, participants' response times were significantly affected by angular differences between self and avatar, with larger angles producing slower responses. This indicates that level-2 VPT involves mental rotation. It is important to note that level-2 VPT differs fundamentally from object mental rotation and should not be considered identical psychological mechanisms (赵杨柯等, 2010). Researchers argue that level-2 VPT is an "embodied" process (Müsseler et al., 2022; Samuel, Salo et al., 2023). "Embodied" broadly means body-related, emphasizing that the body plays an important role in cognition and perception (Wilson, 2002). Substantial evidence demonstrates that level-2 VPT is closely related to body and action representation (Fischer & Demiris, 2020; Kessler & Thomson, 2010; Surtees et al., 2013; Yu & Zacks, 2017). Studies by Kessler and Thomson (2010) and Surtees et al. (2013) found that body posture alignment is crucial for level-2 VPT. In these studies, participants sat on rotating chairs to judge object locations and number appearances from the avatar's perspective. Results showed that for both types of judgments, performance was better when participants' own movement direction aligned with the avatar' s orientation (e.g., when participants rotated rightward and the avatar also faced rightward). This demonstrates that bodily cues are essential for level-2 VPT. During perspective simulation in level-2 VPT, individuals mentally simulate the bodily movements (rotation/translation) necessary to obtain the other' s viewpoint. This mental simulation is not object-based mental rotation but rather an embodied self-rotation that matches one' s own movement patterns to the avatar' s posture, "putting oneself in another' s shoes" (Kessler & Thomson, 2010).

Furthermore, it is important to note that this embodied imagination of rotating oneself to another's perspective is not a "bit-by-bit"rotation but rather a transformation that reorients one' s reference frame to align with a new "principal axis." Wraga et al. (2005) and Negen (2025) used identical materials to compare men-

tal rotation and self-rotation (similar to the perspective-taking rotation process in level-2 VPT). Their test materials consisted of ten three-dimensional objects made of cubes. The mental rotation task required participants to imagine the object rotating to a certain angle; the self-rotation task required participants to imagine themselves moving around the object to a fixed angle before making a judgment. Results revealed that self-rotation performance did not decline monotonically with increasing rotation angle but showed a distinctive "trough-shaped" pattern. At intermediate angles (60°), participants' performance was significantly better than at smaller (30°) and larger angles (90°, 120°). Researchers argue that during self-rotation, participants rotate their reference frame to another axis. When stimulus angles approach the orthogonal principal axes of the participant' s unrotated reference frame (e.g., left-right, front-back axes), transformation becomes easier and less costly, resulting in optimal performance. Thus, embodied self-rotation in level-2 VPT involves aligning one' s own reference frame with others' reference frames.

In summary, during the perspective simulation stage, level-1 and level-2 VPT involve completely distinct mechanisms. Level-1 VPT achieves simulation through non-embodied, self-reference-frame-based gaze tracking, whereas level-2 VPT accomplishes perspective switching primarily through embodied, self-reference-frame-detached self-rotation and reference frame alignment.

3.3 Information Integration and Response Selection Stage: Potentially Shared Understanding of Intentions and Mental States

After simulating others' perspectives, individuals enter the information integration and response selection stage. At this point, information from one' s own perspective, simulated others' perspective information, and other social cues from the scene are integrated. Simultaneously, individuals must inhibit irrelevant perspective information and select the target perspective for final judgment according to task requirements. At this stage, level-1 and level-2 VPT may share a process of integrating information about others' intentions and mental states.

Recent research reveals that social cues, particularly behavioral intentions, can serve as high-level cues that are integrated to facilitate final response stage judgments in both level-1 and level-2 VPT. As previously discussed, gaze is important for level-1 VPT, yet some studies have not found unique gaze effects. Cole et al. (2016) also used a barrier dot-perspective task and found that regardless of whether barriers had windows, participants' self-perspective judgments were unaffected by others' perspectives. Conway et al. (2017) manipulated dot visibility using telescopes and similarly found consistency effects regardless of whether avatars could see the dots. However, these contradictory results do not completely negate the role of gaze in level-1 VPT. 李艺 et al. (2021) suggested that such contradictions may arise from other factors in experimental designs, such as avatar attributes and characteristics or instruction settings. In any case, gaze remains the most powerful influence on level-1 VPT, though it does not completely dominate task performance. Mayrand et

al. (2024) noted that gaze signals typically communicate both viewing direction and the gazing agent' s mental state. Perhaps perceptual cues about the agent' s state and behavioral intentions also influence level-1 VPT. Hu et al. (2025) provided important evidence through a series of level-1 VPT experiments. Results showed that when avatars performed both gaze and reach actions, or gaze alone, they could elicit spontaneous visual perspective taking. However, the strongest perspective-taking effect occurred only when avatars simultaneously displayed gaze and reach behaviors, with gaze preceding reach, because this sequence conforms to the normal human behavior pattern of "perception before action" and is more easily understood. This suggests that when perceptual cues clearly convey others' goal-directed intentions, they more effectively support level-1 VPT, thereby facilitating final judgments about others' visual accessibility.

