Límite | ***Revista Interdisciplinaria de Filosofía y Psicología***
Límite | ***Interdisciplinary Journal of Philosophy & Psychology***
(2022) 17: 4

UNDERSTANDING THE EMBODIED NATURE OF SOCIAL COGNITION THROUGH VISUOSPATIAL PERSPECTIVE-TAKING

COMPRENDER LA NATURALEZA CORPORIZADA DE LA COGNICIÓN SOCIAL A TRAVÉS DE LA TOMA DE PERSPECTIVA VISOESPACIAL

Andrea Wagner-Ferrari*
Universidad Austral de Chile
Valdivia - Chile

Carla Kuschel Ruiz
Asociación Chilena de Seguridad
Santiago - Chile

Recibido febrero de 2021/Received February, 2021
Aceptado noviembre de 2021/Accepted November, 2021

ABSTRACT

Embodied cognition is anchored from experiences of the individual with her environment through sensorimotor capacities that define her perception and action. In research on visuospatial perspective-taking (VSPT), which is a fundamental social ability to imagine the perspective from someone else's spatial location, an experimental paradigm has been used. According to this paradigm, judgments must be made based on what the other person sees, allowing us to discover the specific mechanisms involved in the process. The most important mechanism is mental self-rotation. In this process, the subject mentally rotates her body until she reaches the "position" that allows her to see things as the other person sees them. This also requires the emulation of the other's body posture, such that, when congruent, facilitates the adoption of the new perspective. Recent researches that identify these mechanisms were reviewed, revealing the usefulness of this experimental paradigm in the study of social cognition's embodied character.

Key Words: Visuospatial perspective-taking; embodied social cognition; angular disparity; congruence of body posture; mental self-rotation.

RESUMEN

La cognición corporizada está anclada en las experiencias del individuo con su entorno mediante sus capacidades sensoriomotoras que definen su percepción y acción. En la investigación respecto de toma de perspectiva visoespacial, que es la habilidad social fundamental que permite imaginar el punto de vista de otro desde su localización espacial, se ha utilizado un paradigma experimental en el que se deben realizar juicios de lo que otro ve, que ha permitido descubrir mecanismos específicos involucrados en el proceso. El más importante de estos mecanismos es la autorrotación mental, en donde el sujeto rota mentalmente su cuerpo hasta alcanzar la "posición" que le permite ver las cosas como el otro las ve. Esto también requiere la emulación de la postura corporal del otro, de manera que, cuando esta es congruente, facilita la adopción de la nueva perspectiva. Investigación reciente que identifica estos mecanismos fue revisada, revelando la utilidad de este paradigma experimental en el estudio del carácter corporizado de la cognición social.

Palabras Clave: *Toma de perspectiva visoespacial; cognición social corporizada; disparidad angular; congruencia de la postura corporal; autorrotación mental.*

* Autor correspondiente / Corresponding author: andrea.wagner@uach.cl

INTRODUCTION

Human social cognition refers to the sum of those processes that allow individuals of the same species (conspecifics) to interact with one another (Frith & Frith, 2007). Being able to adopt another person's perspective is an important feature of these processes (Schurz *et al.*, 2013).

Human beings think in a spatial manner (Levinson, 1996), and we have a spatial perspective when we approach how we stand in the world. This is formed by the activities we are involved in (Heidegger cited in Proulx *et al.*, 2016), and where we operate in complex social and spatial environments. In these contexts, other people are particularly important (Surtees *et al.*, 2013), and the ability to take their visuospatial perspectives is necessary (Erle & Topolinski, 2017). First, the fact that what the other individual sees may differ from what we see must be considered. Namely, assume a primary distinction between self and others, and between one's own subjectivity and experience of the world, and the subjectivity and experience of others (Glaveanu, 2019). Second, this requires developing an understanding of how physical space is perceived by another individual, from her own point of view (Fischer & Demiris, 2018; Xie *et al.*, 2018).

Taking the visuospatial perspectives of others allow us to understand where the objects are in relation to their egocentric frames of reference (Michelon & Zacks, 2006), and where our own egocentric frame of reference is located in relation to theirs (Zacks & Michelon, 2005). This capacity allows us to establish common frames of reference from which we can create a shared world, where we can interact with each other (Frith & Frith, 2007). For example, the point to which they direct their attention, provides critical information for the monitoring of social interactions (Amorim, 2003), and even the simplest ones require anticipating others' actions to coordinate their own (Tversky & Hard, 2009). Those common frames of references allow the development of intersubjective understandings about events and things, which are related to the capacity of decentered perception, in the sense of Piaget (Habermas, 1999). This also provides the context in which people can take perspective to reflect upon their own situation, and participate in the situation of others (Gillespie, 2012). In a more sociological level, Gillespie (2012) proposed the idea of position exchange to explain how people can take other's perspective, specifying that each social position sustains a distinctive motor and perceptual perspective, which takes place as long as people are exchanging social positions.

The interaction between shared realities and efforts, and between individual perspectives and roles, that constitute the two levels structure that Tomasello (2019) points out, is the cause of the most distinctive aspects of the human species. This began early in the ontological develop of human beings, who gradually integrate institutions of more complexity. Therefore, people achieve larger targets. But, as the author claims, taking up a role in some larger social enterprise is qualitatively different from interacting with other individuals directly, and that escapes the aims of this manuscript. This article's focus is the interpersonal relation level and its embodied roots.

Visuospatial perspective-taking (VSPT) is part of the infrastructure of the ability to take a psychological perspective of another individual, which is also a component of this capacity (Schurz *et al.*, 2018). In turn, this allows the emergence of social cognitive processes, such as joint attention and collaborative work that involves taking on complementary roles.

