What Do Theory-of-Mind Tasks Actually Measure? Theory and Practice

Quesque François[1,2] and Yves Rossetti[1,2]

[1]Inserm Unité Mixte de Recherche-Santé (UMR-S) 1028, CNRS UMR 5292, ImpAct, Centre de Recherche en Neurosciences de Lyon, Université Lyon 1, and [2]Plateforme "Mouvement et Handicap" and Plateforme NeuroImmersion, Hôpital Henry-Gabrielle, Hospices Civils de Lyon, Saint-Genis-Laval, France

Abstract

In recent decades, the ability to represent others' mental states (i.e., theory of mind) has gained particular attention in various disciplines ranging from ethology to cognitive neuroscience. Despite the exponentially growing interest, the functional architecture of social cognition is still unclear. In the present review, we argue that not only the vocabulary but also most of the classic measures for theory of mind lack specificity. We examined classic tests used to assess theory of mind and noted that the majority of them do not require the participant to represent another's mental state or, sometimes, any mental state at all. Our review reveals that numerous classic tests measure lower-level processes that do not directly test for theory of mind. We propose that more attention should be paid to methods used in this field of social cognition to improve the understanding of underlying concepts.

Keywords

theory of mind, perspective-taking, empathy, mentalizing, social cognition

Every psychologist and psychiatrist, every child development expert, most cognitive scientists and ethologists, as well as most people interested in consciousness, know what *theory of mind* and *empathy* are. And every contributor to this field of social cognition is able not only to provide a definition for these terms but also to propose specific ways to evaluate their content. Unfortunately, however, definitions and assessments are extremely variable. This variability continues despite the unprecedented interest in social processes over recent decades. Fifty years after the emergence of the first tools designed to measure social-cognitive abilities (Hogan, 1969; Carkhuff & Truax, 1965), the very structure of social cognition still suffers from insufficient clarity (e.g., F. Happé, Cook, & Bird, 2017). One obvious reason for this stems from the highly heterogeneous sources of knowledge in this particular field, which was formed by the confluence of incommensurable approaches such as ethology, psychology and psychiatry, and developmental psychology.

In our view, two main and interacting factors have contributed to the insufficient understanding of the functional architecture of social cognition. The first factor has been noted by several authors in recent years (F. Happé et al., 2017; Quesque & Rossetti, 2019): The vocabulary

for sociocognitive abilities is highly heterogeneous and nonspecific (see Fig. 1 for an illustration). In addition, several terms are used to describe a single concept (convergence of meaning). For example, the "ability to distinguish and represent one's own and others' mental states" can be referred to as "theory of mind" (e.g., Premack & Woodruff, 1978), "mentalizing" (e.g., Frith & Frith, 2012), "mindreading" (e.g., Gallese & Sinigaglia, 2011), "perspective-taking" (e.g., Galinsky, Ku, & Wang, 2005), "empathy" (e.g., Preston & de Waal, 2002), "cognitive empathy" (e.g., Baron-Cohen & Wheelwright, 2004), or "empathic perspective-taking" (e.g., de Waal, 1996) depending on the authors and/or contexts. However, a given term can also be used to depict distinct processes (divergence of meaning). For example, Batson (2009) identified at least nine different psychological constructs that are referred to as "empathy," and more

Corresponding Authors:
Quesque François, Inserm U1028, ImpAct, Centre de Recherche en Neurosciences de Lyon, 16 avenue Lépine, 69676 Bron, France
E-mail: francois.quesque@gmail.com