Similarly, research has demonstrated that when participants perceive cues pointing to others' mental states or intentions, level-2 VPT performance is also enhanced. Lukošiūnaitė et al. (2024) compared level-2 VPT performance under "action" (avatar reaching for a cup) versus "no action" (avatar' s hands on lap or one hand beside the cup) conditions, finding that the action condition facilitated level-2 VPT. Brady et al. (2024) also confirmed in their level-2 VPT study that when avatars interacted with target objects, such as making pointing or grasping actions, participants' judgments of objects' left-right positions relative to the avatar became more accurate. Based on these findings, both level-1 and level-2 VPT may involve a process of understanding others' mental states or intentions from perceptual cues. Understanding others' goals and intentions in the current situation can serve as a cue that participates in information integration for both level-1 and level-2 VPT, subsequently influencing final decisions.

In summary, through systematic review using the three-stage framework, behavioral research evidence clearly demonstrates mechanistic similarities and differences between level-1 and level-2 VPT across three stages: During information processing, they differ in representational depth and scope—level-1 VPT represents basic spatial information related to sightline paths with limited scope, while level-2 VPT requires representation of finer spatial relations and forms. In perspective simulation, they exhibit fundamental mechanistic differences: level-1 VPT achieves simulation through non-embodied gaze tracking, while level-2 VPT accomplishes simulation through embodied reference frame rotation and alignment. In information integration and response selection, both VPT types likely share the integration of information about understanding others' behavioral intentions. Next, we delve into the neural mechanism level to examine the neural basis of this behavioral framework.

4 Neuroscientific Studies on the Mechanisms of Level-1 and Level-2 Visual Perspective Taking

Behavioral research demonstrates the similarities and differences between level-1 and level-2 VPT within the three-stage processing model, while cognitive neuroscience research further supports and deepens this model at the neural

level.

In the information processing stage, neuroimaging studies also confirm that level-1 and level-2 VPT commonly activate brain regions responsible for initial processing of visuospatial information, though with differences in activation intensity and extent. Numerous studies have detected increased blood oxygen level-dependent (BOLD) signals in secondary visual cortex, precuneus, superior parietal lobule, and left inferior parietal lobule during level-2 VPT, regions related to spatial perception and visual information processing (Wraga et al., 2005; Zacks et al., 2003). Level-1 VPT also activates left inferior parietal lobule and bilateral precuneus regions involved in visuospatial processing (Zacks & Michelon, 2005). Activation in these regions ensures that both VPT types can perceive and represent necessary visuospatial information. Additionally, a recent functional magnetic resonance imaging (fMRI) study directly compared the neural mechanisms of level-1 and level-2 VPT. Results showed that both significantly activated bilateral occipitoparietal cortex and small right frontal regions, but only level-2 VPT significantly activated a broader network including inferior frontal gyrus, precentral gyrus, inferior parietal lobule, supplementary motor area, insula, and cerebellum. Functional connectivity analysis revealed that level-2 VPT activated the dorsal attention network and frontoparietal control network significantly more strongly than level-1 VPT (Schurz et al., 2025). Researchers noted that although both level-1 and level-2 VPT involve basic visuospatial information processing mechanisms (e.g., bilateral occipitoparietal cortex) and cognitive control regions (e.g., small right frontal regions), level-1 VPT shows relatively limited neural activation, supporting the notion that it primarily relies on rapid sightline tracking. Level-2 VPT, however, has more complex neural processing pathways, activating deeper, more advanced networks to handle complex perspective conflicts and cognitive control demands (e.g., heightened activation of dorsal attention and frontoparietal control networks), and can perform advanced functions such as mental rotation and recoding spatial relations (Schurz et al., 2025). This also suggests that level-2 VPT requires higher cognitive control demands and processing depth during the information integration and response selection stage.