Researchers, who have used the VSPT's experimental paradigm, have allowed the identification of specific mechanisms of this process. With these mechanisms, it is possible to visualize concretely the way in which embodied cognition operates, from the action of the organism over its environment. These also refer to mental self-rotation, from which the point of view of another person is adopted, and the influence of the body posture congruence degree between the corporal posture of who adopts the visuospatial perspective of another individual, and the corporal posture of that other subject, whose visuospatial perspective is adopted in the process of mental self-rotation.

In summary, VSPT is a social cognitive process representative of a high-level perspective-taking in general (Kessler & Wang, 2012), which motor sensory roots can be mapped from the relation of the organism and the space in which it develops.

From the perspective of embodied cognition, it is fundamental to consider that the mind is always located in a body that always occupies a specific place in space and looks towards a specific direction (Tversky & Hard, 2009). Bodily states, situated action, and modal simulations underlie

cognition (Barsalau, 2008) so that, for example, the visuospatial induction of perspective-taking alone can lead to take a psychological perspective, which summarizes the process of obtaining an affective and cognitive perspective (Erle & Topolinski, 2017).

Embodied cognition accounts of social cognition reject cognitivist theories, which provide the idea that mindreading explains how we understand and interact with others (Spaulding, 2012). Those embodied cognition accounts make a turn towards embodied, enactivist approaches to social cognition, which emphasize social interaction, the role of action, and direct social perception in rich environments (e.g., Gallagher 2005, 2008a; Ratcliffe 2007; Reddy & Trevarthen 2004, cited in Gallagher & Allen, 2018), which aim to explain how our embodiment shapes our knowledge of others. This also emphasizes the idea that our normal everyday interactions consist in non-mentalistic embodied engagements (Spaulding, 2012).

Regarding the question about the nature of embodied cognition, research agenda is based on typical examples, which proponents of 4E Cognition (including embodied, embedded, enactive, and extended approaches to cognition) consider associated to basic cognition processes, encompassing those associated to social cognition (Newen *et al*., 2018). In this vein, considering that VSPT is a basic social cognitive process, it is relevant to review how findings related to this process may inform embodied cognition accounts of social cognition.

Therefore, this manuscript aims to review the contributions of researches made through the VSPT experimental paradigm to understand the embodied character of social cognition. This contemplates the review of the main findings, to later analyse its usefulness linked to the study of embodied social cognition.

To achieve this, fundamental contributions of embodied social cognition accounts will be reviewed, considering where are located among other approaches included in 4E Cognition, emphasizing the frameworks to which VSPT can inform and, in turn, be appropriately interpreted. Then, fundamental findings of VSPT will be reviewed, to analyse findings from the point of view of embodied cognition. Consequently, the role representations can or cannot have will be discussed, in relation with VSPT, to later suggest possible further steps in the research agenda. This will help deepen how the embodied character of social cognition can be understood using VSPT experimental paradigm.

EMBODIED COGNITION

Under the conceptual umbrella of embodied cognition, various approaches to cognition come together. Even though the boundaries between those approaches are yet to be defined, at least within their margins, all of them have in common the idea that the body, or the body's interactions with the environment, constitute or contribute to cognition (Shapiro & Spaulding, 2021, forthcoming), rejecting cognitivist accounts of cognition.

Cognitivist accounts "attempted to understand intelligence in terms of production, transformation, and manipulation of inner states that represented properties of the domain that the cognizer was trying to deal with" (Ward *et al*, 2017, p. 365). But, as Hutto and Myin (2018) point out, that kind of processing is "deemed too slow, rigid, and abstractly formatted to properly account the dynamically updated character of real-time intelligent activity" (p. 96).

In their seminal work, Varela, Thompson, and Rosch (1991, cited in Newen *et al.*, 2018), proposed an enactivist account of cognition that emphasized the role of the dynamical coupling of brain-body-environment, supporting a vision of mind as emergent, embodied, and engaged (Ward *et al*, 2017). In addition, proponents of 4E cognition claim that cognitive processes are all dependent on the details of an agent's body; an appropriately structured natural, technological, or social environment; and the agent's active and embodied interaction with this environment (Newen *et al.*, 2018). At this point it is important to consider, despite the distinctions between the four "Es" –embodied, embedded, enactive, and extended– that it is not uncommon to use the label "embodied" to include any or all of these (Shapiro & Spaulding, 2021, forthcoming). Therefore, while reviewing some fundamental elements of embodied cognition approach, this manuscript will emphasize how the label "embodied" frames the approaches that conform 4E cognition.

As enactivism emphasize, self-organize cognitive structures emerge because of interactions between an organism and portions of its environment, and from the details of these interactions, which are sustained by the way the organism is embodied (Ward *et al.*, 2017). Thus, cognition refers to behaviour, rather than computed representations

(Di Paolo *et al.*, 2010, cited in Aizawa, 2014). In this vein, dynamicits as Chemero propose that cognitive scientists must try to understand cognition as intelligent behaviour (2009, cited in Aizawa, 2014), since one's mental life is not sustained in one's brain, but in one's physical activities in the world (Nöe, 2009, cited in Aizawa, 2015).

In this way, 'embodiment' means that cognition is developed by, constituted by, or supervenes upon other parts of the body, not just the brain (Aizawa, 2015). It also means that perception consists in a perceptually guided action, which occurs in virtue of recurrent sensorimotor patterns, from which cognitive structures emerge (Varela *et al.*, cited in Ward *et al.*, 2017). From this approach to cognition, individuals are understood as situated dynamical systems (Hutto & Myin, 2018), which are "always open to reconfigurations, enabling them to self-organize quickly and flexibly to the contextual demands" (Kelso, 1995, cited in Hutto & Myin, 2018, p. 98).