Yves Rossetti, Inserm U1028, ImpAct, Centre de Recherche en Neurosciences de Lyon, 16 avenue Lépine, 69676 Bron, France
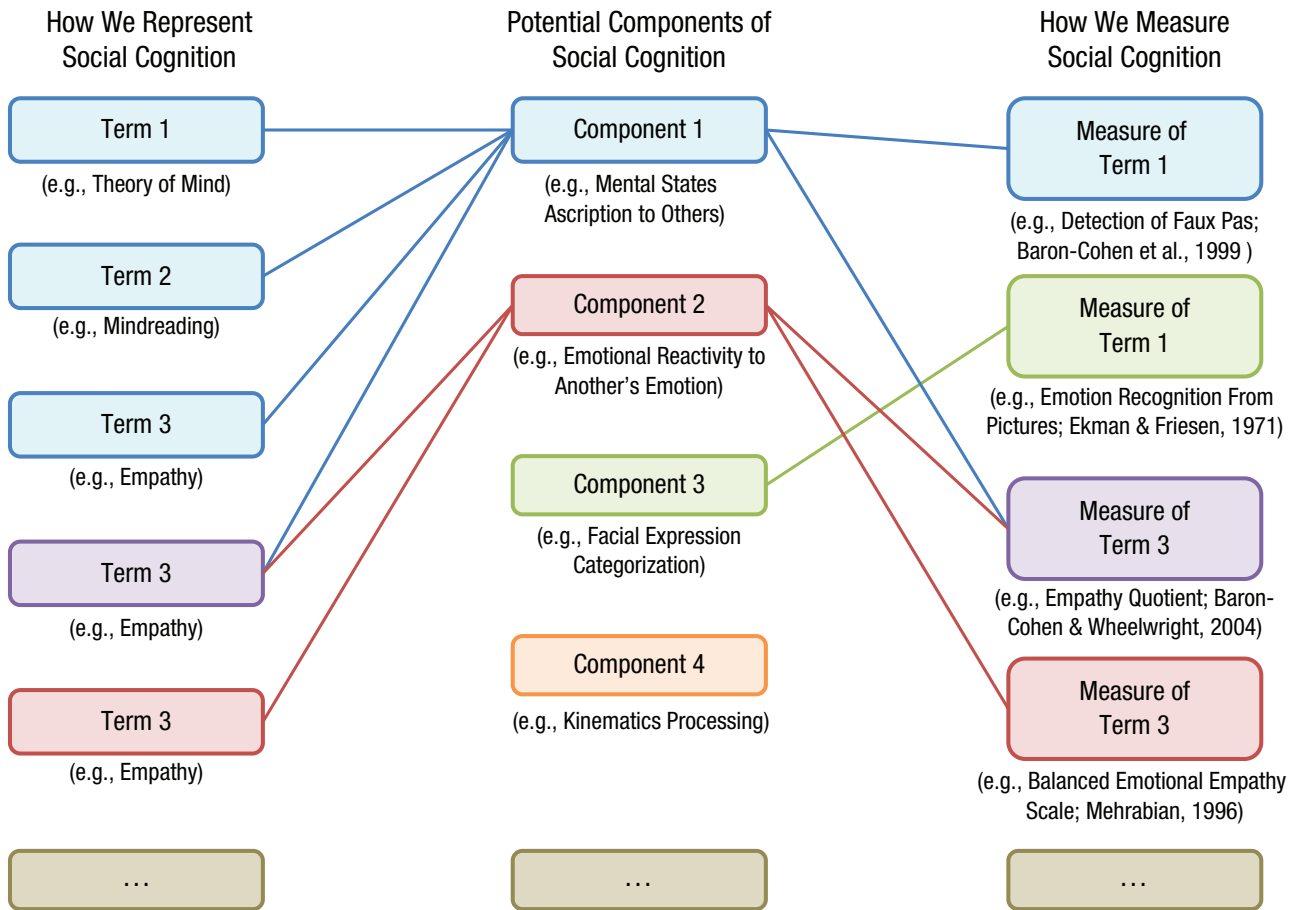E-mail: yves.rossetti@inserm.fr

How We Represent
Social Cognition

Potential Components of
Social Cognition

How We Measure
Social Cognition

Term 1

(e.g., Theory of Mind)

Term 2

(e.g., Mindreading)

Term 3

(e.g., Empathy)

Term 3

(e.g., Empathy)

Term 3

(e.g., Empathy)

...

