Brain mechanisms of visuospatial perspective-taking in relation to object mental rotation and the theory of mind

Anna Gunia\textsuperscript{a, b, *}, Sofia Moraresku\textsuperscript{a, b}, Kamil Vlcek\textsuperscript{a, *}

\textsuperscript{a} Laboratory of Neurophysiology of Memory, Institute of Physiology of the Czech Academy of Sciences, Vídeňská 1083, 142 20 Prague, Czech Republic
\textsuperscript{b} Third Faculty of Medicine, Charles University, Ruská 87, 100 00 Prague, Czech Republic

\textbf{ARTICLE INFO}

\textbf{ABSTRACT}

Visuospatial perspective-taking (VPT) is a process of imagining what can be seen and how a scene looks from a location and orientation in space that differs from one’s own. It comprises two levels that are underpinned by distinct neurocognitive processes. Level-2 VPT is often studied in relation to two other cognitive phenomena, object mental rotation (oMR) and theory of mind (ToM). With the aim to describe the broad picture of neurocognitive processes underlying level-2 VPT, here we give an overview of the recent behavioral and neuroscientific findings of level-2 VPT. We discuss its relation to level-1 VPT, which is also referred to as perspective-tracking, and the neighboring topics, oMR and ToM. Neuroscientific research shows that level-2 VPT is a diverse cognitive process, encompassing functionally distinct neural circuits. It shares brain substrates with oMR, especially those parietal brain areas that are specialized in spatial reasoning. However, compared to oMR, level-2 VPT involves additional activations in brain structures that are typically involved in ToM tasks and deal with self/other distinctions. In addition, level-2 VPT has been suggested to engage brain areas coding for internal representations of the body. Thus, the neurocognitive model underpinning level-2 VPT can be understood as a combination of visuospatial processing with social cognition and body schema representations.

1. Introduction

Visuospatial perspective-taking (VPT) is a complex cognitive process that has two levels. Level-1 VPT, also known as perspective-tracking, is the process of calculating what can and cannot be seen from another location and orientation in space. In contrast, level-2 VPT is a process of imagining how a scene looks from a position that differs from one’s own \cite{1}. The main aim of this review is to give an overview of the cognitive neuroscience behind level-2 VPT. Therefore, first, we discuss different types of VPT and, hereby, we review the neuroscientific background of differentiating and representing self and others’ perspectives.

Second, level-2 VPT is often studied in relation to object mental rotation (oMR) since both of them are understood in terms of the combination of visuospatial processing with social cognition and body schema representations.

\textbf{ARTICLE INFO}

\textbf{Keywords:}
Visuospatial perspective-taking
Object mental rotation
Theory of mind
Embodiment
Temporoparietal junction
Medial prefrontal cortex

Abbreviations: ACC, anterior cingulate cortex; AD, Alzheimer’s disease; aMCI, amnestic mild cognitive impairment; ASD, autism spectrum disorder; ATL, anterior temporal lobe; DMN, default mode network; dmPFC, dorsomedial prefrontal cortex; EBA, extrastriate body area; EF, executive functions; FEF, frontal eye fields; HD-tDCS, high-definition transcranial direct current stimulation; IFG, inferior frontal gyrus; Ins, insula; IPL, inferior parietal lobule; IPFC, lateral prefrontal cortex; M1, primary motor cortex; MFG, middle frontal gyrus; MOG, middle occipital gyrus; mPFC, medial prefrontal cortex; oMR, object mental rotation; PCC, posterior cingulate cortex; PCUN, precuneus; PFC, prefrontal cortex; RT, reaction time; rTPJ, right temporoparietal junction; SFG, superior frontal gyrus; SMA, supplementary motor area; SMG, supramarginal gyrus; SPL, superior parietal lobule; TMS, transcranial magnetic stimulation; ToM, theory of mind; TPJ, temporoparietal junction; vmPFC, ventromedial prefrontal cortex; vOC, ventral occipital cortex; VPT, visuospatial perspective-taking.

\textbf{E-mail addresses:} anna.gunia@fgu.cas.cz (A. Gunia), sofia.moraresku@fgu.cas.cz (S. Moraresku), kamil.vleek@fgu.cas.cz (K. Vlcek).

https://doi.org/10.1016/j.bbr.2021.113247
Received 21 November 2020; Received in revised form 8 March 2021; Accepted 12 March 2021
Available online 18 March 2021
0166-4328/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
concept of theory of mind (ToM), which implies thinking of others’ mental states [1,4–6]. Hence, in the last section of the review, we discuss the links between level-2 VPT and ToM on the behavioral and neural levels.

In regular conditions, we view an environment from our own, first-person (self, egocentric) visuospatial perspective, where our multi-modal experience of space is centered upon our own body [7]. On the other hand, visuospatial perspective-taking (VPT) is the ability to imagine viewing a scene from another, sometimes a third-person perspective. To successfully judge how a scene would look from another perspective, we need to overcome our own egocentric biases and be able to give priority to the perspective that is distinct from our actual one [1].

2. Visuospatial perspective-taking

Observing VPT abilities in humans was pioneered by Piaget using a construction model of three mountains and a doll [8]. During the experiment, children were asked to answer whether the doll could or could not see particular objects sometimes hidden behind the layers of mountains (later termed as level-1 VPT, also referred to as perspective-tracking) or how would the scene of the mountains look from the doll’s perspective (later termed as level-2 VPT). Their results indicated that before the age of 9–10, it is difficult for children to fully overcome their egocentric visuospatial biases and make judgments from another’s perspective [8]. Later, Flavell [9] differentiated level-1 VPT (perspective-tracking) from level-2 VPT and characterized the first one as estimating what is visible and what occluded for others, and the second one as understanding that others might see the same things as we do, but differently [9]. Both processes, perspective-tracking and level-2 VPT, involve judgments about another’s viewpoint. However, perspective-tracking relies on computing a line of sight of another subject [10], and can be performed by drawing a line between an agent and target [11], whereas level-2 VPT requires mental rotation of the self into another’s viewpoint [10]. Perspective-tracking was reported to be developed at the age of 2 years [12]. In contrast, level-2 VPT starts to show around 4–5 years of age as more recent investigations have demonstrated [13]. Perspective-tracking, but not level-2 VPT, has also been observed in apes [14,15] and corvids [16].

Sometimes, spatial and visual perspective-taking are studied separately [1,17]. Spatial perspective-taking refers to the ability to correctly estimate a spatial relationship (actual or imagined) between another person and an object, whereas visual perspective-taking enables us to estimate how an object is seen by another person [17]. In a series of behavioral experiments, level-1 visual perspective-taking was studied by judging what was visible and what was not for an avatar; while, level-1 spatial perspective-taking was studied by estimating what was in front and what behind the avatar. On the other hand, a level-2 spatial perspective-taking task involved the left/right judgments from the avatar’s perspective, while its visual analogue was estimated by judging how alphanumeric symbols would look like from the avatar’s perspective [17]. According to their results, in contrast to level-1 and level-2 perspective-taking, visual and spatial perspective-taking exhibited similar behavioral patterns [1,17]. Level-2 visual and spatial perspective-taking were similarly influenced by the modulation of angular disparity between the observer and the imagined position [17], whereas the disparity had no effect on level-1 visual and spatial perspective-taking [1]. In the following text, we follow the practice of the current studies [5,6,18] that examine and operationalize visual and spatial perspective-taking together but differentiate level-1 and level-2 perspective-taking and term them as perspective-tracking and level-2 VPT, respectively.

Level-2 VPT is executed by transforming visuospatial mental images from our first-person perspective into another perspective. Visuospatial mental images can be viewed as “pictures in the head”, and their transformation depends on representations in various spatial frames. Human cognition is assumed to embrace three major types of reference frames: object-based, environmental, and egocentric [2,19]. Object-based reference frames locate things relative to one or more axes of the particular object (e.g. up/down of a cube). They are used to determine a relationship between parts of an object or place an object relative to another one. Environmental reference frames localize things relative to a fixed space (e.g. north/south/east/west). Egocentric reference frames position things relative to the axes of the self (up/down, back/front, left/right from me) and are important for governing body movements in space, grasping/reaching out an object, and navigating [2]. From multiple types of egocentric reference frames the one that has been associated with perspective transformations is centered upon our eyes (head/gaze) [2]. During level-2 VPT, an observer imagines transforming one’s own visuospatial perspective so that observer’s egocentric reference frame moves relative to a reference frame of an environment and to a reference frame of an observed object, e.g., observer imagines moving around a cube in a room [2].

Level-2 VPT is thought to be an embodied process, suggesting that it depends on a position and posture of a body [10]. Embodied nature of level-2 VPT is closely linked with the notion of the body schema. Body schema is an internal representation of the subject’s body position in space, also referred to as “self” and viewed as a starting point of level-2 VPT [4,20]. The level-2 VPT process is understood as a mental transformation of this body schema into another’s location and orientation in space, also termed as “embodiment” of another’s viewpoint [4,20]. Amorim et al. [21] suggested that two types of embodiment (spatial and motoric) are relevant for mental transformation of the body figures. Spatial embodiment is the process of mapping one’s own body axes onto the reference posture. In contrast, during motoric embodiment, the sensorimotor system of the brain emulates the observed posture by mentally adopting the same posture [21]. By further developing the notion of motoric embodiment, Kessler and Thomson [4], suggested that endogenous motoric embodiment, which is a self-initiated emulation of a body movement (not only a posture), like imagining to rotate a body into a new orientation, is related to level-2 VPT.