From the ecological program originally developed by James Gibson (1979, cited in Rietveld *et al.*, 2018), what we perceive are affordances, which are opportunities to engage with the environment in ways that reflect our needs and plans, and which are mediated for our capacities (Ward *et al.*, 2017). The notion of affordances was recreated by Rietveld *et al.* (2018) in their Skilled Intentionality Framework (SIF). This is an ecological-enactive approach to cognition that brings together the aforementioned ecological program and the embodied/enactive program (Chemero, 2009; Thompson 2007; cited in Rietveld *et al.*, 2018).

SIF focuses on the agent's active and embodied interaction with this environment, through skilled action, and defines skilled intentionality as the selective engagement of the agent, with multiple affordances occurring simultaneously in a concrete situation (Rietveld *et al.*, 2013; Bruineberg & Rietveld, 2014; Kiverstein & Rietveld, 2015; Van Dijk & Rietveld, 2017; in Rietveld *et al.*, 2018). This approach offers a framework to understand social interaction, among other processes of skilled action. It also connects relevant abilities to be able to act on an affordance, which human beings can display, and affordances present in the socio-material environment that human beings build while mutually interacting. This article reviews some approaches of embodied cognition to the understanding of social cognition and, among them, some contributions that SIF makes to it. This analysis is presented below.

EMBODIED SOCIAL COGNITION

Only recently the association between embodied cognition and social cognition has been acknowledged (Marmolejo-Ramos & D'Angiulli, 2014). From cognitivism, Theory Theory and the Simulation Theory approaches intend to explain social cognition processes. While the former argues that we explain and predict behaviour by employing folk psychological theories about how mental states inform behaviour, simulation theorists argue that we explain and predict a target's behaviour by using our own minds as a model (Spaulding, 2012). However, they agree that the explanation about how we understand others and interact with them is the mindreading process (Spaulding, 2012), which is defined as the human cognitive ability to represent the psychological states of either self or others (Jacob, 2009).

Embodied cognition theorists have rejected this construal of social cognition, arguing that mindreading is a specialized and rarely used skill in ordinary social interactions (Gallagher 2020; Gallagher 2008; Hutto and Ratcliffe 2007, cited in Shapiro and Spaulding, 2021, forthcoming). In contrast, Embodied Social Cognition accounts hold that "what underlies our ability to understand and interact with others is the capacity for basic, non-mentalistic, interactive embodied practices" (Spaulding, 2012, p. 433). These aim to explain how our embodiment shapes our knowledge of others, emphasizing the idea that our normal everyday interactions consist in non-mentalistic embodied engagements, and the social cognition processes that are carried out online (Spaulding, 2012).

Most social interactions require only basic social cognitive capacities that are known as primary and secondary intersubjectivity (Trevarthen 1979, cited in Shapiro and Spaulding, 2021, forthcoming), which are sufficient for navigating most typical social interactions (Shapiro and Spaulding, 2021, forthcoming). The former refers to the innate or early developing capacity to interact with others, manifested at the level of perceptual experience, so that we perceive in the other person's bodily movements, facial gestures, eye direction, and so on, what they intend and feel (Gallagher, 2005, cited in Spaulding, 2012). From embodied cognition approach, this does not involve the representation

of those features (Shapiro & Spaulding, 2021, forthcoming). In contrast, it simply requires certain practical abilities, which are sensitive to certain bodily cues (Spaulding, 2012).

Regarding secondary intersubjectivity, in this stage children develop the capacity to communicate with others about objects and events. To share their perception of the world, children must align their focus of attention with other. This process is called "joint attention" and is typically achieved by pointing at an object and/or following another person's eye gaze direction (Frith & Frith, 2007). The ability to share attention in this way can be observed in infants as young as 12 months (Liszkowski *et al.*, 2004; Tomasello *et al.*, 2005; cited in Frith & Frith, 2007), and, according to embodied social cognition, this social understanding is still non-mentalistic (Gallagher 2005, cited in Spaulding, 2012). Finally, as was pointed out by several researchers, despite primary and secondary intersubjectivity have been studied primarily in developmental terms, according to embodied social cognition, these practices are our primary mode of social cognition as adults (Fuchs 2012; Gallagher 2008; cited in Shapiro & Spaulding, 2021, forthcoming).

In this vein, Direct Social Perception approach to social cognition (Krueger, 2018) claims that we directly perceive mental states which are embodied in behaviour. As Gallagher and Zahavi argue, in seeing actions and expressive movements of other persons, one already sees their meaning, and no inference to a hidden set of mental states is necessary (2008, cited in Krueger, 2018). From this approach, one way to explain the relation between mind and behaviour is to characterize the former as perceptually co-present with the latter (Krueger, 2018). In this sense, although we see just a perspective of an object, we experience it as three-dimensional, since perception is a process structured by anticipatory appresentations of what we would see if we turned around our head, for example (Krueger, 2018). The relation between what we have just seen and what we will see as we move is a kind of knowledge acquired through actively exploring the space, which conforms our quotidian socio-material environments, and from which stable patterns that relate to movement and vision are established. Consequently, "sensorimotor contingencies" are formed, which in turn mediate our vision as a mode of exploration of the world (O'Regan & Nöe, 2001). Applying the principle then-if, which is implicit in the "sensorimotor contingencies" approach of perception to direct social perception, Smith argues that, although we only see the behaviour of others, we experience their mental states as co-present in relation to such behaviour (2010, in Krueger, 2018).