Component 1

(e.g., Mental States
Ascription to Others)

Component 2

(e.g., Emotional Reactivity to
Another's Emotion)

Component 3

(e.g., Facial Expression
Categorization)

Component 4

(e.g., Kinematics Processing)

...

Measure of
Term 1

(e.g., Detection of Faux Pas;
Baron-Cohen et al., 1999)

Measure of
Term 1

(e.g., Emotion Recognition From
Pictures; Ekman & Friesen, 1971)

Measure of
Term 3

(e.g., Empathy Quotient; Baron-
Cohen & Wheelwright, 2004)

Measure of
Term 3

(e.g., Balanced Emotional Empathy
Scale; Mehrabian, 1996)

...

Fig. 1. Schematic representation of the current heterogeneity and nonspecific aspects in the conceptualization of social cognition and its measures. Heterogeneity: Different terms are currently used to refer to the same theoretical construct (e.g., Terms 1, 2, and 3), and tests that are supposed to measure different constructs actually investigate the same component (e.g., measure of "Term 1" and measure of "Term 3"). Nonspecificity: The same term can be used to refer to distinct constructs (e.g., Terms 3). The same term can also be used to simultaneously include different constructs (e.g., purple Term 3). Tests that are conceived to quantify a particular construct actually measure different components of social cognition (e.g., Measures of Term 1), and some of these tests simultaneously measure different constructs (e.g., purple Measure of Term 3). A nonexhaustive list of examples (in parentheses) illustrates the current heterogeneity and nonspecific aspects of social cognition.

recently, Cuff, Brown, Taylor, and Howat (2016) distinguished 43 different definitions proposed for this term.

The second factor that we identified has received less critical attention and originates from the measures themselves. As is the case for the vocabulary and definitions, it turns out that classic measures of social-cognition mechanisms are also heterogeneous and nonspecific (for an illustration, see Fig. 1). The semantic divergence and convergence described above for terminology also occur at the level of practical evaluation. Obviously, numerous tests coexist to estimate theory of mind (for a review, see Achim, Guitton, Jackson, Boutin, & Monetta, 2013). Some of these tests (e.g., The Reading the Mind in the Eyes test, RMET; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001) are however also frequently used as indexes of empathy (Chapman et al., 2006), emotion decoding (Maurage et al., 2011), or even the precise

ability to read a person's mind through their eyes (Declerck & Bogaert, 2008).

What Is Theory of Mind and How Do We Believe We Measure It?

Among the numerous components of social cognition (Fiske & Taylor, 2013; Goldman & de Vignemont, 2009), some have benefited from privileged attention from scientists. This privileged attention is typically the case for the ability to represent other mental states. Despite the aforementioned terminological heterogeneity, researchers seem to agree on a definition (Apperly, 2012). Theory of mind is classically defined as the ability to impute mental states to oneself and others (Wimmer & Perner, 1983) or the ability to attribute mental states (such as emotions, intentions, or beliefs) to other

persons (Gallese & Sinigaglia, 2011). The term *theory of mind* was originally used to qualify the ability of nonhuman primates to infer other agents' intentions (Premack & Woodruff, 1978). Subsequent studies investigated a wide range of populations (newborns, younger and older infants, adults, as well as numerous other animal species), which led to the development of an important variety of tests and experimental measures. This variety encouraged us to question whether theory of mind depicts a single entity or refers to a large family of abilities in terms of the breadth, homogeneity, and specificity of the functions involved (Apperly, 2012).