3. Neural and behavioral perspectives of visuospatial perspective-taking

3.1. Behavioral perspective

Behavioral experiments demonstrated that participants perform differently during self-visuospatial perspective, perspective-tracking, and level-2 VPT judgments [5,22,23]. In regular conditions, performing a task from the visuospatial perspective of oneself seems to take less effort than perspective-tracking [22,23], or level-2 VPT [23,24]. Besides, level-2 VPT seems more difficult to perform than perspective-tracking, as the former is characterized by longer reaction times (RT) and more errors [5,17].

The embodied nature of level-2 VPT was suggested by the series of behavioral experiments conducted by Kessler and Thomson [4]. In their experiments, participants made the left/right judgments from an avatar’s perspective. In every trial, the avatar was sitting at a round table, but the angular disparity between a participant and the avatar was varying from trial to trial. Depending on a trial-type, the avatar was presented with 0°, 40°, 80°, 120°, or 160° angular disparity from the participant. The 0° angular disparity between the participant and the avatar was similar to the self-perspective of the participant. In addition to the varying angular disparity, the effect of participants’ body posture on the level-2 VPT process was measured by asking participants to turn the chair they were sitting on in the clockwise or counterclockwise direction, while their gaze remained directed to the computer screen. Also, the avatar, sitting at different angles around the table, was presented in a clockwise or counterclockwise direction from the participant’s self-perspective. Thus, participants’ body position was either congruent or incongruent with the direction of the avatar. The results showed significant main effects of the avatar presentation angle and
participant’s body posture congruence on the RT, and a significant interaction between angle and congruence. The participants’ body posture incongruence systematically slowed their performance in those trials in which the avatar was presented above 40°. The trials, in which the avatar was presented at 160° and the participants’ body posture was incongruent, were the most time-consuming. These results indicated that the participants’ body posture, therefore, the match/mismatch of the proprioceptive information with the target, in addition to the cognitive load introduced by varying angular disparity had an effect on the level-2 VPT process [4].

Further experiments demonstrated that larger angular disparity and participants’ body posture incongruence with avatar’s position had a negative influence only on level-2 VPT and had no significant effect on perspective-tracking [5,6,10,17]. Therefore, the effect of angular disparity and posture incongruence suggested that level-2 VPT, but not perspective-tracking, is an embodied process.

3.2. fMRI perspective

Increased blood-oxygen-level-dependent (BOLD) signal was detected in response to level-2 VPT in various brain areas associated with visuospatial reasoning, namely in the precuneus, superior parietal lobule (SPL) [2,23], and in the left inferior parietal lobule (IPL) [23]. Additionally, brain areas involved in social cognition, such as the temporoparietal junction (TPJ) [23,24] and medial prefrontal cortex (mPFC) [24], were activated. The brain areas reported during level-2 VPT tasks also include the premotor area in the middle frontal gyrus (MFG), inferior frontal gyrus (IFG) [23,24], and the brain areas, such as the insula, supplementary motor area (SMA), and extrastriate body area (EBA) located in the middle occipital gyrus (MOG) [25], which among other functions are also involved in coding for body representations [26,27] (Fig. 1).

Contrary to level-2 VPT, the fMRI study measuring neural substrates of perspective-tracking did not reveal the activation in the TPJ [22]. However, some of the visuospatial brain areas observed during level-2 VPT, such as the left IPL and bilateral precuneus, and also frontal brain areas, MFG and IFG, were detected during perspective-tracking as well [22] (Fig. 1).

Thus, level-2 VPT and perspective-tracking rely on partially distinct neurocognitive processes. On top of the overlapping brain areas, level-2 VPT recruits additional brain region reported in visuospatial tasks, such as the SPL, the brain areas playing various roles in body representations, such as the insula, SMA, and EBA, and also the brain region associated with social reasoning - the TPJ. This is in line with the behavioral accounts showing that level-2 VPT is a more complex cognitive process, which takes more effort to perform and is encountered later in human development than perspective-tracking.

Performing a task from a self-visuospatial perspective compared to third-person perspective-tracking and level-2 VPT, resulted in stronger activation in the ventromedial PFC (vmPFC) [23,24], insula [22,24],...
and posterior cingulate cortex (PCC) [22,23]. These brain regions are part of the default mode network (DMN), observed during the resting state and often discussed in connection with the neural basis of self-consciousness [22,23,28]. Self-consciousness has been reported to unite the consciousness of one’s own perceptions, attitudes, opinions, and intentions [22]. As suggested by developmental psychologists, it develops through the process of comparing and differentiating oneself from others [29]. Viewing an environment from the self-visuospatial perspective enables us to multimodally perceive and experience a space that is centered upon our own body [7]. Besides, representing one’s own visuospatial perspective implies the process of distinguishing between one’s own and others’ perspectives [22]. Therefore, being aware of one’s own visuospatial perspective was considered as a part of self-consciousness [22,23]. The higher vmPFC and PCC activity observed during representing a first-person perspective compared to level-2 VPT can also be attributed to a less cognitive demand that causes less deactivation of the DMN brain areas [23]. In contrast to performing a task from a self-perspective, taking a third-person perspective (level-2 VPT) seems to involve social processing with considerable visuospatial cognitive load, as an observer needs to describe a scene not as one sees it, but is required to imagine how it looks from another’s visuospatial perspective (Fig. 2).

3.3. Electrophysiological perspective

Important accounts on the neural substrates of level-2 VPT came from electrophysiological studies. Results of several EEG studies correspond to the fMRI investigations, as they also stressed the active involvement of the TPJ in level-2 VPT. Arzy et al. [30] suggested that level-2 VPT was accompanied by event-related potential (ERP) responses originating from the TPJ with a mean onset of 367 ms [30]. Another study also indicated that the ERPs recorded over the TPJ and supposedly over the premotor area in the time interval of 330–420 ms represent a level-2 VPT task performance manifestation [31]. The ERP source localized in the TPJ during level-2 VPT has also been reported by Thirioux et al. [32], but with a bit later timing of 517–628 ms [32]. However, this later onset can be due to a different experimental design [32].

Other insights came from MEG studies. The theta band oscillations accompanying the level-2 VPT have been localized to a brain network including the somatosensory area - supramarginal gyrus (SMG), SMA, EBA in the bilateral MOG, IFPC, and the right TPJ (rTPJ) [5]. Importantly, the study results demonstrated that the rTPJ was crucial for embodied processes underlying level-2 VPT, as the primary cortical origin of the theta power increase in response to both effects, increasing angular disparity and body posture incongruence, was located in the right hemisphere TPJ. Functional connectivity studies also showed the important role of the rTPJ in level-2 VPT. The greater angular disparity between a viewer’s and an avatar’s perspective in a level-2 VPT task was reported to be followed by a higher theta band power in the right hemisphere, in the rTPJ, right lateral PFC (IFPC), and the right anterior cingulate cortex (ACC), in the interval of 0–650 ms after stimulus onset. At the same time, oscillation-based functional connectivity between the rTPJ and both IFPC and ACC was also increased in the theta band [6]. Interestingly, the theta oscillatory power in response to level-2 VPT first emerged in the rTPJ in the time period of 0–500 ms, and only then in the right IFPC and the right ACC at 200–500 ms post-stimulus. This time course suggests the engagement of the rTPJ throughout the process of
level-2 VPT, while the IFPC and ACC get involved later on. The causal influence in this frequency band, analyzed by Granger causality, was observed from the right IFPC and right ACC to the rTPJ, but not vice versa. This directed connectivity taking place at a later phase of the task performance was suggested to correspond to the “top-down” executive function processes that inhibit the self-perspective and give priority to the imagined perspective of others [6]. Executive functions (EF) refer to cognitive processes that underpin goal-directed behavior. Three basic EF are following: 1) inhibition - the ability to inhibit task-irrelevant information and predominant responses; 2) switching or shifting, also referred to as cognitive flexibility - the ability to switch among tasks, operations or mental sets; 3) working memory or updating - the ability to refresh information temporarily stored in working memory [33,137].

A whole-brain phase-coherence theta band analysis of MEG data from the same study suggested a close relation between the cognitive processes underlying level-2 VPT and ToM on the one hand, and level-2 VPT and the body schema representation on the other hand. In response to the increased angular disparity, the rTPJ increased its phase-coherence with the dorsal parts of the mPFC, and PCC [6]. These brain regions have been associated with the ToM network [3,34,35]. The increased theta band phase-coherence was also observed between the rTPJ and brain regions associated with the body schema representations [36], namely, the SMG, posterior parietal cortex (PPC), and the supplementary motor area ( SMA) [6].