Considering the relevance of the agent's active and embodied interaction with this environment on the embodied cognition approach, Skilled Intentional Framework (SIF) brings an account of embodied cognition as skilled engagement with multiple affordances (Rietveld *et al*, 2018), and provides a framework to understand social cognition. SIF recreates Gibson's notion of affordances, which means a set of possibilities for action provided to us by the environment (Gibson, 1979; Chemero, 2003, 2009; Michaels, 2003; Reed, 1996; Costall, 1995; Heft, 2001; Rietveld and Kiverstein, 2014, cited in Rietveld *et al*, 2018), on which an individual can act in virtue of the possession of a relevant ability. According to this framework, the socio-material environment offers all sorts of possibilities to human beings, including possibilities for social interaction.

This framework defines a landscape of affordances as the ecological niche of certain form of life (Wittgenstein, 1953, cited in Rietveld *et al.*, 2018), where a specific form of life, (an animal, for instance), deploys practices that characterize it. These are constituted by coordinated patterns of behaviour, common to the individuals of that kind. As a form of life, human beings are characteristically social, and others and their behaviours are part of their landscape of affordances (Rietveld *et al.*, 2018), where coordinated patterns of social behaviour are deployed. Following the direct social perception approach to social cognition, those distinctive patterns of social behaviour refer to sensorimotor regularities, from which we can experience sensory cues connected to co-present mental states, during social interactions. Typically, mental states that are co-present in social interactions represent the other subject's intentions. Thus, through following her eye gaze and take her visuospatial perspective, we experience not only what the other subject observes, but also why they observe it. (Tomasello, 2008).

SIF also provides a structure to understand the social behaviour of an individual in a concrete situation, called "field of relevant affordances". This structure explains how an individual can be selectively open to the landscape of affordances that characterised her form of life. This field of

relevant affordances is formed by solicitations, which "are the affordances that show up as relevant to a situated individual and generate bodily states of action readiness" (Rietveld *et al*, 2018, p. 52). Tomasello (2008) posits the Cooperation Model to explain that human beings have unique ways to engage with each other socially. This model allows the visualization of how an affordance becomes a solicitation in a social context. Thus, from this model, "the communicative context is what is 'relevant' to the social interaction, that is, what each participant sees as relevant and knows that the other sees as relevant as well" (Tomasello, 2008, p. 74). In this intersubjective context, the common frame of reference is constituted through processes of joint attention, in which those who participate in it orient their visuospatial perspectives to aspects of the context. These aspects turn out to be relevant for them, since they constitute relevant affordances to act on, namely, solicitations. This article will review the fundamental findings of VSPT, which can inform aspects of embodied social cognition accounts delineated above. This analysis is presented below.

VISUOSPATIAL PERSPECTIVE-TAKING

The ability to take the visuospatial perspective of others refers to the ability to understand the spatial relationship between the individual and the objects of their environment (Surtees *et al.*, 2013). This capacity leads us to change our visuospatial frame of reference to theirs and, from there, to establish common frames of reference, which form the socio-cognitive and motivational infrastructure to interact with others (Liszkowski, 2018).

A frame of reference is a group of axes from which to consider the location of objects (Levinson, 1996), which can be absolute or relative. The latter is defined by the position of the objects in relation to who sees them (Surtees *et al.*, 2013). It is possible to distinguish between ego and allocentric frames of reference; the first is the one in which one denotes the location of something in reference to an agent, and this representation of the location is directly linked to the actions that the agent can perform towards the object, focusing on that agent (Frith & De Vigdemont, 2005). In contrast, an allocentric frame of reference is the one in which the location of something has a reference to other objects, to which agency is not attributed and does not change if the agent moves (Frith & De Vigdemont, 2005).

Flavell (1988) noted that the ability to understand the visual perspective of another individual occurs at two levels; the first one reflects an understanding of what is in the line of sight of another, and the second one involves the capacity to mentally adopt another individual's point of view and understanding how the world looks from this new perspective. The first level of visual perspective-taking is also called "perspective tracking" and it is described as the process by which an observer understands what the other can or cannot perceive (i.e., what it is visible or hidden for them). This capacity is reached around the age of two years old (Lempers *et al.*, 1977). The second level of visual perspective-taking is a more complex process, which requires to understand the perspective, and this implies that different people can have different points of view of the same state of things, and, in addition, be able to intentionally change perspective (Schurz *et al.*, 2013). This second level has a posterior ontogenetic development, around the age of four or five years old (Gzesh & Surber, 1985).

Surtees *et al.* (2013, cited in Xie *et al.*, 2018) extends this distinction from visual perspectives to experiences of spatial relationships by pointing out that, if an objective is ahead of the agent (visible) or behind it (invisible), it constitutes a first level visuospatial perspective. On the other hand, knowing if the objective is on the left or right side of the agent's perspective constitutes a second level visuospatial perspective. First level visuospatial perspective-taking is sensitive to the distance between the agent and the object from which perspective is taken and involves a tracking line of sight. Meanwhile, second level visuospatial perspective-taking is used when a judgement must be made regarding the position in which a specific object is placed, right or left side from the agent's position.

The ability to take second-level visuospatial perspectives has been investigated using the experimental paradigm of visuospatial perspective-taking, in which participants perform tasks where they are asked to make judgments about the visuospatial perspective of another individual. These tasks have been created based on the complex Test of the Three Mountains (Piaget & Inhelder, cited in Mounoud, 2001), developed to assess a child's ability to considerate another individual's point of view, in relation to several objects (three mountains in a model), indicating how they think a doll sees the three mountains. Currently, 3D graphic

computing tools have allowed the model to be replaced by a virtual environment, and the doll by an avatar that inhabits the environment. With this method, the observer can change from an "out of body" perspective of their avatar to an "embodied" perspective of looking at the world through the eyes of their avatar (Amorim, 2003). From this paradigm, aspects such as the angular disparity existing between the position of the agent and the avatar, and the effect of the avatar's body posture, have been systematically manipulated to measure the cognitive effort associated with visuospatial perspective-taking, and unravel the mechanisms that explain it (Schurz *et al.*, 2018).