Classic definitions suppose that theory of mind includes belief, intention, and emotional inferences (Frith & Frith, 2006). Recent correlational (Erle & Topolinski, 2015; Kanske, Böckler, Trautwein, Parianen Lesemann, & Singer, 2016; Mattan, Rotshtein, & Quinn, 2016) and experimental (Erle & Topolinski, 2017), as well as clinical (Hamilton, Brindley, & Frith, 2009) evidence, however, validates that theory of mind also encompasses the ability to represent how another would perceptually represent the surrounding world.[1] Supporting this idea, early studies reported that the efficient representation of others' false beliefs (e.g., Wimmer & Perner, 1983) and others' visuospatial perspectives (Flavell, Everett, Croft, & Flavell, 1981) emerged around the same age during child development. In addition, it has been possible to identify brain areas (e.g., the dorsal part of the temporo-parietal junction) responsible for representing other perspectives in a domain-general fashion (Aichhorn, Perner, Kronbichler, Staffen, & Ladurner, 2006; Schurz, Aichhorn, Martin, & Perner, 2013; Zaitchik et al., 2010). Integrating these findings, theory of mind would correspond to the general ability to infer others' mental states, regardless of which precise function they support, even if it is possible that different subcomponents of social cognition (kinematics processing, mirroring, stereotypes, etc.) are recruited depending on the type of judgment (emotional, intentional, etc.) and on available stimuli (full body, gaze, verbal information, etc.).

Assuming that theory of mind is conceived of as a unitary process that relies on assorted lower level mechanisms (Gallese & Goldman, 1998; Gangopadhyay & Schilbach, 2012; Rizzolatti & Craighero, 2004), it remains to be determined which aspects are common to all of the relevant types of social inference. According to Epley and Caruso (2009; see also Erle & Topolinski, 2017), all kinds of perspective-taking processes rely on the same set of abilities; they all require the ability to represent mental states that differ from what is directly experienced in the here and now, distinguishing one's own from others' mental states. This ability to corepresent—or to switch between—different perspectives seems to represent the core component of all types of

theory-of-mind judgments. As a practical implication, it would be inappropriate to speak about theory of mind in cases in which there is no evidence for this ability. In accordance, two main criteria should be systemically met by measures of theory of mind. First, a valid assessment of theory of mind should necessitate more than just attributing a mental state to another person. Importantly, it should also imply that the respondents maintain a distinction between the other's mental state and their own (we refer to this as the "nonmerging criterion"). In the particular case of applying theory of mind to the self, the distinction that has to be maintained is between the present and the imagined mental state (for a congruent account concerning the emergence of the ability to pretend, see Leslie, 1987). Although crucial, this is rarely the case in theory-of-mind tasks. Second, lower-level processes (e.g., attention orientation, associative learning) should not possibly account for successful performance on any theory-of-mind task ("mentalizing criterion"; for discussion, see Heyes, 2014). When these simpler processes can provide sufficient explanatory value, one should definitively favor the more parsimonious explanation when interpreting performances. In our view, if a task does not meet these two criteria ("mentalizing" and "nonmerging"), it should no longer be discussed as a measure of theory of mind.

Emotional attribution from others' faces is often used as an index of theory of mind (see Table 1). Success in this type of task may, however, be interpreted as mere visual discrimination (when the task consists of categorizing pictures between different categories) or as emotional contagion (in situations where the same emotional state is shared by the observer). These two cognitive operations also represent sociocognitive mechanisms but certainly should not be regarded as involving theory of mind. It is interesting that such caution is classically evoked when conducting experiments with nonhuman animals. In nonhuman animals, emotion discrimination from selected parts of the human face is interpreted as mere discrimination and not as a manifestation of theory of mind or other higher-level sociocognitive mechanisms (Müller, Schmitt, Barber, & Huber, 2015).

When dealing with humans, scientists sometimes tend to be less parsimonious in their interpretations (e.g., Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997), presumably because we all have naive folk ideas about the way our brains work (e.g., "If I can remember your phone number, then I have a memory," or "If I can recognize your emotion, then I have a theory of mind"). When we see a fish changing direction and following another fish swimming quickly, we do not imagine that it is a manifestation of the follower's intentions (at the

best, we will consider that it learned, by conditioning, that this behavior favors survival). It is striking that when we observe humans' responses to tests, we seem to frequently fall in the trap of less parsimonious interpretations. For example, as noted by Obhi (2012), human performance on two-alternative forced-choice categorization of action kinematics is classically interpreted as evidence for intention reading, whereas such results actually inform us only about humans' visual-discrimination abilities. Obviously, scientists should actively struggle to avoid such interpretation biases. A simple rule to apply would be to systematically consider explanations at the simplest level before considering the involvement of any higher-level cognitive processes.