On the other hand, cognitive neuroscience research supports the unique embodied mechanism of level-2 VPT during the perspective simulation stage. For example, an early classic fMRI study confirmed that when individuals imagined rotating around an object (i.e., embodied self-rotation in level-2 VPT), activation primarily occurred in middle occipital gyrus, insula, supplementary motor area, extrastriate body area, and right superior parietal lobule (Wraga et al., 2005). These regions, among other functions, participate in encoding body representation (Fontan et al., 2017), but such activation was not found in level-1 VPT, supporting the hypothesis that only level-2 VPT involves embodied simulation (Gunia et al., 2021).

However, neuroimaging research remains controversial regarding the behavioral hypothesis that both level-1 and level-2 VPT involve understanding others' men-

tal states and intentions from perceptual cues. This controversy primarily centers on the right temporoparietal junction (rTPJ) and dorsomedial prefrontal cortex (dmPFC). The rTPJ and dmPFC are important hubs of the "social brain," highly interconnected with various social cognition regions and participating in theory-of-mind processes (Schurz et al., 2014). Most studies suggest that rTPJ and dmPFC play important roles in level-2 VPT (Aichhorn et al., 2006; Santiesteban et al., 2012; Seymour et al., 2018). For instance, Seymour et al. (2018) used a left-right judgment level-2 VPT paradigm combined with magnetoencephalography (MEG) and fMRI, finding that as angular differences between self and other perspectives increased, participants' response times lengthened significantly, and theta band (3-6 Hz) power in rTPJ, right anterior cingulate cortex, and right dorsolateral prefrontal cortex increased significantly. The rTPJ played a core coordinating role, connecting visual regions, mentalizing networks, and motor/body schema networks via theta oscillations (Seymour et al., 2018). Additionally, dmPFC shows significant activation in level-2 VPT (David et al., 2006; Lieberman et al., 2019; Wittmann et al., 2016). Mazzarella et al.' s (2013) fMRI study found that during level-2 VPT tasks, dmPFC BOLD signals increased linearly with angular differences between avatar and participant. However, these studies did not observe similar activation patterns in rTPJ and dmPFC during level-1 VPT tasks (Seymour et al., 2018; Wang et al., 2016). Martin et al. (2020) used High-Definition transcranial Direct Current Stimulation (HD-tDCS) to directly compare dmPFC and rTPJ roles in level-1 versus level-2 VPT, finding that stimulating either rTPJ or dmPFC produced no significant effects on level-1 VPT task performance.

However, in another HD-tDCS study, Martin et al. (2019) found that dmPFC did affect explicit level-1 VPT tasks (explicit tasks being those where participants were explicitly instructed to respond from their own or others' perspective). Additionally, Rochas et al. (2023) used high-density EEG to analyze neural mechanisms in dot-perspective tasks. Results showed that during level-1 VPT self-tasks, early brain activation primarily concentrated in basic visual attention and information integration regions, including parahippocampal gyrus, occipital lobe, and lingual gyrus. In other-tasks, brain activation began later but involved broader high-level networks: from 500 ms onward, activation significantly expanded to mentalizing networks related to theory of mind (precuneus, posterior cingulate cortex), angular gyrus, and frontoparietal executive networks responsible for cognitive control; notably, right frontal and temporoparietal region activation was particularly prominent, consistent with two previous EEG studies (Beck et al., 2018; McCleery et al., 2011). Yao et al. (2021) conducted a meta-analysis of 13 non-invasive brain stimulation studies on level-1 and level-2 VPT, finding that excitatory stimulation of rTPJ improved participants' performance in level-1 VPT other-tasks, particularly when self and other perspectives were inconsistent. This suggests that rTPJ can also inhibit self-perspective interference in level-1 VPT, facilitating judgments about others' visual accessibility; dmPFC stimulation similarly impaired participants' performance in self-tasks, possibly enhancing the salience of irrelevant other-perspective information and

thereby interfering with self-perspective judgments. Surprisingly, however, this study found no significant effects of rTPJ and dmPFC stimulation on level-2 VPT.

Bukowski (2018) noted that task type differences may be an important cause of these inconsistent results. Some studies used explicit tasks, others used implicit tasks, and some included both self- and other-tasks. However, when interpreting results, these studies did not deeply analyze why certain brain regions were only activated in one task type (Bukowski, 2018; Yao et al., 2021). As Martin et al. (2019) demonstrated, explicit level-1 VPT tasks showed activation of social cognition-related brain regions, suggesting that perhaps only when task demands or situational cues explicitly include others' states or intentions (e.g., explicit reaching actions), or when participants are explicitly required to consider others (e.g., in other-tasks), will level-1 VPT recruit such social cognition networks. Additionally, as previously mentioned, some EEG studies show that level-1 VPT only involves theory-of-mind-related brain regions at later stages (Beck et al., 2018; McCleery et al., 2011; Rochas et al., 2023), while Seymour et al. (2018) also noted that as level-2 VPT progresses, rTPJ gradually reduces connectivity with visual regions and strengthens connections with mentalizing and body schema regions. Therefore, this shared mechanism for understanding mental states and intentions more likely exists at later stages of both level-1 and level-2 VPT, particularly functioning during the information integration and response selection stage.