Angular Disparity

To evaluate the effect of angular disparity on the ability to take visuospatial perspective of another individual, the consistency between the participant's perspective and the agent's perspective has been systematically manipulated. Using this mechanism, Zacks & Michelon (2005) reported that the speed and precision of visuospatial perspective-taking decrease as angular disparity between participant's point of view and avatar's point of view increases, which reveals the increase in cognitive effort associated with the increase of angular disparity. From these results, it has been possible to hypothesize that visuospatial perspective-taking of another individual is based on an embodied cognitive process called "mental self-rotation". This process allows the visuospatial perspective-taking and the exchange of perspectives, and, therefore, of the frames of reference between one subject and another.

In this way, it is possible to affirm that visuospatial perspective-taking, which refers to a transformation (or more precisely, a translocation) of the egocentric perspective, is a process strongly rooted in the representations of the body, and in its repertoire of movements (Kessler & Thomson, 2010). This involves a mental reorientation of oneself (Kozhevnikov *et al.*, 2006), and where the perspective taker simulates the rotation of the whole body by situating itself in the position of the agent (Kessler *et al.*, 2014).

However, the relationship between angular disparity and the cognitive effort involved is not strictly linear. In their research, Kessler & Thomson (2010) account for two alternative mechanisms that are activated differently, depending on the amplitude of the angular disparity between the participant and the agent. They concluded that, at broader angles, mental self-rotation is necessary to take the visuospatial perspective of the other subject, while in less broad angles a process of visual matching is conducted, since the perspective of the agent is still broadly aligned with the egocentric perspective of who should make a judgment about the perspective of the other individual. This explains the abrupt increase of response times in visuospatial perspective-taking tests between 60° and 90° of angular disparity (Michelon & Zacks, cited in Kessler & Thomson, 2010). On the other hand, to determine whether the effect of angular disparity differs for each level of visuospatial perspective-taking, Kessler & Thomson (2010) and Kessler & Rutherford (2010) found, as Michelon & Zacks (2006) did before, that in widest angular disparities, response times for second level visuospatial perspective-taking increase as angular disparity increases, while response times for first level perspective-taking were not affected by angular disparity and remained constant. In conclusion, in second level visuospatial perspective-taking there is an increase in cognitive effort with an increase of angular deviation between the egocentric perspective and that of the agent. This is not the case for the first level perspective-taking.

Body Posture's Congruence

Besides angular disparity, regarding the second level visuospatial perspective-taking, the effect of congruence of body postures between the participant and the agent has been studied. To distinguish the effect of angular disparity and the congruence of posture, Surtees *et al.* (2013) manipulated both angular disparity between the position of the avatar and the position of the participant, and the congruence of body posture between both. To achieve this, participants sat in a rotating chair, but kept their faces fixed in front of the screen. This allowed them to change the position of their body, but not their visual perspective, which remained fixed while varying angular disparity in relation to the avatar. The authors found that participants tended to perform better in tasks of visuospatial perspective-taking when the position of their body was aligned with the position of the avatar, which confirmed the results of a previous study designed by Kessler & Thomson (2010) to determine the differentiated effect of angular disparity and body posture. In this study, they concluded that when the test involves an avatar, that is, when there is

a body position to emulate, before being able to rotate, the participant must mentally adopt the position of the avatar before beginning the process of rotation from it. Thus, posture emulation is the first embodied mechanism of visuospatial perspective-taking, which is consistent with the perspective of embodied transformation (embodied transformation account). It indicates that the cognitive effort implied by the angular disparity in wider angles decreases when the body posture is congruent and increases when it is incongruent (Kessler & Thomson, 2010). Additionally, Kessler & Rutherford (2010) observed that the congruence of the avatar's posture accelerates the response times in angular disparities of 120° and 160°, and, in general, it is associated with fewer errors. This led them to conclude that an adjustment between rotated self and body posture provides a stop sign for the rotation process.

The influence of posture emulation is also consistent with the direct-matching hypothesis, which states that a perceived action or posture is directly emulated by the observer, which requires an exogenous visual input to resonate with. Kessler & Thomson (2010) suggest defining this type of motor embodiment as exogenous, referring to the fact that the same motor representations and their neural correlations are activated during processes as different as observation, imagination, or actions (Amorim *et al.*, 2006). In contrast, they propose the term "endogenous motoric embodiment" when a visual input (that is, a body posture to emulate) is not necessarily required and, instead, a movement that is already in the participant's motor repertoire is emulated. This endogenous process of motor embodiment is evident when the test does not involve adopting another individual's perspective but imagining how the participant will represent the world in a visuospatial manner if they occupy a new position (for example, in an empty chair) (Kessler & Thomson, 2010). In this way, it is possible to conclude that the default strategy for second-level visuospatial perspective-taking consists in simulating the rotation of the body (Kessler & Thomson, cited in Kessler & Rutherford, 2010), facilitated by the congruence of body posture, which is possible even when there is no position to emulate.

DISCUSSION

In the review of the most fundamental contributions of visuospatial perspective-taking experimental paradigm to understand the embodied character of human social cognition, references to representations in explanations of cognitive processes, which underlie this behaviour, were found. However, there were no specifications to what kind of representations those findings refer to exactly. Regarding that the role of representations constituted a fundamental difference between cognitivist and embodied accounts of cognition and, considering the affirmation that those findings inform the last ones, it is important to discuss what kind of representations it is possible to infer, to which those findings allude.