Compared to level-2 VPT, perspective-tracking did not show the oscillatory theta activity in the brain network comprising the TPJ, somatosensory, and motor-related areas [5]. Instead, the theta activity in response to perspective-tracking was recorded in the frontal eye fields (FEF, located in the vicinity of the MFG), suggesting the potential role of this brain region in inferring another’s line of sight [5]. Also, the theta power increase in response to perspective-tracking was observed within the medial parts of the ventral occipital and parietal cortices in the interval of 0–650 ms [6]. Contrary to level-2 VPT, these theta oscillations were not reported to be affected by the change in the angular disparity between an observer and an avatar [5,6], nor by the change in participant’s body posture [5]. These results are in concordance with the behavioral results that did not show the influence of the angular disparity and participant’s body posture on the perspective-tracking performance [5,6,10,17] (Fig. 1).

However, reports about the involvement of the TPJ in tracking another’s perspective do not coincide [5,6,37,38]. McCleery et al. [37] have reported the ERP component source localized over bilateral TPJ at ~400–500 ms during perspective-tracking. In this study, participants were asked to respond either from their own or from a third-person’s perspective interchangeably within each block of trials. On the other hand, the experimental designs used by Wang et al. [5] and Seymour et al. [6], which only required the third-person perspective judgments, did not demonstrate the involvement of the TPJ during the perspective-tracking tasks. According to McCleery et al. [37], in their perspective-tracking study, the role of the TPJ was to distinguish and represent one’s own and another’s perspective [37]. Yet, the involvement of the TPJ in another’s perspective judgments was not shown in a transcranial magnetic stimulation (TMS) study measuring perspective-tracking, which also required responding from self or another’s visuospatial perspective [38]. Instead, Santiesteban et al. [38] suggested that the role of the TPJ in perspective-tracking tasks could be explained by domain-general attentional processes. In this study, participants performed the same task as the one used by McCleery et al. [37]. But, Santiesteban et al. [38] split the trials so that throughout each block participants were required to respond only from their own or from another perspective. After applying disruptive repetitive TMS on the rTPJ, the performance was impaired on those trials where participants were responding how many dots were visible from their own perspective. Crucially, these trials required the participants to disregard an avatar’s perspective that was looking at the different number of dots. When an arrow was used instead of the avatar in control trials, the rTPJ activity disruption also resulted in impaired performance. In contrast, participants’ performance was not affected by disrupting the rTPJ activity in those trials where they were required to represent a third-person perspective [38]. According to Santiesteban et al. [38], on self-perspective trials participants are required to overcome an attentional cueing effect caused by the avatar’s gaze or the pointing direction of the arrow. Participants need to reorient their attention to the whole scene to detect all the dots that are presented on the screen, not only the ones that are in the visual field of the avatar or are pointed by the arrow. On the other hand, in other-perspective trials, participants are not required to reorient their attention from that side of the scene, which is already directed by the central stimuli (the avatar or arrow) [38].

Neuroscientific literature suggests that the TPJ among many other processes is also related to attentional reorientation [39]. According to these results, the supposed role of the rTPJ in perspective-tracking can be explained by the domain-general attentional processes, but not by representing another’s perspective [38].

In summary, electrophysiological studies add credibility to the view that perspective-tracking and level-2 VPT are two distinct neuro-cognitive processes. Level-2 VPT, but not perspective-tracking, seems to involve the transformation of the self in an embodied manner into another’s location, and the TPJ apparently plays a crucial role in this process. One EEG study reported the engagement of the TPJ in perspective-tracking as well [37]. However, another study using TMS demonstrated that the observed engagement of the TPJ in certain types of perspective-tracking tasks can be explained by its role in domain-general attentional processes instead of representing another’s perspective [38]. According to the recent MEG results, the neural signature of level-2 VPT seems to be as follows: the theta band power increase in the network of the TPJ and EF brain areas, namely the IFPC and ACC; the theta band synchronization between the TPJ and other ToM related brain areas, such as the mPFC and PCC; and also the phase-coupling in the same band between the TPJ and brain areas associated with the body schema representations, such as the SMG, PPC, and SMA [6] (Fig. 1).

3.4. Neuropsychological perspective

Visuospatial abilities were shown to be diminished in patients with Alzheimer’s disease (AD), the condition characterized by neurodegeneration of the medial and posterior temporal, parietal, cingulate, and frontal cortices [40,41]. Patients with amnestic mild cognitive impairment (aMCI), with memory deficits and a high probability of developing AD, also exhibit histopathological and hypometabolic changes in the temporal and posterior parietal cortex. While patients with AD were significantly impaired in both a ground (i.e. regular horizontal) and overhead-view planes, the aMCI group showed diminished level-2 VPT abilities compared to controls only in the ground plane [42]. Their impairment supports the involvement of the posterior temporal-parietal brain regions in the level-2 VPT tasks and stresses the importance of further examining the role of the viewing plane in level-2 VPT.

Right cerebral lesions have often been linked with topographical disorientation. More specifically, the right posterior parietal, but not frontal, lesions were documented to negatively affect spatial updating of an egocentric direction [43,44]. Recently, 63 patients with the right hemisphere lesions (overlapping mostly in the insular cortex, putamen, and around the superior temporal sulcus) were tested on the level-2 VPT task. The patients showed significantly poorer performance than healthy controls and were prone to solve the task from their own perspective [45]. However, it is difficult to declare the right hemisphere dominance in level-2 VPT in the absence of direct comparison with the left hemisphere lesions.

Some support for the causal role of the right hemisphere, namely, the rTPJ, in level-2 VPT also came from studies applying transcranial stimulation, which enables to temporarily potentiate or inhibit specific
brain areas. As discussed above, Wang et al. [5] detected by MEG the rTPJ as the core brain area for transforming self-perspective into another’s location and orientation in space in an embodied manner. Then, using dual pulse TMS, they inhibited the rTPJ activity, which resulted in weakened embodied processing [5]. In particular, compared to a no stimulation condition, participants’ body posture congruence/incongruence had a lesser effect on the performance [5]. In another study, the rTPJ activity was potentiated by applying anodal high-definition transcranial direct current (HD-tDCS) over this brain area [18]. Here, the body posture congruence/incongruence with an avatar’s direction had a significantly stronger effect on the performance than during a no stimulation condition [18]. In one more study, TMS of the rTPJ at the theta frequency improved level-2 VPT task performance [20]. Thus, the rTPJ seems to be associated with the embodied processes underlying level-2 VPT. These transcranial stimulation studies focused only on the right and not the left hemisphere, based on the MEG time-frequency and connectivity results discussed above. Another study [46] used a different experimental design employing the widely used Director task [47], with versions measuring either perspective tracking or level-2 VPT. In this study, participants were required to move objects to the left/right from a director’s perspective; therefore, it was measuring level-2 VPT [46]. However, in contrast to the above-mentioned MEG and transcranial stimulation studies, the analysis in the study of Santiesteban et al. [46] was focused not on the main effects of increasing angular disparity and body posture congruence, but on the contrast between perspective-taking and no perspective-taking trial types. By estimating the effect of anodal stimulation of both hemispheres on level-2 VPT, Santiesteban et al. [46] demonstrated that the stimulation of the right, as well as the left, TPJ had an effect on level-2 VPT abilities. Therefore, it can be inferred that the rTPJ is predominantly important for embodied processes underlying level-2 VPT, whereas the left TPJ is involved in differentiating another’s from self-perspective. Supporting this view, after measuring the theta power in both hemispheres in response to level-2 VPT, Seymour et al. [6] in their MEG experiment detected that the observed theta power was stronger in the rTPJ, compared to the left TPJ, when the angular disparity between self and avatar’s perspectives was the largest. In addition, it has also been reported that the left TPJ exhibited widely distributed functional connectivity with the superior temporal sulcus and anterior temporal lobe [48]. These brain areas are associated with processing semantic knowledge about the self; therefore, they might be playing a role in the self/other distinction processes. On the other hand, the rTPJ was connected to the brain areas linked with body representations, namely with the insula and SMA [48].

The role of the mPFC has often been emphasized in level-2 VPT. Interestingly, stimulation of the dorsomedial PFC (dmPFC) with anodal direct current did not show any effect in the same study that demonstrated the role of the rTPJ during level-2 VPT [18]. Importantly, in this experiment, only the third-person level-2 VPT abilities were tested, as participants were not asked to represent their own perspective [18]. On the other hand, the role of the dmPFC was demonstrated in their previous study involving self-visuospatial perspective judgments [49]. The study consisted of a series of experiments. In one of them, participants were explicitly instructed to make judgments depending on a trial-type either from their own, i.e., self-perspective or from another perspective. During the self-perspective judgments, Martin et al. [49] compared trials with a similar response, i.e., congruent, from one’s own and another’s perspective, and trials with incongruent response. They observed that anodal stimulation of the dmPFC resulted in a greater difference in RTs between these two types of trials, with shorter RT in the congruent and longer RT in the incongruent trials [49]. In other words, the stimulation of the dmPFC increased the influence of another’s visuospatial perspective during self-visuospatial perspective judgments and decreased egocentric bias. In another experiment of the same study, participants were asked to represent only their own perspective in all trials. Thus, they were not explicitly asked to switch between their own and another’s perspective. Interestingly, in this type of experiment, anodal stimulation of the dmPFC showed no effect on the performance [49]. Therefore, the activation of the dmPFC seems to predominantly influence self-visuospatial perspective judgments, specifically those, where discrimination between one’s own and another’s perspective is explicitly required. According to the authors’ interpretation, stimulation of the dmPFC causes an increase in the integration of information about another’s perspective into our own [49].