When one considers the theoretical arguments listed above and the need for parsimonious and unbiased interpretations, it seems of critical importance to verify whether each of the classically used theory of mind tests actually necessitates the ability to switch from an ego-centered perspective. For those tests in which this ability is not required, we may have to redefine what they actually measure. As a first step in that direction, we examined the tests and experimental procedures commonly used to assess theory of mind (see Table 1). For each task, we assessed (a) whether success in that task could be attributed to lower-level processes rather than to a mental state (mentalizing criterion) and, critically, (b) whether the task requires representing a mental state that differs from that of the respondent, implying that the participant needs to distinguish between their own and others' mental states (nonmerging criterion).

What Do We Actually Measure?

What do classic theory-of-mind tasks and tests measure? Table 1 presents the most commonly used tests and tasks for evaluating theory of mind. To underline how tasks that do not meet the two abovementioned criteria differ from tasks that do, here, we arbitrarily focused on two measures. First, as noted by Heyes (2014), when discussing the nonspecificity of most implicit tasks of theory of mind (but see also Kulke, Johannsen, & Rakoczy, 2019; Kulke, Reiß, Krist, & Rakoczy, 2018; Kulke, von Duhn, Schneider, & Rakoczy, 2018; Schuwerk, Priewasser, Sodian, & Perner, 2018 for recent experimental evidence), it is crucial that success in tasks cannot be explained by lower-level processes. A typical example of a task that would not meet this first criterion would be the knowledge-access task (e.g., Povinelli, Nelson, & Boysen, 1990). Participants must choose between two contradictory sources of information (two agents) to determine the location of a hidden

item. Typically, one of the agents attended to the placement of the item, and the other did not. Success in such tasks is sometimes interpreted as evidence for belief ascription (e.g., "this agent knew the actual item's location"), but basic associative learning mechanisms would allow the production of the very same behavior (e.g., "this agent was presented at the same time as the item").

Second, as emphasized earlier, a valid measure of theory of mind should require the participant to represent a mental state that differs from the one experienced by the respondent. A typical example of a task that would not meet this second criterion would be the "ascription of intention from previous rational action" task (e.g., Brunet, Sarfati, Hardy-Baylé, & Decety, 1990). In this task, participants are presented with an open-ended story involving an agent, and they have to select a suitable ending. Again, if success in such tasks can be interpreted as evidence for intention ascription (e.g., "this agent wanted to grasp that item"), there is no evidence that participants distinguish their own intentions from the agent's (e.g., "I now want to grasp that item"). Such merging with others' minds or bodies could be compared with what occurs when we watch movies: We project ourselves onto the character and experience their intentions and emotions at the first-person level, sometimes even losing contact with reality. We may experience the same mental states as the character (interestingly, not the same states as the actor!) and thus may be primed to act in a congruent way, leading us to successfully pass classic tests of theory of mind.

In fact, it seems that evaluations that (a) involve mental state representation and (b) actually require a respondent to distinguish between representations of the self and those of others are not evenly distributed among the different types of mental-state inferences. Some types of judgments are addressed by several tasks that positively meet our nonmerging criterion; for example, this is the case for belief ascription and for level 2 visuospatial perspective-taking (i.e., representing how the world is seen by another person; Flavell et al., 1981). Conversely, there are at least three types of mental-state inferences for which the tasks currently in use suffer from a lack of specificity and do not meet the two abovementioned criteria: visual accessibility judgments, emotion ascription, and intention ascription tasks.