In summary, existing cognitive neuroscience research further supports and extends behavioral hypotheses about the internal mechanisms of the two VPT types. First, during the information processing stage, both VPT types involve mechanisms for processing and representing visuospatial information (e.g., bilateral occipitoparietal cortex), but due to different task demands, level-2 VPT additionally activates higher-level brain regions processing deeper information (e.g., heightened activation of dorsal attention and frontoparietal control networks). Second, during the perspective simulation stage, level-2 VPT uniquely involves several body-encoding brain regions (e.g., insula, supplementary motor area, extrastriate body area, and right superior parietal lobule) compared to level-1 VPT, confirming that only level-2 VPT has an embodied process. During information integration and response selection, both level-1 and level-2 VPT involve brain regions for cognitive control (e.g., right prefrontal cortex) that require inhibiting irrelevant information to judge target perspectives, though level-1 VPT demands lower cognitive load. However, controversy remains regarding whether both VPT types activate social cognition-related brain regions, though online cues with clear intentions likely stabilize the shared mechanism of understanding others' intentions and mental states in both level-1 and level-2 VPT.

Figure 1

Figure 1: Figure 1

5 Hypotheses on the Internal Mechanisms of Level-1 and Level-2 Visual Perspective Taking

This article has progressively examined the mechanistic similarities and differences between level-1 and level-2 VPT. First, through analysis of existing mindreading theories, we suggested that level-1 and level-2 VPT may involve both distinct and shared mechanisms, proposing the three-stage model. Next, we explored behavioral and cognitive neuroscience research to identify specific similarities and differences across stages. Behavioral research found that during information processing, the two VPT types differ in processing depth and scope; during perspective simulation, level-1 VPT relies more on rapid gaze tracking mechanisms, while level-2 VPT includes embodied reference frame alignment through self-rotation—a crucial internal process difference. Additionally, during information integration and response selection, both may need to integrate others' mental states and intentions from cues. Neuroscience research provides supporting evidence: both level-1 and level-2 VPT activate brain regions for processing and representing visuospatial information, but level-1 VPT more heavily involves basic visual feature processing, supporting behavioral hypotheses about the information processing stage and providing evidence for rapid sightline matching mechanisms in perspective simulation. Furthermore, level-2 VPT additionally activates body-encoding brain regions, aligning with behavioral findings of embodied mechanisms. Finally, level-2 VPT activates more advanced cognitive control mechanisms, indicating higher cognitive load during information integration and response selection. While controversy exists over whether both VPT types involve social cognition-related brain regions (primarily rTPJ and dmPFC), this controversy suggests that when cues contain explicit intentions, both level-1 and level-2 VPT likely share the mechanism of understanding others' mental states and intentions.

Based on systematic integration of the aforementioned behavioral and neuroscientific evidence, we formally propose the "three-stage model" of internal mechanisms for level-1 and level-2 VPT. The processing of both level-1 and level-2 VPT can be parsed into three stages, with distinct yet similar mechanisms operating within them (see Figure 1

):

Figure 1 Internal Mechanisms of Level-1 and Level-2 Visual Perspective Taking

The first stage is information processing. Both level-1 and level-2 VPT require processing and representing visuospatial information, such as directional cues (sightlines, body orientation) and objective environmental cues, to establish a spatial region encompassing self, other, and object. At this stage, level-1 and level-2 VPT differ in representational depth and processing scope. Level-1

VPT only needs to focus on physical conditions related to sightline paths (e.g., distance, lighting, occlusion) (Michelon & Zacks, 2006) and self-other sightline and head directions (Furlanetto et al., 2016; Marshall et al., 2018; O'Grady et al., 2020). Level-2 VPT, however, requires representing objects' specific forms and locations and others' body orientations, involving deeper processing (Samuel, Cole et al., 2023; Samuel, Salo et al., 2023; Schurz et al., 2025; Wang et al., 2025).