Overall, based on the results from the neurodevelopmental/lesion and transcranial stimulation studies, the posterior temporoparietal brain regions, especially the TPJ, seem to be essential for transforming one’s own visuospatial perspective into another’s location and orientation in space. Although many studies reported the significance of the right hemisphere, some studies showed that the left hemisphere is also involved in level-2 VPT [46,50]. However, the right hemisphere TPJ can be specifically important for the embodiment processes underlying level-2 VPT.

The mPFC seems to be more involved in self-perspective judgments, with the dorsal and ventral parts of the mPFC apparently playing distinct roles in this process. In particular, the activation of the dmPFC might have an effect on those self-perspective judgments, where the distinction between self and others’ perspective is explicitly required, and one needs to switch between one’s own and others’ perspectives. Here, the overactivation of the dmPFC seems to increase the merging of others’ viewpoint-related information with one’s own and interferes with egocentric judgments. In contrast, as demonstrated by fMRI experiments, the vmPFC can be crucial for representing the egocentric view from a self-visuospatial perspective [23,24] (Fig. 2).

4. Comparison of level-2 VPT with object mental rotation from behavioral and neural perspectives

4.1. Behavioral perspective

Similarly to level-2 VPT, object mental rotation (oMR) is also performed by the transformation of visuospatial mental images. However, in contrast to level-2 VPT, oMR implies imagining to rotate the object-based reference frame [2].

One of the firsts who started studying oMR were Shepard and Metzler [51]. In their original experiment, two differently oriented 3D objects were presented simultaneously. Participants should have answered as fast as possible whether the two stimuli were identical or a mirror-reversed version of each other. Results showed that the RTs became longer with the increase in rotational angular disparity between the two objects, indicating that the task was performed by forming a mental image of the stimulus and then mentally rotating it [51].

Behavioral results comparing level-2 VPT task performance with oMR demonstrated that in a ground plane participants were faster in level-2 VPT tasks compared to oMR [4,25,52,53]. However, when participants performed both tasks in other planes, no significant difference in RTs was observed [52,54]. This might have two explanations: First, compared to other planes, level-2 VPT in the ground plane can be easier to perform, as performing visuospatial perspective transformations in the ground plane is more natural and faster for us as ground-dwelling species [2,4]. Second, oMR in the ground plane can be more difficult to perform compared to other planes. In support of the first explanation, participants were more error-prone during level-2 VPT in other planes compared to level-2 VPT in the ground plane [52]. Also, a comparison of level-2 VPT abilities in various planes conducted in another experiment demonstrated that participants were faster in the ground plane compared to other planes [55].

Both level-2 VPT and oMR are more difficult to perform with larger angular disparities [4,53]. However, it was observed that during oMR,
angular disparity increased RTs linearly, whereas during level-2 VPT, a significant increase was associated only with angles larger than 60° and 90° [4, 53]. During oMR the angular disparity is measured between a reference and a target object; during level-2 VPT, it is measured between a participant’s egocentric reference frame and a target perspective [2].

Another discrepancy between oMR and level-2 VPT seems to be related to executive functions. From the three basic EF, oMR seems to be related to working memory [56]. One study showed a relation between oMR and performance on an EF task, measuring inhibition and working memory, but not between oMR and another EF task that measured inhibition and switching [56]. Another study demonstrated a significant correlation between inhibition and level-2 VPT performance, but not between inhibition and oMR [57]. Even though, level-2 VPT task performance did not show a significant correlation with the test scores measuring switching EF [58], the effect of the switch cost, that is the measure of additional cognitive demand required to switch between tasks [59], was demonstrated when participants were required to switch between their own and another’s visuospatial perspective [60, 58]. Also, a significant correlation was detected between level-2 VPT and working memory EF task performances [58]. Thus, oMR and level-2 VPT employ EF differently. While various studies show a correlation of level-2 VPT with all three basic executive functions, according to the reviewed literature, oMR seems to be related to working memory EF.

Apparently, aging has a distinct influence on oMR and level-2 VPT. One study examined the oMR and level-2 VPT abilities in young (age range: 18–29), middle-aged (30–59), and elderly (60 and older) adults [61]. The middle-aged group showed a significantly decreased number of correct responses compared to young participants in both tasks. However, the elderly were significantly more impaired on a level-2 VPT than on an oMR task compared to the middle-aged group. This result suggests that the ability to mentally transform one’s own visuospatial perspective declines with age more than the ability to mentally rotate objects [61].

One more distinction between level-2 VPT and oMR seems to be the embodiment. Participants’ body posture had an effect only on level-2 VPT, but not on oMR [4]. Thus, it was argued that, compared with level-2 VPT, oMR does not imply transformations of the whole body schema [4]. Overall, the behavioral results show that oMR and level-2 VPT are underpinned by different cognitive processes.

4.2. Neural perspective

Attempts to localize brain activity underpinning oMR have pointed to the recruitment of motor and visuospatial brain areas. An fMRI study comparing oMR directly with level-2 VPT showed higher activity for oMR in the brain regions responsible for visuospatial reasoning (the bilateral IPL and SPL), in the brain area associated with EF, including working memory, namely, the lateral PFC, and also in the left primary motor cortex [25].

Importantly, the SPL and the left IPL, as well as the IPFC are also involved in level-2 VPT tasks, but they seem to be more actively engaged during oMR. The stronger activation in the brain area associated with working memory EF is also in line with the behavioral results showing that oMR is more difficult to perform than level-2 VPT. The reversed effect, level-2 VPT vs. oMR, showed increased BOLD signal in the dorsal parts of the mPFC corresponding to the superior frontal gyrus (SFG), in the insula, SMA, and in a region of the MOG corresponding to the EBA [25] (Fig. 3).

Mental rotation of different types of stimuli apparently recruits distinct brain structures. The right angular gyrus (part of the TPJ), insula, SFG, and cingulate cortex were activated during mental rotation of bodily-related stimuli (e.g., images of hands or full bodies) compared to non-bodily-related stimuli [62]. Importantly, it has been demonstrated that when in the Shepard and Metzler [51] paradigm the whole human-body-shaped stimuli were used instead of the 3D objects, the underlying cognitive processes of the task performance corresponded to

![Fig. 3. Brain areas involved in level-2 VPT and oMR. Brain areas involved in oMR are given in orange (oMR). Brain areas involved in level-2 VPT are shown in blue (level-2 VPT). Brain areas involved in both oMR and level-2 VPT are depicted in purple (Both); however, when directly compared, bilateral IPFC, SPL, and left IPL were more active during oMR.](image-url)
level-2 VPT, not oMR [21]. In other words, mental rotation of such stimuli category as full human-body requires the mental transformation of the egocentric reference frame that occurs during level-2 VPT [4,21]. Thus, the involvement of the part of the TPJ reported in response to mental rotation of full bodies most likely depicts the neural processes underpinning level-2 VPT. Mental rotation of the hand images induced activation in the left precentral gyrus (incl. SMA) compared to the rotation of the whole body [62].

EEG studies have also revealed some aspects of neural processes underlying mental rotation. Mental rotation seems to have a more or less robust ERP signature, the “rotation-related negativity”. It has been observed over the parietal electrodes at ~350–800 ms post-stimulus in response to mental rotation of various types of stimuli [63–67]. Besides, the response was reported to be modulated by angular disparity: greater rotation angle induced stronger ERP signal [67–70].

Similar to level-2 VPT, the right lateralization of oMR is also an equivocal topic. Although some studies have claimed a dominant engagement of the right hemisphere during oMR [71], more recent studies have not confirmed it. ERP associated with oMR was found in both hemispheres [70]. However, in the right hemisphere, parietal negativity in response to oMR started at 400 ms after stimulus onset and lasted till 550 ms. In contrast, in the left hemisphere, the same effect started at 610 ms after stimulus onset [70]. These results indicated that both hemispheres were engaged in oMR, but the main difference was in the timing. In particular, the right hemisphere was the first to engage and faster to process. Furthermore, the source modeling showed that the cortical areas involved in oMR processing did not seem to differ in size between the hemispheres [70].