Since visuospatial perspective-taking of another individual requires "mental self-rotation", which allows the exchange of perspectives; and posture incongruence requires a mental alignment with the avatar's corporal posture, and both of those cognitive processes implied a cognitive effort, the following question arises: To what kind of cognitive processes this effort must be attributed?

From cognitivism, which "takes representation as their central notion" (Varela *et al.*, 1991, p. 172, in Hutto, 2013), it seems obvious that what occurs through those cognitive efforts refers to the construction and manipulation of representations.

Specifically, from the perspective of cognitivist simulation-based approaches to social human cognition, it is assumed that processes of motor mirroring (action-mirroring), exemplified by the activity of mirror neurons, constitute instances of third-person mindreading by virtue of which an observer can represent an agent's intention (Jacob, 2009). Thus, several theorists posit body-formatted representations based on motor mirroring (Gallese & Sinigaglia, 2011; Goldman, 2012; cited in Hutto, 2013), which are behind social cognitive processes, and can be the type of representations to which revised VSPT findings allude. Embodied simulation refers to "the reuse of mental states and processes involving representations that have a bodily format" (Gallese & Sinaglia, 2011, p. 515), which have been characterized as a distinctive class of mental representations, by virtue of their format (with motor, visceromotor, and somatosensory profiles) rather than their content (Gallese & Sinaglia, 2011).

However, the recruitment of their own motor representations would only provide observers with an "emulation" of the motor commands suitable for achieving the goals and intentions of agents, not goals and intentions itself (Csibra, 2007; Jacob,

2009; cited in Gallese & Sinaglia, 2011), because instances of action-mirroring could not constitute (in a strong sense) instances of third-person mindreading (Jacob, 2009).

According to Goldman (2012, in Hutto, 2013), "body-formatted representations have a specific kind of content and a distinctive format for interfacing directly with certain machinery of mind" (p. 146), so that special vehicles –namely vehicles that can only be used by kinds of mental machinery– carry special contents and stand out precisely because they cannot be used by just any cognitive systems. However, despite neural re-use inclusion of the family of mirroring phenomena, in which motoric formats are redeployed for social cognition tasks, there is also a re-use of circuits associated with motor control functions in higher-level tasks of language comprehension, and, also, in memory tasks (Pulvermuller, 2005; Glenberg & Kaschak, 2002; Hubbard *et al.*, 2005, cited in Hutto, 2013). In this way, it does not seem to be appropriate to affirm that the re-use of motoric formats for social cognition tasks is evidence of the specificity of body formatted representations to certain kind of content. Instead, the "talk of representations [at least in] coupled systems may be too cheap, or too arbitrary, and, thus, adds little or nothing to an explanation of how these systems work" (Shapiro, 2011, p. 147, in Hutto, 2013).

Moreover, O'Regan & Nöe (2001) claim that it is not necessary to appeal to mysterious explanatory devices to understand how we perceive and how those perceptions guide our actions, because while we explore our environments, "sensorimotor contingencies" are established, which can be conceived as co-present while we perceive. In the same vein, following Krueger (2018) Direct Social Perception account to social cognition, while we perceive others' behaviours, we detect social cues, for example, their visuospatial perspectives, to which possibilities to act on are co-present.

NEXT STEPS

Regarding VSPT findings reviewed above and how they inform embodied sociocognitive processes, it is possible to affirm that, while the relationship between one's visuospatial frame of reference and the other's visuospatial frame of reference changes, different sensorimotor contingencies are automatically activated, since sensory cues that conduct our social behaviour are detected. In addition, considering contextual factors which can be present in our everyday life, and factors associated to each other participant in a situated social interaction, it may be possible, at least hypothetically, to establish which of those factors determine when the visuospatial perspective of an agent is perceived as an affordance, and when it is perceived as a solicitation, and through which sensory cues this is established. Also, from this point it is possible to visualize what kind of behaviours the sensory cues trigger, and to which neural activity they relate to.

From a phenomenologist perspective, it is important to ask the participants how they experience each condition of experimentation, specifically referring to co-presence of mental states that inform social interaction, from the Direct Social Perception approach.

CONCLUSIONS

The main findings of visuospatial perspective-taking experimental paradigm were reviewed to understand this social cognitive process, and to investigate whether those findings can inform, and how, to social embodied cognition accounts. Also, the role mental representations have, if any, in the explanation of this process, from embodied social cognition approach perspective, were discussed.

Regarding the necessity to understand the processes that underlie visuospatial perspective-taking, the experimental paradigm of visuospatial perspective-taking has allowed the manipulation of mechanisms involved in this process, identifying its embodied character. The systematic manipulation of the angular disparity has made possible to determine that the embodied mechanism of mental self-rotation, which underlies the second level visuospatial perspective-taking, begins to be used in major angular differences. This mechanism is also exclusive of this level of perspective-taking, making mechanisms that require less cognitive effort available for first level perspective-taking, such as following the line of sight. Regarding the effect of body posture on visuospatial perspective, it is possible to determine that it operates even when there is no posture to emulate, modulating the effect of angular disparity when it reaches maximum levels. Thus, the posture congruence seems to lessen the effect of angular disparity on the cognitive effort

associated with mental self-rotation. This allows the adoption of the other individual's position, whose visuospatial perspective is pretended to be adopted, noting that mental rotation of self has reached its goal.