Visual accessibility judgments (i.e., representing what is and what is not visible to another person, without considering how this representation will be perceived), which is also referred to as level 1 visuospatial perspective-taking (Flavell et al., 1981), is typically estimated through tasks that are independent from another person's frame of reference (mentalizing criteria). Yaniv and

Table 1. Descriptions of the Different Tests and Experimental Tasks Used to Estimate Theory of Mind

Classic tasks used to measure theory of mind	Task description	Necessitates representing mental states? *(Mentalizing criterion)*	Necessitates distinguishing one's own/others' mental states? *(Nonmerging criterion)*	Perspective of the respondent
Detection of "faux- pas" (e.g., Baron-Cohen, O'Riordan, Stone, Jones, & Plaisted, 1999)	Detect if a person made a "faux-pas" in a conversation.	Yes	Yes	Third person
Detection of deceptive intentions from kinematics (e.g., Sebanz & Shiffrar, 2009)	Categorize the intentions (deceptive or not) of another person from kinematic information. These tasks are classically operationalized using forced-choice categorization.	No	No	Second or third person
Detection of others' thoughts (e.g., Privilege knowledges; Keysar, 1994)	Predict how a naive recipient would interpret an ambiguous message. Participants have access to privileged knowledge, and it is clear for them that the message is intended to be sarcastic, but this privileged information is not available to the message recipient.	Yes	Yes	Third person
Emotion recognition from pictures (e.g., Ekman & Friesen, 1971)	Infer the emotions of other people from their faces. These tasks are classically operationalized using forced-choice categorization.	No	No	Third Person
Emotion recognition from voices (e.g., RMVT; Golan, Baron-Cohen, Hill, & Rutherford, 2007)	Infer the emotions of other people from their voices. These tasks are classically operationalized using forced-choice categorization.	No	No	Third person
False belief attribution (e.g., The Sally & Ann task; Wimmer & Perner, 1983)	Infer the belief of a person who has a false belief about a particular scene (which is not the case of the participants who have an updated view of that scene).	Yes	Yes	Third person
Inference of spatial orientation (e.g., Hegarty & Waller, 2004)	Participants are presented with a bird's-eye view of a scene that includes several objects and are asked to place one of these objects in its actual location in a second (rotated) view of the scene, by projecting themselves into the central object.	Yes	Yes	Third person
Intention ascription from movie (e.g., Premack & Woodruff, 1978)	Initially implemented with nonhuman animals, participants are presented with a movie in which an actor unsuccessfully tries to perform an action. Different objects are displayed near the participant, the idea being to test if the participant will choose an object that would allow the actor to successfully perform the action.	Yes	No	Third person
Interactive scene description (e.g., The director task; Wu & Keysar, 2007)	Follow the instructions given by a person that does not share the same visual experience of the ambiguous scene of interest.	Yes	Yes	Second person

Table 1. (Continued)

Classic tasks used to measure theory of mind	Task description	Necessitates representing mental states? *(Mentalizing criterion)*	Necessitates distinguishing one's own/others' mental states? *(Nonmerging criterion)*	Perspective of the respondent
Knowledge access task (e.g., Povinelli, Nelson, & Boysen, 1990)	Initially implemented with nonhuman animals, participants have to choose between two sources of information (two experimenters) about the location of hidden food; one of the experimenters was present during the placement of the food, and the other was absent.	No	No	Second person
Level 1 representation of another's visual experience (e.g., Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010)	Determine, as fast as possible, the number of dots present in a room in which an agent is standing who may or may not share the same perspective as the participants. An increase in decision time when the number of dots visually accessible through the perspectives of the participant and the agent is incongruent is interpreted as proof for spontaneous visual perspective-taking.	No	No	Third person
Level 2 representation of another's visual experience (Piaget & Inhelder, 1956)	Represent or describe how a particular scene would look from another person's point of view.	Yes	Yes	Second or third person
Mental state inferences from stories (e.g., Strange stories; F. G. Happé, 1994)	Provide context-appropriate mental state explanations for a character's behaviors.	Yes	Yes	Third person
Mental state ascription from ecological movie scenes of social interaction (e.g., MASC; Dziobek et al., 2006)	Infer the mental state (feelings, thoughts, and intentions) that drives a movie character's behaviors.	Yes	Yes	Third person
Mental state attribution from animated shapes (e.g., Heider & Simmel, 1944)	Participants are presented with short animated movies of several geometrical shapes and are asked to describe the scenes they assisted. In some cases, participants are instructed to think about what the shapes are doing and thinking; in others, no additional instruction is given.	No	No	Third person
Mental state attribution from face pictures (e.g., RMET; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001)	Participants have to infer the mental state (emotional and intentional) of other persons from their faces (or their gaze). These tasks are classically operationalized using forced-choice categorization.	No	No	Third person
Motor intention ascription from previous rational action (e.g., Brunet, Sarfati, Hardy-Baylé, & Decety, 2000)	Participants are presented with an open-ended movie or comic strip and have to imagine how the character would act at the end.	Yes	No	Third person