According to the brain-imaging results reported in this and previous sections, oMR and level-2 VPT both recruit visuospatial brain areas like the SPL and IPL, and also the IFPC that is involved in EF. However, the direct comparison demonstrated that these brain areas are more active during oMR. This might correspond to the behavioral results showing that, in general, oMR is more difficult to perform than level-2 VPT. In contrast to oMR, mental rotation of human-body parts has more areas overlapping with level-2 VPT. These additional overlapping areas are the ones associated with self-consciousness and body representations, such as the SFG, insula, cingulate cortex, and SMA. The right hemisphere dominance has often been discussed in response to level-2 VPT, as well as during oMR. However, recent findings have shown that the left hemisphere also exhibits a considerable role in both processes. Overall, level-2 VPT seems to engage functionally more diverse brain areas than oMR. Therefore, level-2 VPT apparently is a complex cognitive process, encompassing not only visuospatial and EF, but also social and embodiment processes.

5. Comparison of level-2 VPT with theory of mind from behavioral and neural perspectives

5.1. Behavioral perspective

While level-2 VPT implies judgments on one’s own or others’ imagined visuospatial perspectives, theory of mind (ToM) incorporates reasoning on one’s own and others’ mental states and attributing emotions, intentions, or beliefs to oneself and others [72,73]. The “Puppet play” paradigm developed by Wimmer and Perner [74], later named as “Sally-Anne test” [75], is one of the earliest tasks measuring ToM abilities in children. Their results indicated that at the age of 4–5, more than half of children could correctly infer that others can hold different beliefs about a location of a doll than they do themselves [74]. However, some studies suggest that in the experiments where children were not required to process verbal information, toddlers already around 2–3 years of age could recognize false beliefs and intentions of others [76–78].

ToM abilities, like inferring others’ mental states, also called mentalizing, have been considered to develop earlier in humans than level-2 VPT [79]. However, for healthy adult participants, level-2 VPT seems to be easier, as they were significantly more accurate in a level-2 VPT task than during mentalizing [80]. In contrast, adult participants with Asperger syndrome showed longer RTs and more errors during mentalizing tasks, but not during level-2 VPT tasks, compared to controls [81]. These results point out that the cognitive processes underlying level-2 VPT and ToM could be partially distinct.

However, according to some authors, there is a strong correlation between level-2 VPT and ToM, and both of them rely on the process of switching between one’s own and others’ perspectives (visuospatial in one case and mental in another) [73,82]. When studying level-2 VPT abilities in children with low-functioning autism spectrum disorder (ASD), Hamilton et al. [82] reported that level-2 VPT abilities were directly correlated with ToM scores [82]. In these children with ASD, the level-2 VPT abilities were significantly more impaired than oMR abilities compared to typically developing children (TD) [82]. In addition, the regression analysis demonstrated that in the TD children, the ToM scores were able to significantly predict performance on level-2 VPT task, whereas the oMR performance contributed only marginally [82]. Another study using a similar experimental paradigm to that used by Hamilton et al. [82] did not show a correlation between level-2 VPT and ToM [83]. Instead, Pearson et al. [83] demonstrated that in the ASD group, but not in the TD group, a level-2 VPT performance was predicted by oMR scores. However, one more study also measuring level-2 VPT abilities in ASD and TD children did not show that oMR scores could significantly predict the level-2 VPT abilities in none of these groups [84]. The reason for contrasting results regarding the correlation between level-2 VPT and ToM scores reported by Hamilton et al. [82] and Pearson et al. [83] might have two reasons. First, in contrast to Hamilton et al. [82], Pearson et al. [83] measured ToM scores in the ASD group only, but not in the TD group. The second reason could be a distinct severity of ASD in the group of children participating in these two studies. The children with ASD participating in the experiment of Hamilton et al. [82] were diagnosed with low functioning ASD and were significantly more impaired on the ToM scores than the TD children. In comparison, the children with ASD participating in the study of Pearson et al. [83] showed intact performance on the same battery of the ToM tests.

Furthermore, a recent review of the assessment of ToM abilities outlined two defining criteria that any task measuring ToM abilities should meet [73]. First, a ToM task should require more than just attributing a mental state to others, but it should also necessitate that participants distinguish between their own and others’ mental states, or distinguish between their own actual and imagined mental states. Second, successful performance on a ToM task should not depend on just lower-level processes like attention reorientation or associative learning. According to the authors, level-2 VPT tasks can be used to assess ToM abilities since they meet both of these defining criteria [73].

Supporting the assumption that related processes underlie level-2 VPT and ToM, both of them seem to employ similar EF. It is well documented that the ToM process to understand mental states of oneself and others requires working memory, inhibition, and switching [85–89]. At the same time, as discussed above, level-2 VPT also requires working memory function [58] and inhibition [57,58]. Also, the switch cost was observed when participants were switching between self and another’s visuospatial perspective [60,58]. It has been suggested that aging affects ToM abilities by making older adults more prone to egocentric judgments; therefore, it is more difficult for them to represent the mental states of others [90,91]. As it seems, aging has a similar influence on level-2 VPT too since older adults showed significantly decreased level-2 VPT abilities and made egocentric errors while representing another’s visuospatial perspective [61]. A recent study has also demonstrated that with older age, it was more difficult for participants to switch to another’s visuospatial perspective [58].

Another intriguing question is whether level-2 VPT can be related to
all types of ToM, or only to certain types. ToM has been differentiated into cognitive and affective ToM. Cognitive ToM deals with a cognitive understanding of the difference between one’s own and others’ beliefs and intentions. It has been demonstrated that level-2 VPT task performance was in correlation with cognitive ToM [82]. Affective ToM, in addition to cognitive understanding, incorporates empathizing with others’ emotional states [92–94]. It has also been related to embodied processes [95]. As level-2 VPT is an embodied process and several studies have also demonstrated its correlation with affective ToM and level-2 VPT. However, a recent study did not find a significant correlation between affective ToM and level-2 VPT [58]. But, in this study affective ToM was measured by the Reading the mind in the eyes test [98], which, although frequently used in assessing affective ToM, according to some authors, should not be identified as measuring ToM abilities since it can be performed by a lower-level process as it is perceptual emotion recognition [73,99].

In contrast to level-2 VPT, the interaction between perspective-tracking and ToM is doubted [5,6,73,100]. Interestingly, the perspective-tracking ability was reported to be intact in children on the autistic spectrum [100]. Also, it was demonstrated that perspective-tracking is related to EF differently than ToM [101,102]. Although the switch cost [60] as well as the working memory load was observed during perspective-tracking [102]; the main cognitive process underlying perspective-tracking – calculation of another’s line of sight – does not seem to employ inhibition [101]. On the other hand, aging seems to have a comparable pattern of influence on perspective-tracking and ToM. It has been suggested that during a perspective-tracking task, older adults prioritized self over a third-person perspective more than younger adults [103]. Another study also showed that compared to younger adults, it was more difficult for older adults to switch to tracking another’s perspective [58]. However, according to the criteria of ToM tasks outlined by Quesque and Rossetti [73], perspective-tracking tasks do not measure ToM abilities as they can be performed by the lower-level automatic process as it is the calculation of another’s line of sight [73].

5.2. Neural perspective

Neuroimaging studies suggested that the mPFC, TPJ, precuneus, and the posterior cingulate cortex are the most important brain regions of the ToM network [104,105]. In addition, ToM abilities seem to be maintained in cooperation with other brain areas that among different functions are involved in EF and spatial processes [48].

Interestingly, it has been indicated that ToM and level-2 VPT share overlapping neural circuits [50]. The meta-analysis of brain regions reported in both level-2 VPT and ToM tasks has revealed that the left TPJ, as well as EBA, are engaged in both tasks [50]. Besides, other brain areas observed in response to level-2 VPT and in the ToM judgments include the left IPL, the premotor area in the bilateral MFG, and the somatosensory area - the SMG [80], as well as the IFG [106,107], dmPFC [6,34], and executive control brain areas, such as the bilateral ACC and IPFC [6,108,109]. The insula was also reported to be involved in a certain type of ToM task [110] and also in level-2 VPT [25]. After insular lesions, the performance was impaired during understanding the emotions of others, whereas performance was intact in the task, which required understanding others’ beliefs [110]. According to the meta-analysis, the brain area reported in response to ToM tasks, but not during level-2 VPT, is the anterior temporal lobe [50]. It has been suggested that the anterior parts of the temporal lobe are involved in those ToM tasks that require an understanding of communicative intents (gestures, facial expressions, verbalizations, or written messages). Importantly, their activity was not determined by narrative comprehension but was correlated with social cognition, specifically with processing social conceptual knowledge [111] (Fig. 4).

Fig. 4. Brain areas involved in level-2 VPT and ToM. Brain areas involved in ToM are given in green (ToM). Brain areas involved in level-2 VPT are given in blue (level-2 VPT). Brain areas involved in ToM judgments and also during level-2 VPT are depicted in red (Both).
As discussed above, the right TPJ plays a crucial role in transferring one’s own visuospatial viewpoint into another’s perspective in an embodied manner [5,6,18,20]. At the same time, the involvement of the bilateral TPJ has been reported to modulate mentalizing abilities during ToM tasks [35,48,105,112]. Similarly to level-2 VPT [6], the phase-coherence in the theta band was observed among the TPJ and mPFC in response to mentalizing [113].