The second stage is perspective information simulation. After processing basic information, participants must simulate others' perspective information. At this point, level-1 and level-2 VPT operate through two different simulation mechanisms. Level-1 VPT tracks others' sightline paths to judge whether occlusion exists between objects and others; when objects exist within others' spatial regions without being blocked, they are visible (Creem-Regehr, 2013). This process is more based on physical properties of visual scenes and individuals' visual-spatial perception abilities, thus having an automatic foundation, though actual performance remains influenced by task goals and executive functions (Todd et al., 2017). Level-2 VPT is more complex. First, individuals must match their own body movement patterns to the avatar' s body posture—for example, shifting from one' s own standing posture to the avatar' s sitting posture. This simulation process consumes relatively few cognitive resources, primarily relying on postural control mechanisms that integrate visual, proprioceptive, and vestibular signals (Yeh et al., 2021). Next, individuals must determine the spatial transformation path from self-location to other-location—for example, mentally simulating leftward rotation when the other is located to the left. Finally, individuals set movement directions to align reference frames, matching their own principal axis with others' principal axes (Fischer & Demiris, 2020; Kessler & Thomson, 2010; Surtees et al., 2013; Yu & Zacks, 2017).

The third stage involves integrating information, inhibiting irrelevant perspectives, and selecting target perspectives. At this stage, individuals must not only integrate self and other perspective information but also understand intentions and mental states from cues, though this understanding may only be integrated when cues are explicit (Brady et al., 2024; Ford et al., 2024; Hu et al., 2025; Mayrand et al., 2024; Ueda et al., 2021). Integration of this information jointly influences level-1 and level-2 VPT' s inhibition of irrelevant perspective information and selection of target perspectives for judgment (Ciorli & Pia, 2023). Level-2 VPT requires stronger working memory and inhibitory control capabilities than level-1 VPT to execute this stage (Schurz et al., 2025).

6 Summary and Outlook

VPT is a critical process for individuals to understand others' perceptual experiences and is essential for social interaction and social cognition. Through systematic analysis of theoretical controversies, behavioral and neuroscientific evidence on level-1 and level-2 VPT, this article proposes a novel "three-stage integrative model."This model provides a unified and powerful framework for un-

derstanding mechanistic similarities and differences between level-1 and level-2 VPT. Future research should continue to advance along several key directions:

First, future behavioral research should focus on designing experimental paradigms that can truly dissociate the three stages and explore which stages different influencing factors affect in level-1 versus level-2 VPT. Process dissociation procedure (PDP) studies have already attempted this. PDP is an analytical method that assumes all task performance is simultaneously controlled by resource-demanding controlled processes and fast, unintentional automatic processes. In VPT self-tasks, the automatic process computes others' perspectives, while the controlled process selects self-perspective. In consistent conditions, correct responses represent successful operation of both automatic and controlled processes (since other-perspective matches target perspective); in inconsistent conditions, error responses represent controlled process failure with automatic process success. Substituting individuals' behavioral data from these two conditions into specific mathematical models can calculate the strength of both processes, thereby exploring how different influencing factors affect these processes (Qureshi & Monk, 2018; Todd et al., 2019; Todd et al., 2021). PDP research can examine how the same influencing factor affects different VPT processes—for example, Todd et al. (2017) used PDP to find that time pressure weakened VPT' s controlled process but not its automatic process. However, PDP does not truly separate different processes within the same task programmatically, and its division into automatic and controlled stages is overly simplistic and mechanistic. Therefore, future research should strive to develop experimental paradigms that can dissociate different VPT stages and further explore whether the same influencing factor produces identical effects on different stages of level-1 and level-2 VPT, thereby providing further evidence for mechanistic differences between the two VPT types.

Second, future neuroimaging research should enhance temporal dynamics studies and examine task types to resolve controversies about social cognition brain regions. We have proposed model hypotheses for the three stages of level-1 and level-2 VPT, but are these stages strictly serial or partially overlapping? What are their respective temporal relationships? Answering these questions requires increased application of high-temporal-resolution techniques (e.g., EEG/MEG). By analyzing activation sequences and functional connectivity patterns of specific brain networks (e.g., body representation networks, social brain networks) across different time windows, we can test the temporal dynamics of the three-stage model and provide more direct evidence. Furthermore, substantial controversy remains regarding whether core social cognition brain regions such as rTPJ and dmPFC jointly participate in both VPT types and under what conditions (Martin et al., 2020; Yao et al., 2021). This inconsistency likely stems from insufficient consideration of task paradigm heterogeneity. Future research should consciously compare explicit versus implicit tasks and conditions with versus without explicit intention cues (e.g., pointing, grasping actions) to clarify the conditions for social cognition brain region involvement.