Table 1. (Continued)

Classic tasks used to measure theory of mind	Task description	Necessitates representing mental states? *(Mentalizing criterion)*	Necessitates distinguishing one's own/others' mental states? *(Nonmerging criterion)*	Perspective of the respondent
Social intention ascription from kinematics (e.g., Lewkowicz, Quesque, Coello, & Delevoye-Turrell, 2015)	Participants must identify the intention that drives another person's actions from kinematic information. These tasks are classically operationalized using forced-choice categorization.	No	No	Second or third person
Visual accessibility judgments (e.g., Masangkay et al., 1974)	Participants are asked to judge what is visible (and what is not) from another person's point of view.	No	No	Second or third person
Scene description (e.g., Quesque, Chabanat, & Rossetti, 2018)	Participants are asked to describe an ambiguous component (e.g., a number that can be seen as a 6 or a 9) of a visual scene that contains another agent. This test measures the spontaneous use of the other agent's perspective, in that this agent is nonrelevant for task completion and is not mentioned in the instructions.	No		Third person
Social Spatial Compatibility (e.g., Freundlieb, Kovács, & Sebanz, 2016)	The participant and a partner perform a simple stimulus-response compatibility task sitting at a 90° angle to each other. If participants adopt the partner's visuospatial perspective, a spatial compatibility effect should be observed on their reaction time.	No	Possibly[a]	Third Person
Spontaneous influence of a bystander's beliefs on the decision-making process (e.g., Kovács, Téglás, & Endress, 2010)	Participants have to make perceptual decisions in the presence of an agent that may (or may not) hold false beliefs about the response they provide. An increase in decision time when the agent holds a belief that is incongruent with the participant's is interpreted as evidence for spontaneous belief ascription.	No		Third person
Spontaneous influence of a character's beliefs on anticipatory looking behavior (e.g., Surian & Geraci, 2012)	Participants passively attend to an animated scene in which a character may (or may not) hold a false belief about an object's location. Using eye-tracking technology, differences in the orientation of anticipatory looking behaviors in conditions in which the character holds true or false beliefs is classically interpreted as a marker of theory of mind.	No		Third person

[a]Because no instruction to consider the other agent perspective is given, these tests do not necessarily require to distinguish our own mental state from that of another. These tests are considered as a measure of how spontaneously people would consider others' visuospatial perspectives and not as a measure of how accurate or difficult this judgment is. When participants endorse the perspective of the agent in their response, researchers classically interpret this behavior as a form of theory of mind. However, it is possible that when responding in that way, participants do not distinguish between others' and their own mental states (this effect could be conceived as the visuospatial equivalent of emotional contagion). It is, however, worth noting that in some tasks (e.g., Kovács, Téglás, & Endress, 2010), comparing trials in which the agent has congruent or incongruent beliefs could provide evidence for theory of mind, as it does for responses using "double perspective" in other tasks (e.g., Quesque, Chabanat, & Rossetti, 2018).

Shatz (1990) proposed that computing the line of sight of another agent is analogous to actually drawing a line from the agent to the target object. As a consequence, visual accessibility tasks have been parsimoniously described as relying predominantly on egocentric processes (Kessler & Rutherford, 2010).