In conclusion, considering the sensorimotor contingencies construct and its application to understand the Direct Social Perception approach fundamental proposals, as well as Skilled Intentionality Framework applied to social cognition processes understanding, it is possible to realize how possible it is to act on an object through the direct perception of sensory keys. With these, we experience this possibility as co-present, in such a way that it is presented as an affordance to us. Similarly, it is concluded that the possibility of taking another's visuospatial perspective is noticeable through sensory keys -eye gaze direction of the avatar in the virtual space, and its relation with our own egocentric visuospatial frame of reference; and body posture of the avatar and the relation it maintains with our own body posture- regarding how we apply our knowledge about sensorimotor keys that account not only for what we directly see, but for the possibilities for action that accompanies what we see. In this sense, another's person visuospatial perspective becomes an affordance, a possibility to take her visuospatial perspective, while in the concrete situation of the visuospatial perspective-taking experimental paradigm, the visuospatial perspective of the avatar becomes a solicitation, since the task expressly requires such a response.

The tendency to perceive the visuospatial perspective of others is innate. This implies that we configure our socio-material environments automatically in a specific situation, considering those visuospatial perspectives where the mere presence of others activates action readiness responses, considering the relevance of the communicational context factors concerned.

Contemplating the previous statements, it is concluded that both mechanisms, mental self-rotation (as it is shown through manipulation of angular disparity) and emulation of body posture, which are exogenous and endogenous (as it is shown through its effect on angular disparity), prove the embodied nature of visuospatial perspective-taking. Given the relationship between this form of perspective-taking and other forms of psychological perspective, there are new possibilities to develop systematic research on the embodied character of more abstract socio-cognitive skills, whose path to the sensorimotor origins of embodied cognition has been difficult to establish. In this context, it is considered that the embodied cognition approach, operationalized in the experimental paradigm of visuospatial perspective-taking and its mechanisms of mental self-rotation and adjustment of body posture, provides an adequate matrix to propose the existence of embodied cognitive processes, common to both types of perspectives, visuospatial and psychological, facilitating the understanding of how they occur and the existing relationship between them, submitting them to verification.

REFERENCES

Aizawa, K. (2014). The Enactivist Revolution. *AVANT*, *5*(2), 19-42, https://doi.org/10.12849/50202014.0109.0002

Aizawa, K. (2015). What is this cognition that is supposed to be embodied? *Philosophical Psychology*, 28 (6), 755-775, https://doi.org/10.1080/09515089.2013.875280

Amorim, M. (2003). "What's my avatar seeing?": the coordination of "out of body" and "embodied" perspectives for scene recognition across views. *Visual Cognition*, 10 (2), 157-199, https://doi.org/10.1080/713756678

Amorim, M., Isableu, B., & Jarraya, M. (2006). Embodied Spatial Transformations: "Body Analogy" for the Mental Rotation of Objects. *Journal of Experimental Psychology: General*, 135 (3), 327-347, https://doi.org/10.1037/0096-3445.135.3.327

Barsalou, L. (2008). Grounded Cognition. *Annual Review of Psychology*, 59, 617-645, https://doi.org/10.1146/annurev.psych.59.103006.093639

Erle, T. & Topolinski, S. (2017). The Grounded Nature of Psychological Perspective-Taking. *Journal of Personality and Social Psychology*, 112 (5), 683-695, https://doi.org/10.1037/pspa0000081.

Fischer, T. & Demiris, Y. (2018). A Computational Model for Embodied Visual Perspective-taking: From Physical Movements to Mental Simulation. In S. Ramalingam & M. Salzmann, *IEEE Conference on Computer Vision and Pattern Recognition Workshop on Vision Meets Cognition*. Conference held in 1st International Workshop on Computer Vision for Physiological Measurement (CVPM), Salt Lake City, Utah, USA.

Flavell, J. H. (1988). The development of children's knowledge about the mind: from cognitive connections to mental representations. In J.W. Astington, P.L. Harris and D.R. Olson (Eds.). *Developing theories of mind* (pp. 244-267). Cambridge University Press.

Frith, U. & De Vigdemont, F. (2005). Egocentrism, allocentrism, and Asperger syndrome Consciousness and Cognition. *Elsevier*, 14 (4), 719-738, https://doi.org/10.1016/j.concog.2005.04.006

Frith, U. & Frith, C. (2007). Social cognition in humans. *Current Biology,* 17 (16), 724-732, https://doi.org/10.1016/j.cub.2007.05.068

Gallagher, S. & Allen, M. (2018). Active inference, enactivism and the hermeneutics of social cognition. *Synthese*, *195*(6), 2627-2648, https://doi.org/10.1007/s11229-016-1269-8

Gallese,V. & Sinigaglia, C. (2011). What is so special about embodied simulation? Trends in Cognitive Sciences, 15, 512-519, https://doi.org/10.1016/j.tics.2011.09.003

Gillespie, A. (2012). Position exchange: The social development of agency. *New ideas in Psychology,* 30, 32-46, https://doi.org/10.1016/j.newideapsych.2010.03.004

Glaveanu, V. (2019). Being other: Intersubjectivity, allocentrism and the possible. *Journal for Theory of Social Behavior*, 1-17, https://doi.org/10.1111/jtsb.12221

Gzesh, S. M. & Surber, C. F. (1985). Visual perspective-taking skills in children. *Child Development*, 56 (5), 1204-1213, https://doi.org/10.2307/1130235

Habermas, J. (1999). *Teoría de la Acción comunicativa I*. Grupo Santillana Ediciones.

Hutto, D. (2013). Exorcising action oriented representations: ridding cognitive science of its Nazgûl. *Adaptive behavior*, 21 (3), 142-150, https://doi.org/10.1177/1059712313482684

Hutto, D & Myin, E. (2018). Going Radical. In A. Newen, L. de Bruin & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition* (pp. 95-115). Oxford University Press.