Finally, future research can continuously expand the model hypotheses proposed here to broaden their applicability and explanatory power. For instance, deeper investigation of the characteristics and internal processes of level-2 VPT' s embodied mechanisms is needed. Although level-2 VPT is widely recognized as having this embodied self-rotation mechanism, the conditions under which this mechanism is mobilized remain unclear. Janczyk (2013) conducted a series of psychological refractory period (PRP) tasks. PRP is a classic dual-task interference paradigm where participants perform two tasks simultaneously. If both tasks require central processing system resources, then when the two tasks are presented in rapid succession, the second task' s response time is significantly prolonged because the central processing stage has a "bottleneck" that can only handle one task' s "decision" component at a time (Pashler, 1994). Notably, Janczyk (2013) found that when only small-angle rotations were required, the interval between the two tasks did not affect level-2 VPT performance, indicating that level-2 VPT was unconscious and could proceed in parallel with other processes without being limited by central capacity. Only when large angles exceeding 60° needed processing did level-2 VPT begin to occupy central capacity. This suggests that the body-reference-frame-based simulation mechanism may not be fully triggered in low-difficulty tasks, raising the question of what processing mechanism individuals rely on to complete level-2 VPT under such conditions—currently lacking direct evidence. Therefore, when constructing mechanistic hypotheses, this article referenced Yeh et al.' s (2021) proposal to preliminarily divide the mental simulation stage into two subprocesses: automated posture matching and subsequent rotation alignment. This division remains theoretical speculation awaiting future empirical support.

Additionally, "embodied" represents "putting oneself in another' s shoes." Existing research has primarily examined the roles of visual information and body representation, but whether other sensory information also promotes embodied processes remains unexplored—such as auditory, tactile, or olfactory information. Guo et al. (2024) showed that during level-2 VPT, auditory information can also be integrated when establishing an other-centered reference frame. However, few researchers have examined cross-modal interactions in other-centered reference frames. Beyond this expansion, future research should not be limited to current hypotheses but should more deeply explore characteristics of other processes.

References

李鸿锴. (2025). 读心系统理论之争：双系统还是单系统?. 系统科学学报, 33(1), 46-51.

李艺, 肖风. (2021). 自动观点采择：内隐心智化与潜心智化的争议. 心理科学进展, 29(10), 1887-1900.

邵雨婷, 李伟健, 孙炳海, 张文海. (2020). 视觉空间观点采择对教师共情的影响: 自我表征抑制和自我视空转换的不同作用. 心理科学, 43(4), 871-878.

吴梦慧, 谢久书, 邓铸. (2022). 视觉观点采择中自我中心性偏差的抑制和归因之争. 心理科学进

展, 30(1),

肖承丽, 隋雨檠, 肖苏衡, 周仁来. (2021). 空间交互研究新视角：多重社会因素的影响. 心理科学进展, 29(5),

张越, 葛贤亮, 田志强, 葛列众. (2018). 基于空间的一级和二级视角转换的行为研究及理论综述. 心理科学, 41(2), 504-510.

赵婧, 王璐, 苏彦捷. (2010). 视觉观点采择的发生发展及其影响因素. 心理发展与教育, 26(1), 107-111.

赵杨柯, 钱秀莹. (2010). 自我中心视角转换——基于自身的心理空间转换. 心理科学进展, 18(12),

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

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? Psychological Review, 116(4), 953-970.

Apperly, I., Riggs, K., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? Psychological Science, 17(10), 841-844.

Baker, L. J., Levin, D. T., & Saylor, M. M. (2016). The extent of default visual perspective taking in complex layouts. Journal of Experimental Psychology: Human Perception and Performance, 42(4), 508-516.

Beck, A. A., Rossion, B., & Samson, D. (2018). An objective neural signature of rapid perspective taking. Social Cognitive and Affective Neuroscience, 13(1), 72-79.

Bohl, V., & van den Bos, W. (2012). Toward an integrative account of social cognition: Marrying theory of mind and interactionism to study the interplay of type 1 and type 2 processes. Frontiers in Human Neuroscience, 6,

Brady, N., Leonard, S., & Ní Choisdealbha, Á. (2024). Visual perspective taking and action understanding. Acta Psychologica, 249, 104467.

Bukowski, H., (2018). The neural correlates of visual perspective taking: A critical review. Current Behaviour Neuroscience, 5, 189-197.

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. Mind & Language, 28(5),

Carruthers, P. (2016). Two systems for mindreading?. Review of Philosophy and Psychology, 7(1), 141-162.

Ciorli, T., & Pia, L. (2023). Spatial perspective and identity in visual awareness of the bodily self-other distinction. Scientific Reports, 13(1), 14994.