The observed functional dichotomy between the dorsal and ventral parts of the mPFC in response to visuospatial perspective judgments seems to be also relevant for mental perspective judgments. The dmPFC inhibition by TMS resulted in stronger egocentric bias and reduced influence of another’s mental perspective during ToM judgments [104]. Similarly, anodal stimulation of the dmPFC weakened egocentric bias during self-visuospatial perspective judgments and increased the influence of another’s visuospatial perspective [49]. On the other hand, the vmPFC has been reported to be involved in representing the egocentric view during self-mental perspective judgments in ToM tasks [34,104], as well as during self-visuospatial perspective judgments [7,23,24].

Overall, these findings show that there is an interaction in the neural underpinnings of level-2 VPT and ToM. The involvement of brain regions such as precuneus and IPL, which are usually attributed to visuospatial perspective transformations, suggests that these areas might encode both mental and visuospatial types of perspective transformations [50,80]. The recruitment of the premotor, somatosensory area, and EBA might indicate that both ToM and level-2 VPT imply some type of motor-imagery, such as visualizing one’s own or others’ movements [50,80]. The EF brain areas, the ACC and IFG [109], as well as the IFG [114] were suggested to be involved in inhibiting and overcoming egocentric biases. These processes are important aspects of taking someone else’s visual or mental perspectives. The role of the insula in certain types of ToM tasks can be emotion recognition in others [110]. Besides, during level-2 VPT, the insula might play a role in the embodiment processes as this brain area has also been associated with processing information related to one’s own body [115]. The dmPFC seems to play a role in the integration of information about others’ mental or visuospatial perspectives into our own, while the TPJ (apparently bilaterally) might be the key hub of transforming our own mental or visuospatial representations into the ones of others [50,116]. As for the vmPFC and PCC, they seem to modulate the representation of our own visuospatial and mental states [7,23,34].

6. Conclusion

Level-2 VPT is characterized by mentally transforming one’s own visuospatial viewpoint into another’s location and orientation in space in an embodied manner. It engages various cognitive processes and functionally distinct brain regions, related to oMR on the one hand and to the ToM on the other. Neuroimaging studies revealed that brain regions traditionally involved in visuospatial reasoning (SPL, IPL) and in executive functions, including working memory (IFPC), are also engaged during level-2 VPT and oMR. However, they seem to be more active during oMR, which corresponds to the behavioral results indicating that in healthy conditions and in the ground plane, oMR is more difficult to perform than level-2 VPT. On the other hand, the brain regions associated with social cognition are commonly observed during level-2 VPT, as well as in the ToM tasks. The crucial structure for transforming our own visuospatial or mental perspective and representing another’s one seems to be the TPJ. The process of level-2 VPT was reported to be modulated by the theta band power increase in the network of the TPJ and EF brain areas (IFPC, ACC), as well as by the theta band synchronization between the TPJ and other ToM related brain areas, such as the mPFC and PCC.

The embodiment of another’s viewpoint has been suggested to be the important component of level-2 VPT. The neural signature of embodied processes underlying level-2 VPT can be the theta band phase-coupling detected between the right TPJ and brain areas associated with the body schema representations (SMG, PPC, SMA), as well as increased BOLD activity in the insula, SMA, and in the MOG corresponding to the EBA.

level-2 VPT is differentiated from perspective-tracking that does not seem to require a mental transformation of one’s own visuospatial perspective and the body schema. Instead, perspective-tracking, apparently, relies on calculating another’s line of sight. Participants generally perform better during perspective-tracking compared to level-2 VPT. Perspective-tracking activates some of the brain areas that are involved during level-2 VPT. Specifically, it is reported to recruit visuospatial (precuneus, IPL), as well as middle and inferior frontal brain areas. In contrast to the level-2 VPT, increased theta power or phase-coupling was not observed in the brain areas associated with ToM or the body schema representations. Instead, increased theta power was detected in the ventral occipital and parietal cortices and in the frontal eye fields.

When asked to imagine how a scene looks from another perspective, an observer needs to make a distinction between one’s own actual visuospatial perspective and imagined perspective. Therefore, differentiating and representing a self-visuospatial perspective is often studied together with level-2 VPT. Representing the self-visuospatial perspective induces activations in the brain areas associated with self-consciousness (vmPFC, PCC, insula). Being conscious of one’s own perceptions, including visuospatial, is the constituent part of self-consciousness. In this sense, representing the self-visuospatial perspective has been related to self-consciousness.

Overall, level-2 VPT is a diverse cognitive phenomenon implying various neural structures. The brain areas reported to be involved in level-2 VPT are following: The ACC, dmPFC, MOG, IFG, insula, IPL, IFPC, MFG, PCC, precuneus, SMA, SMG, SPL, and TPJ. Therefore, the neurocognitive model underpinning level-2 VPT can be understood as a network combining the visuospatial and executive control processes with the body schema representations and social cognition processes. The social cognition aspects of level-2 VPT seem to have more neurocognitive load than the visuospatial processes, as ToM abilities predict level-2 VPT performance better than oMR, at least in those tasks where a human-like shape is presented as a stimulus.

7. Future directions

The TPJ seems to be the brain structure which activity was most often reported as distinguishing level-2 VPT from oMR and linking level-2 VPT with ToM. However, it was also shown that in an experiment focusing on changes in visuospatial perspective, the TPJ activity only increased when human-body parts, presented from a third-person perspective, were well recognizable, compared to trials with blurred images [117]. Another experiment that measured perspective-tracking abilities with and without the presentation of a human-like avatar reported that the TPJ was involved in those conditions, in which an avatar was presented compared to an arrow [118], but see [38]. Considering these two results, it seems, at first sight, that the involvement of the TPJ in visuospatial perspective tasks could be influenced by detecting another subject as an active agent, i.e., by the presentation of human-body-related stimuli in object-directed action. Even though, it should also be mentioned that none of these studies measured the level-2 VPT abilities per se. The first study involved viewing of object-directed actions either from the first-person or from the third-person perspective [117], whereas the second study measured perspective-tracking [118]. Besides, in the second study, although the ROI analysis showed an interaction of perspective by avatar presentation in the TPJ, the involvement of the TPJ during perspective-tracking in the whole-brain analysis is questionable [118]. However, to better understand the role of the TPJ activation in level-2 VPT, further experiments are needed comparing conditions involving human avatar with conditions involving one’s own visuospatial perspective transfer to another abstract point in space without the presence of a human-like avatar.

Behavioral experiments demonstrated that viewing planes have a considerable effect on level-2 VPT, such that participants generally
perform better in the ground plane than in other planes [2,52,55]. In addition, level-2 VPT in the ground plane, but not in the overhead-view plane, can be used in early AD diagnostics [42]. However, the neural background of the influence of the viewing planes on level-2 VPT is not well documented. Therefore, it is important to better understand the role of the viewing planes in level-2 VPT from a neuroscientific perspective.

Several issues might cause confusion in the level-2 VPT, perspective-tracking, and mental rotation research. Properly distinguishing perspective-tracking from level-2 VPT can avoid confusion in the associated brain activation. Some neuroimaging studies measure only perspective-tracking abilities but refer to the general concept of VPT (without differentiating level-1, i.e., perspective-tracking, from level-2 VPT) and do not demonstrate activation of such a crucial ToM network brain area as the TPJ [22]. At the same time, the TPJ activation was robust in other studies directly measuring level-2 VPT [6,23,24]. Another confusion has been about the mental rotation of whole human-body images. As it has been suggested, mental rotation of this stimulus category can be distinguished from the mental rotation of objects (oMR) as only the former one requires a mental transformation of the egocentric reference frame similar to that which occurs during level-2 VPT [4,21]. Hence, it might be useful to make a clearer distinction between these two types of mental rotation.

The roles of the two crucial brain areas involved in self and another’s visuospatial perspective representation process ask for better clarification. In particular, it is still questionable whether the major role of the TPJ in level-2 VPT is distinguishing between any other and self-perspective (irrespective of another agent’s presence), detecting another as an agent and distinguishing between another agent’s and self-perspective, suppressing self-perspective and giving priority to another perspective, or if it is the embodiment during transforming self-perspective into another’s location and orientation in space. Besides, the dorsal and ventral parts of the mPFC have been reported to play predominant roles in self-visuospatial perspective judgments compared to third-person perspective judgments. However, they might play distinct roles in this process. Namely, the involvement of the dmPFC seems to cause the increase in the integration of another’s perspective into our own and elimination of the egocentric bias when judgments are made from a self-perspective [49], whereas the vmPFC seems to be specialized in representing the egocentric view from a self-perspective [23], but this needs more testing and clearer classification. In particular, the role of the dmPFC was demonstrated only in those conditions where participants were explicitly asked to make a distinction and to switch between their own and another’s perspective [49]. Hence, it will be worth examining whether the vmPFC, in contrast to the dmPFC, would be more actively engaged in experimental designs where one is asked to represent only a self-perspective and no switch between one’s own and another’s perspective is required. Also, the role of the dmPFC during the self-perspective judgments was demonstrated by applying the anodal stimulation to it, which interfered with the self-perspective judgments. Thus, it is questionable if the dmPFC is involved in self-visuospatial perspective judgments in regular conditions, without external stimulation. Other than that, if the stimulation of the dmPFC causes a reduction of an egocentric bias and increases the influence of another’s perspective during self-perspective judgments, the question arises why this stimulation does not significantly improve also the third-person perspective judgments [18,49]. Considering the reported involvement of the dorsal parts of the mPFC during representing the third-person perspective [6], the engagement of the dmPFC during third-person perspective judgments needs more testing. Therefore, there is a need to have a close look at the brain structures active during representing self and another’s visuospatial perspective with technology enabling high temporal and spatial resolution. Intracranial EEG could be one of the solutions here by providing information about the selective involvement of neuronal populations with the millimeter scale spatial precision and with the millisecond scale timing of their engagement [119].