Jacob, P. (2009) The tuning-fork model of human social cognition: a critique. Conscious. Cogn. 18, 229-243.

Kessler, K. & Rutherford, H. (2010). The two forms of visuospatial perspective-taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology,* 1 (213), 1-12, https://doi.org/10.3389/fpsyg.2010.00213

Kessler, K. & Thomson, L. (2010). The embodied nature of spatial perspective-taking: Embodied Transformation versus sensoriomotor interference. *Cognition*, 114 (1), 72-88, https://doi.org/10.1016/j.cognition.2009.08.015

Keesler, K. & Wang, H. (2012). Spatial perspective-taking is an embodied process, but not for everyone in the same way: Differences predicted by sex and social skills acore. *Spatial cognition & Computation,* 12 (2-3), 133-158, https://doi.org/10.1080/13875868.2011.634533

Kessler, K.; Cao, L.; O'Shea, K.; Wang, H. (2014). A cross-culture, cross-gender comparison of perspective-taking mecanisms. *Proceedings of The Royal Society B: Biological Sciences,* 1-9, https://doi.org/10.1098/rspb.2014.0388

Kozhevnikov, M.; Motes, M.; Rasch, B.; Blajenkova, O. (2006). Perspective-taking vs. Mental Rotation Transformations and How They Predict Spatial Navigation Performance. *Applied Cognitive Psychology*, 20, 397-417, https://doi.org/10.1002/acp.1192

Krueger, J. (2018). Direct Social Perception. In A. Newen, L. de Bruin & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition* (pp. 95-115). Oxford University Press.

Lempers, J. D.; Flavell, E. R. & Flavell, J. H. (1977). The development in very young children of tacit knowledge concerning visual perception. *Genetic Psychology Monographs,* 95 (1), 3-53, https://www.ncbi.nlm.nih.gov/pubmed/849832

Levinson, S. (1996). Language and space. *Annual Review of Anthropology,* 25, 353-382, https://doi.org/10.1146/annurev.anthro.25.1.353

Lizskowski, U. (2018). Emergence of shared references and shared minds in infancy. *Current Opinion in Psychology*, 23*,* 26-29, https://doi.org/10.1016/j.copsyc.2017.11.003

Marmolejo-Ramos, F. & D'Angiulli, A. (2014). Current research topics in embodied social cognition. *Cognitive Processing*, *15*(3), 235-236, https://doi.org/10.1007/s10339-014-0624-2

Michelon, P., Zacks, J. (2006). Two kinds of visual perspective-taking. *Perception & Psychophysics,* 68 (2), 327-337, https://doi.org/10.3758/BF03193680

Mounoud, P. (2001). El desarrollo cognitivo del niño: desde el descubrimiento de Piaget hasta las investigaciones actuales. *Contextos educativos,* 4*,* 53-77, https://dialnet.unirioja.es/servlet/articulo?codigo=209682

Newen, A.; de Bruin, L. & Gallagher, S. (2018). Historical roots of the debate. In A. Newen, L. de Bruin & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition* (pp. 3-15). Oxford University Press.

O'Regan, J., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences,* 24 (5), 939-973, https://doi:10.1017/S0140525X01000115

Proulx, M., Todorov, O., Taylor, A. & de Sousa, A. (2016). Where I am? Who I am? The Relation Beetween Spatial Cognition, Social Cognition and Individual Differences in the Built Environment. *Frontiers in Psychology,* 7 (64), 1-23, https://doi.org/10.3389/fpsyg.2016.00064

Rietveld, E., Denys, D. & Van Westen, M. (2018). Ecological-enactive approach cognition as engaging with a field of relevant affordances. In A. Newen, L. de Bruin & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition* (pp. 41-70). Oxford University Press.

Shapiro, L. & Spaulding, S. (Fall 2021). "Embodied Cognition", *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition), Edward N. Zalta (ed.), forthcoming URL = <https://plato.stanford.edu/archives/fall2021/entries/embodied-cognition/>.

Schurz, M.; Kronbichler, M.; Weissengruber, S. & Surtees, A. (2018). Measuring visual perspective-taking in the brain with avatars and arrows: Which question are we asking? *Neuroimage,* 1 (181), 814-817, https://doi.org/10.1016/j.neuroimage.2018.07.048

Spaulding, S. (2012). Introduction to debates on embodied social cognition. *Phenomenology and the Cognitive Sciences,* 11, 431-448, https://doi.org/10.1007/s11097-012-9275-x

Schurz, M.; Aichhorn, M.; Martin, A. & Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective-taking: a meta-analysis of functional brain imaging studies. *Frontiers in Human Neuroscience*, 7 (712), 1-14, https://doi.org/10.3389/fnhum.2013.00712

Surtees, A.; Apperly, I. & Samson, D. (2013). The use of the embodied self-rotation for visual and spatial perspective-taking. *Frontiers in human neuroscience*, 7 (698), 1-12, https://doi.org/10.3389/fnhum.2013.00698

Tomasello, M. (2008). Origins of Human Communication. MIT press.

Tomasello, M. (2019). The role of roles in uniquely human cognition and sociality. *Journal for Theory of Social Behavior*, 1-18, https://doi.org/10.1111/jtsb.12223

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

Ward, D., Silverman, D., & Villalobos, M. (2017). Introduction: the varieties of enactivism. *Topoi,* 36 (3), 365-375, https://doi.org/10.1007/s11245-017-9484-6

Xie, J., Cheung, H., Shen, M. & Wang, R. (2018). Mental rotation in false belief understanding. *Cognitive Science,* 42, 1179-1206, https://doi.org/10.1111/cogs.12594

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