Cole, G. G., & Millett, A. C. (2019). The closing of the theory of mind: A critique of perspective-taking. Psychonomic Bulletin & Review, 26(6), 1787-1802.

Cole, G. G., Atkinson, M., Le, A. T. D., & Smith, D. T. (2016). Do humans spontaneously take the perspective of others? Acta Psychologica, 164, 165–168.

Cole, G. G., Millett, A. C., Samuel, S., & Eacott, M. J. (2020). Perspective-taking: In search of a theory. Vision, 4(2), 30.

Conway, J. R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (2017). Submentalizing or mentalizing in a level 1 perspective-taking task: A cloak and goggles test. Journal of Experimental Psychology: Human Perception and Performance, 43(3), 454–465.

Creem-Regehr, S. H., Gagnon, K. T., Geuss, M. N., & Stefanucci, J. K. (2013). Relating spatial perspective taking to the perception of other' s affordances: Providing a foundation for predicting the future behavior of others. Frontiers in Human Neuroscience, 7, 596.

David, N., Bewernick, B. H., Cohen, M. X., Newen, A., Lux, S., Shah, N. J., & Vogeley, K. (2006). Neural representations of self versus other: Visual-spatial perspective taking and agency in a virtual ball-tossing game. Journal of Cognitive Neuroscience, 18(6), 898–910.

Elekes, F., Varga, M., & Király, I. (2016). Evidence for spontaneous level-2 perspective taking in adults. Consciousness and Cognition, 41, 93–103.

Fischer, T., & Demiris, Y. (2020). Computational modeling of embodied visual perspective-taking. IEEE Transactions on Cognitive and Developmental Systems, 12(4), 723–732.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children' s knowledge about visual perception: Further evidence for the level 1–level 2 distinction. Developmental Psychology, 17(1), 99–103.

Fontan, A., Cignetti, F., Nazarian, B., Anton, J. L., Vaugoyeau, M., & Assaiante, C. (2017). How does the body representation system develop in the human brain?. Developmental Cognitive Neuroscience, 24, 118–128.

Ford, B., Monk, R., Litchfield, D., & Qureshi, A. (2024). Agent-object relationships in level-2 visual perspective taking: An eye-tracking study. Journal of Cognition, 7(1), 72.

Furlanetto, T., Becchio, C., Samson, D., & Apperly, I. A. (2016). Altercentric interference in level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. Journal of Experimental Psychology: Human Perception and Performance, 42(2), 158–163.

Gómez-Tabares, A.-S. (2023). Is there continuity from implicit recognition of intentional action in infants to explicit mindreading in preschoolers? Systematic review of longitudinal evidence and theoretical implications. Journal for the Study of Education and Development, 46(4), 950–982.

Gunia, A., Moraresku, S., & Vlček, K. (2021). Brain mechanisms of visuospatial perspective-taking in relation to object mental rotation and the theory of mind.

Behavioural Brain Research, 407, 113247.

Guo, G., Wang, N., Sun, C., & Geng, H. (2024). Embodied cross-modal interactions based on an altercentric reference frame. Brain Sciences, 14(4), 314.

Hu, X., Xu, H., Chen, H., Shen, M., & Zhou, J. (2025). Good to see you R2-D2: Inducing spontaneous perspective-taking towards non-human agents through human-like gaze and reach. Cognition, 259, 106101.

Jacob, P. (2019). Challenging the two-systems model of mindreading. In A. Avramides & M. Parrott (Eds.), Knowing Other Minds (pp. 79–106). Oxford University Press.

Janczyk M. (2013). Level 2 perspective taking entails two processes: Evidence from PRP experiments. Journal of Experimental Psychology: Learning, Memory, and Cognition, 39(6), 1878–1887.

Kelly, J. W., Beall, A. C., & Loomis, J. M. (2004). Perception of shared visual space: Establishing common ground in real and virtual environments. Presence, 13(4), 442–450.

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

Kloo, D., Kristen-Antonow, S., & Sodian, B. (2020). Progressing from an implicit to an explicit false belief understanding: A matter of executive control? International Journal of Behavioral Development, 44(2),

Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M., & Tan, K. M. (2019). Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): Causal, multivariate, and reverse inference evidence. Neuroscience and Biobehavioral Reviews, 99, 311–328.

Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. Child Development Perspectives, 10(3), 184–189.

Lukošiūnaitė, I., Kovács, Á. M., & Sebanz, N. (2024). The influence of another' s actions and presence on perspective taking. Scientific Reports, 14(1), 4971.

Marshall, J., Gollwitzer, A., & Santos, L. R. (2018). Does altercentric interference rely on mentalizing?: Results from two level-1 perspective-taking tasks. Plos One, 13(3), e0194101.