Multivariate pattern analysis (MVPA) unites diverse methods for neuroimaging data analysis. It can be applied to EEG/MEG [120] as well as to fMRI [121] data. MVPA is widely used in neural decoding and provides the possibility to predict which task a person is engaged in based on patterns of brain activation. Also, since the common element of the MVPA methods is that they consider the relationships between multiple variables (voxels in fMRI, channels in EEG/MEG); therefore, they make it possible to spot shared or independent neuronal activity patterns within certain brain areas [122]. In an fMRI experiment, MVPA has been successfully used to find the brain areas selectively encoding the actions observed from a first-person, but not from a third-person, and other areas encoding actions regardless of perspective [123]. In another fMRI MVPA study, authors managed to detect certain areas within the insula that encoded an aversive state in a general manner, irrespective of whether it was experienced by oneself or by another. At the same time, they also detected other areas within the insula that encoded the aversive state in a modality-specific manner [122]. Considering these examples, MVPA could be used to estimate self/other visuospatial perspective-specific neuronal representation within certain brain areas.

Functional connectivity studies can also provide valuable information about the engagement of different brain areas in level-2 VPT. As discussed above, oscillation-based functional connectivity was observed by MEG in a level-2 VPT task among the rTPJ, IFPC, and ACC in the theta-band [6]. However, this study did not incorporate responding from the self-perspective; therefore, it did not give evidence on the functional connectivity change between representing self and another’s visuospatial perspective. Recently, one fMRI functional connectivity study, measuring ToM abilities, demonstrated distinct functional connectivity when participants were representing others’ thoughts compared to representing their own [124]. It would be of interest to observe also in a level-2 VPT task how representing self and another’s visuospatial perspective would alter brain wiring and if results could be comparable to the ones obtained from ToM studies. Another fMRI study detected that healthy subjects, patients with ASD, and patients with schizophrenia exhibit distinct functional connectivity in response to a perspective-tracking task [125]. It would be interesting to measure the functional connectivity change in response to representing a self-visuospatial perspective and level-2 VPT in these groups of patients as well.

Finally, VPT research can vastly benefit from elaborating more naturalistic paradigms. There are two ways of bringing neuroscientific studies to more real-world-like situations, and therefore increase their ecological validity [126]. One way is to leave a lab and switch to a fully naturalistic real-world environment. This enables direct testing of how much the lab-generated models work in real-life situations. The second way is to incorporate more naturalistic stimuli in the lab research. This approach could provide more precise accounts on natural perception and action and at the same time maintain a certain level of stimulus control [126]. With advancements in technology, it becomes possible to study brain activity outside the lab. Portable EEG heads have already been successfully used to record brain-to-brain synchrony among students during their natural social interaction [127,128]. On the other hand, among various ways of simulating real-world-like situations in the lab, one way is to create experiments in an immersive virtual reality (VR) environment. In this setup, participants tend to experience a sense of presence, i.e., even though they know that the VR environment is not real, they still experience the presented stimuli as much more real than those presented on a less immersive 2D screen, and they feel “being there” in VR [129–131]. Similar VR has been already actively used in studying level-2 VPT, ToM, and embodiment [132,133,134,138]. Using immersive VR in combination with brain-imaging technologies could be beneficial for studying brain substrates of representing one’s own visuospatial perspective, as well as making judgments on another’s visuospatial perspective in more real-world-like scenarios.
about the minds of others, in: J.B. BENSON (Ed.), Advances in Child Development and Behavior, JA1, 2017, pp. 185–226, https://doi.org/10.1016/bs.acdb.2016.10.005.

[79] H. Moll, D. Kadiapasagou, The primacy of social over visual perspective-taking, Front. Hum. Neurosci. 7 (2013), https://doi.org/10.3389/fnhum.2013.00558.

[80] N. David, C. Aumann, N.S. Santos, B.H. Bewernick, S.B. Eickhoff, A. Newen, N.J. Schlyter, G.R. Fink, K. Vogeley, Developmental involvement of the posterior temporal cortex in mentalizing but not perspective-taking, Soc. Cogn. Affect. Neurosci. 3 (2008) 279–289, https://doi.org/10.1093/scan/nsm023.

[81] N. David, C. Aumann, B.H. Bewernick, N.S. Santos, F.-G. Leibarth, K. Vogeley, Investigations of mentalizing and visuospatial perspective taking for self and other in Asperger syndrome, J. Autism Dev. Disord. 40 (2010) 290–299, https://doi.org/10.1007/s10803-009-0867-4.

[82] H. DeC. Hamilton, R. B. Frith, Visual perspective taking impairment in children with autistic spectrum disorder, Cognition 113 (2009) 37–44, https://doi.org/10.1016/j.cognition.2009.07.007.

[83] A. Pearson, L. Marsh, D. Roper, A. Hamilton, Cognitive mechanisms underlying visual perspective taking in typical and ASC children, Autism Res. 9 (2016) 121–130, https://doi.org/10.1002/aur.1501.

[84] R. Cardillo, C. Erbi, L.C. Mammarella, Spatial perspective-taking in children with autism spectrum disorders: the predictive role of visuospatial and motor abilities, Front. Hum. Neurosci. 14 (2020), https://doi.org/10.3389/fnacn.2020.00208.

[85] B. Muter, M.B. Alcorn, M. Welsh, Theory of mind and executive function: working-memory capacity and inhibition control as predictors of false-belief task performance, Percept. Mot. Skills 102 (2006) 819–835, https://doi.org/10.2466/pms.102.4.819-835.

[86] M.A. Sabbagh, F. Xu, S.M. Carlson, L.J. Moses, K. Lee, The development of executive functioning and theory of mind, Psychol. Sci. 17 (2006) 74–81, https://doi.org/10.1111/j.1467-9280.2005.00693.x.

[87] R. Bull, L.H. Phillips, C. Corballis, The role of control functions in mentalizing: dual-task studies of Theory of Mind and executive function, Cognition 107 (2008) 663–672, https://doi.org/10.1016/j.cognition.2007.07.015.

[88] E.E.F. Bradford, I. Jentsch, J.-C. Gomez, From self to social cognition: theory of mind mechanisms and cortical reconfiguration and interference control, Psychol. Bull. 125 (2016) 2119–2146, https://doi.org/10.1037/abn0000182.

[89] M. Wade, H. Prime, J.M. Jenkins, K.O. Yeates, T. Williams, K. Lee, On the relation between theory of mind and executive functioning: a developmental cognitive neuroscience perspective, Cogn. Brain Res. 25 (2008) 206–211, https://doi.org/10.1016/j.cogbrainres.2008.02.004.

[90] L. Broaders, Bro, A. Neuroethological framework for the representation of minds, J. Cogn. Neurosci. 4 (1992) 107–118, https://doi.org/10.1162/jocn.1992.4.1.107.

[91] S.G. Shamay-Toyosh, S. Shur, L. Barcai-Goodman, S. Medlovich, H. Harari, Y. Levkovitz, Dissociation of cognitive from affective components of theory of mind in schizophrenia, Psychiatry Res. 149 (2007) 11–23, https://doi.org/10.1016/j.psychres.2007.04.001.

[92] C.L. Sebastian, N.M.G. Fontaine, G. Bird, S.-J. Blakemore, S.A.D. Brito, E.J. Mcprucy, E. Viding, Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults, Soc. Cogn. Affect. Neurosci. 7 (2012) 53–63, https://doi.org/10.1093/scan/nss023.

[93] K. Jose, A. Fiedl, M. Leinfelder, The interaction between embodiment and empathy in facial expression recognition, Soc. Cogn. Affect. Neurosci. 13 (2018) 203–215, https://doi.org/10.1093/scan/nos005.

[94] P.C. Gunthom, M. Flynn, C.J. Edmonds, M.R. Gardner, Empathic and non-empathic routes to visuospatial perspective-taking, Conscious. Cogn. 21 (2012) 494–500, https://doi.org/10.1016/j.concog.2011.12.004.

[95] V. Sulpizio, G. Committetti, E. Metta, S. Lambrey, A. Berthoz, G. Galati, Visuospatial transformations and personality: evidence of a relationship between visuospatial perspective taking and self-reported emotional empathy, Exp. Brain Res. 233 (2015) 2091–2102, https://doi.org/10.1007/s00221-015-4280-2.