Martin, A. K., Huang, J., Hunold, A., & Meinzer, M. (2019). Dissociable roles within the social brain for self-other processing: A HD-tDCS study. Cerebral Cortex, 29(8), 3642–3654.

Martin, A. K., Kessler, K., Cooke, S., Huang, J., & Meinzer, M. (2020). The right temporoparietal junction is causally associated with embodied perspective-taking. The Journal of Neuroscience, 40(15), 3089–3095.

Mayrand, F., Capozzi, F., & Ristic, J. (2024). Gaze communicates both cue direction and agent mental states. Frontiers in Psychology, 15, 1472538.

Mazzarella, E., Ramsey, R., Conson, M., & Hamilton, A. (2013). Brain systems for visual perspective taking and action perception. Social Neuroscience, 8(3), 248–267.

McCleery, J. P., Surtees, A. D. R., Graham, K. A., Richards, J. E., & Apperly, I. A. (2011). The neural and cognitive time course of theory of mind. Journal of Neuroscience, 31(36), 12849–12854.

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. Perception & Psychophysics, 68(2),

Müsseler, J., von Salm-Hoogstraeten, S., & Böffel, C. (2022). Perspective taking and avatar-self merging. Frontiers in Psychology, 13, 714464.

Negen J. (2025). Mental rotation, perspective taking, and performance profiling. Cognitive Processing, 26(3),

O' Grady, C., Scott-Phillips, T., Lavelle, S., & Smith, K. (2020). Perspective-taking is spontaneous but not automatic. Quarterly Journal of Experimental Psychology, 73(10), 1605–1628.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. Psychological Bulletin, 116(2),

Pesimena, G., & Soranzo, A. (2023). Both the domain-general and the mentalising processes affect visual perspective taking. Quarterly Journal of Experimental Psychology, 76(3), 469–484.

Quesque, F., Chabanat, E., & Rossetti, Y. (2018). Taking the point of view of the blind: Spontaneous level-2 perspective-taking in irrelevant conditions. Journal of Experimental Social Psychology, 79, 356–364.

Qureshi, A. W., & Monk, R. L. (2018). Executive function underlies both perspective selection and calculation in level-1 visual perspective taking. Psychonomic Bulletin & Review, 25(4), 1526–1534.

Rochas, V., Montandon, M.-L., Rodriguez, C., Herrmann, F. R., Eytan, A., Pegna, A. J., Michel, C. M., & Giannakopoulos, P. (2023). Mentalizing and self-other distinction in visual perspective taking: The analysis of temporal neural processing using high-density EEG. Frontiers in Behavioral Neuroscience, 17, 1206011.

Samson, D., & Apperly, I. A. (2010). There is more to mind reading than having theory of mind concepts: New directions in theory of mind research. Infant and Child Development, 19(5), 443–454.

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 compu-

tation of what other people see. Journal of Experimental Psychology: Human Perception and Performance, 36(6), 1255–1266.

Samuel, S., Cole, G. G., & Eacott, M. J. (2023). It' s not you, it' s me: A review of individual differences in visuospatial perspective taking. Perspectives on Psychological Science, 18(2), 293–308.

Samuel, S., Erle, T. M., Kirsch, L. P., Surtees, A., Apperly, I., Bukowski, H., Auvray, M., Catmur, C., Kessler, K., & Quesque, F. (2024). Three key questions to move towards a theoretical framework of visuospatial perspective taking. Cognition, 247, 105787.

Samuel, S., Salo, S., Ladvelin, T., Cole, G. G., & Eacott, M. J. (2023). Teleporting into walls? The irrelevance of the physical world in embodied perspective-taking. Psychonomic Bulletin & Review, 30(3), 1011–1019.

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2012). Enhancing social ability by stimulating right temporoparietal junction. Current Biology, 22(23), 2274–2277.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. Neuroscience and Biobehavioral Reviews, 42, 9–34.

Schurz, M., Tholen, M. G., Kronbichler, M., Perner, J., & Surtees, A. D. R. (2025). Comparing level 1 and level 2 visuo-spatial perspective-taking in the brain: Evidence from fMRI. Social Neuroscience. Advance online publication.https://doi.org/10.1080/17470919.2025.2490574

Seymour, R. A., Wang, H., Rippon, G., & Kessler, K. (2018). Oscillatory networks of high-level mental alignment: A perspective-taking MEG study. NeuroImage, 177, 98–107.

Surtees, A. D., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of level-2 perspective-taking in children and adults. The British Journal of Developmental Psychology, 30(1), 75–86