[96] S. Baron-Cohen, S. Wheelwright, J. Hill, J. Raste, I. Plumb, The ‘reading the mind in the eyes’ test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism, J. Child Psychol. Psychiatry 42 (2001) 241–251, https://doi.org/10.1111/1469-7617.00715.

[97] B.F.M. Oakley, R. Brewer, G. Bird, C. Catmur, Theory of mind is not theory of emotion: a cautionary note on the reading the Mind in the Eyes test, J. Abnorm. Psychol. 125 (2016) 818–823, https://doi.org/10.1037/abn0000182.

[98] K. Kessler, H. Wang, Spatial perspective taking is an embodied process, but not for everyone in the same way: differences predicted by sex and social skills score, S. Calp. Cogn. 12 (2012) 133–158, https://doi.org/10.1016/j.scogn.2012.08.001.

[99] A.W. Qureshi, L.A. Apperly, D. Samson, Executive function is necessary for perspective selection, not level-1 visual perspective calculation: evidence from a dual-task study of adults, Cognit. Sci. 117 (2010) 230–236, https://doi.org/10.1016/j.cognition.2010.07.015.

[100] A.W. Qureshi, R.L. Monk, Executive function underlies both perspective selection and calculation in level-1 visual perspective taking, Psychon. Bull. Rev. 25 (2018) 1526–1534, https://doi.org/10.3758/s13423-018-1496-4.
T. Grootswagers, S.G. Wardle, T.A. Carlson, Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data, J. Cogn. Neurosci. 29 (2016) 677–697, https://doi.org/10.1162/jocn_a_01068.

A. Gunia et al. from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data, Nat. Neurosci. 21 (2018) 474–483, https://doi.org/10.1038/s41593-018-0108-2.

D. Vistoli, M.-A. Lavoie, S. Sutliff, P.L. Jackson, A.M. Achim, Functional MRI examination of empathy for pain in people with schizophrenia reveals abnormal activation related to cognitive perspective-taking but typical activation linked to affective sharing, J. Psychiatry Neurosci. 42 (2017) 262–272, https://doi.org/10.1503/jpn.160136.

D. Schubring, M. Kraus, C. Stolz, N. Weiller, D.A. Keim, H. Schupp, Virtual reality potentiates emotion and task effects of Alpha/Beta brain oscillations, Brain Sci. 10 (2020) 387, https://doi.org/10.3390/brainsci10030387.

B. Sörgels, D.J. Barr, S. Garrod, K. Kessler, Conversational interaction in the scanner: mentalizing during language processing as revealed by MEG, Cereb. Cortex 25 (2015) 3219–3234, https://doi.org/10.1093/cercor/bhu116.

U. Toelch, D.R. Bach, R.J. Dolan, The neural underpinnings of an optimal exploitation of social information under uncertainty, Soc. Cogn. Affect. Neurosci. 9 (2014) 1746–1753, https://doi.org/10.1093/scan/nst173.

B. Berluci, S.M. Aglioti, The body in the brain revisited, Exp. Brain Res. 200 (25) (2009), https://doi.org/10.1007/s00221-009-1970-7.

A.K. Martin, J. Huang, A. Hunsold, M. Meinzer, Dissociable roles within the social brain for self-other processing: a HD-DCS Study, Cereb. Cortex 29 (2019) 3642–3654, https://doi.org/10.1093/cercor/bhy238.

Mf Wurm, Ri Schubotz, The role of the temporoparietal junction (TPJ) in action observation: agent detection rather than visuospatial transformation, Neuroimage 165 (2018) 48–55, https://doi.org/10.1016/j.neuroimage.2017.09.064.

M. Schurz, M. Kronichler, S. Weissengruber, A. Surtees, D. Samson, J. Perner, Clarifying the role of theory of mind areas during visual perspective taking: issues of spontaneity and domain-specificity, Neuroimage 117 (2015) 386–396, https://doi.org/10.1016/j.neuroimage.2015.04.031.

J. Parvizi, S. Kastner, Promises and limitations of human intracranial electroencephalography, Nat. Neurosci. 21 (2018) 474–483, https://doi.org/10.1038/s41593-018-0108-2.

T. Grootswagers, S.G. Wardle, T.A. Carlson, Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data, J. Cogn. Neurosci. 29 (2016) 677–697, https://doi.org/10.1162/jocn_a_01068.

J.V. Haxby, Multivariate pattern analysis of fMRI: the early beginnings, Neuroimage 62 (2012) 852–855, https://doi.org/10.1016/j.neuroimage.2012.03.016.

C. Corradi-Dell’acqua, A. Tusche, P. Vuilleumier, T. Singer, Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex, Nat. Commun. 7 (2016) 10904, https://doi.org/10.1038/ncomms10904.

N.N. Oosterhof, S.P. Tipper, P.E. Downing, Viewpoint (in)dependence of action representations: an MVPA study, J. Cogn. Neurosci. 24 (2012) 975–989, https://doi.org/10.1162/jocn_a_00175.

R.I. Hashimoto, T. Ishashi, H. Ohta, T. Yamada, K. Kanai, M. Nakamura, H. Watanabe, Kato-N, Altered effects of perspective-taking on functional connectivity during self- and other-referential processing in adults with autism spectrum disorder, Soc. Neurosci. 12 (2017) 661–672, https://doi.org/10.1080/17470919.2016.1224202.

S.M. Eack, J.A. Wojtalek, M.S. Keshavan, J.N. Minshew, Social-cognitive brain Function and connectivity during virtual perspective-taking in Autism and Schizophrenia, Schizophr. Res. 183 (2017) 102–109, https://doi.org/10.1016/j.schres.2017.03.009.

P.J. Matusz, S. Dikker, A.G. Huth, C. Peridon, Are we ready for real-world neuroscience? J. Cogn. Neurosci. 31 (2019) 327–338, https://doi.org/10.1162/jocn_a_01276.

D. Bevilacqua, I. Davidecso, L. Wan, K. Chaloner, J. Rowland, M. Ding, D. Poeppel, S. Dikker, Brain-to-brain synchrony and learning outcomes vary by student-teacher dynamics: evidence from a real-world classroom electroencephalography study, J. Cogn. Neurosci. 31 (2019) 401–411, https://doi.org/10.1162/jocn_a_01274.

I. Davidecso, Brain-to-brain synchrony in the STEM classroom, CBE Life Sci. Educ. 19 (2020) 1–6, https://doi.org/10.1187/cbe.19-11-0252.

S.E. Kobert, J. Kurzmann, C. Neuper, Cortical correlate of spatial presence in 2D and 3D interactive virtual reality: an EEG study, Int. J. Psychophysiol. 83 (2012) 365–374, https://doi.org/10.1016/j.ijpsycho.2011.12.003.

J. Diemer, G.W. Alpers, H.M. Peperkorn, Y. Shiban, A. Mühlberger, The impact of perspective and presence on emotional reactions: a review of research in virtual reality, Front. Psychol. 6 (2015), https://doi.org/10.3389/fpsyg.2015.00026.

D. Schubring, M. Kraus, C. Stolz, N. Weiller, D.A. Keim, H. Schupp, Virtual reality potentiates emotion and task effects of Alpha/Beta brain oscillations, Brain Sci. 10 (2020) 537, https://doi.org/10.3390/brainsci10080537.

K. Kätlin, R. Groten, M. Slater, The sense of embodiment in virtual reality, Presence: CBRM 21 (2012) 373–387, https://doi.org/10.1162/PRES_a_00142.

P. Gunalp, T. Moosmann, M. Hegarty, Spatial perspective taking: effects of social, directional, and interactive cues, Mem. Cognit. 47 (2019) 1031–1043, https://doi.org/10.3758/s13421-019-00910-y.

F. Schoeller, P. Bertrand, L.J.erry, G. Jain, A. Horowitz, F. Zenatti, Combining virtual reality and biofeedback to foster empathic abilities in humans, Front. Psychol. 9 (2019), https://doi.org/10.3389/fpsyg.2018.02741.

N. Tsouros-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, M. Joliot, Automated anatomical labeling of activations in SPM using a Macroscopic Anatomical Parcellation of the MNI MRI single-subject brain, NeuroImage 15 (2002) 273–289, https://doi.org/10.1006/nimg.2001.0978.

M. Xia, J. Wang, Y. He, BrainNet Viewer: a network visualization tool for human brain connectomics, PLoS One 8 (2013), e66910, https://doi.org/10.1371/journal.pone.0068910.

A. Diamond, Executive Functions, Annu. Rev. Psychol. 64 (2013) 135–168, https://doi.org/10.1146/annurev-psych-113011-143750.

A. Leon, J. Ballonnet, J. Zaki, J. Bostick, R. Willer, Virtual reality perspective-taking increases cognitive empathy for specific others, PLoS One 13 (2018), e0202442, https://doi.org/10.1371/journal.pone.0202442.