Emotion ascription also suffers from the same problem. The majority of the tasks exploring emotional ascription require recognizing emotions, or merely categorizing them, from facial expressions, voices, and animations. Such tasks are likely to assess lower-level processes such as perceptual emotion recognition rather than genuine theory-of-mind abilities (mentalizing criteria). A critical test for this interpretation has actually been conducted by comparing the performances of clinical populations known to present specific impairment in theory of mind or emotion recognition on the Reading the Mind in the Eyes test (RMET; Baron-Cohen et al., 2001), which is the most used test of theory of mind for emotional judgments. Compatible with our current interpretation, the results suggested that the RMET measures emotion recognition rather than theory-of-mind ability (Oakley, Brewer, Bird, & Catmur, 2016).

Finally, most *intention-ascription* tasks (and some *emotion-ascription* tasks) also present an important limit because they do not require the distinction between one's own and others' mental states (nonmerging criterion). Success in these tasks may be obtained on the mere basis of mirroring processes such as motor contagion, which would, in fact, involve a merging between representations of the self and others (Brass, Ruby, & Spengler, 2009). Because no distinction is made between the observer's own mental state and the character's mental state, it is rather unwise to assume that we ascribe a particular mental state to the character, and we should consequently avoid referring to "theory of mind" in this context.

A Necessary Shift: What We Need to Change Moving Forward

In this last section, we discuss the changes that could be made to overcome the current lack of specificity in many "tests of theory of mind," as well as their conceptual and theoretical benefits. First, we will see how the general call for more ecological validity when studying social processes (Schilbach et al., 2013) would address many of the presently raised issues. Second, we will examine how the suggested paradigm shift would encourage terminological clarity in social cognition. Finally, we will review how the use of the mentalizing and nonmerging criteria would allow the conciliation of findings that may appear contradictory.

In recent years, several researchers have called for a shift in the methods used to investigate social cognition, supporting an approach based on actual interactions and emotional engagements between people rather than mere observation (e.g., Schilbach et al., 2013). This strategy is obviously at odds with classic paradigms in which the participants are presented with written or verbal stories, puppets, comic strips, or movies (i.e., always from a third-person, or outsider's, perspective). The initial motivation for a shift toward second-person perspective studies originates from the idea that social cognition is fundamentally different when we are directly engaged with another person compared with when we remain an external observer (Gallotti & Frith, 2013). For example, recent studies demonstrated that when we are involved in an interaction with another person, we spontaneously represent the motor affordances of the surrounding environment from their perspective, which is not the case when observing a passive partner (Coello, Quesque, Gigliotti, Ott, & Bruyelle, 2018; Freundlieb, Kovács, & Sebanz, 2016).

In the present case, one extremely important outcome of the recommendation to examine first-person engagement in social interactions is that such a paradigm shift would also allow the aforementioned limits (e.g., lack of specificity, distinction between the mental states of the self and others) of most classic tests of theory of mind to be overcome. As underlined by Barsalou (2013), our social interactions require significantly more complementary actions than mirrored actions. When facing a character expressing anger, most participants experience fear (not anger). When facing a character throwing a ball at them, participants are primed to catch (not to throw) the ball. Therefore, their own mental state differs from that of the observed character, even though they will have correctly inferred their emotion or intention. In addition, directly involving participants in tasks would constitute a means to limit alternative lower-level explanations (e.g., motor contagion) to participants' performances, in addition to enhancing ecological validity. As a representative example, the director task, used by Wu and Keysar (2007), requires participants to interpret the message (e.g., "give me the big book") of a partner who has a different point of view (e.g., only two books are visually accessible to the partner, whereas a third book that is even bigger can be seen only from the participants' perspective) and to act accordingly. In this case, participants should not only represent the point of view of another person but also distinguish between what they see and what the partner sees. This uncommon feature for level 1 visuospatial perspective-taking tests allows for an efficient exclusion of low-level interpretations of participants' performance.
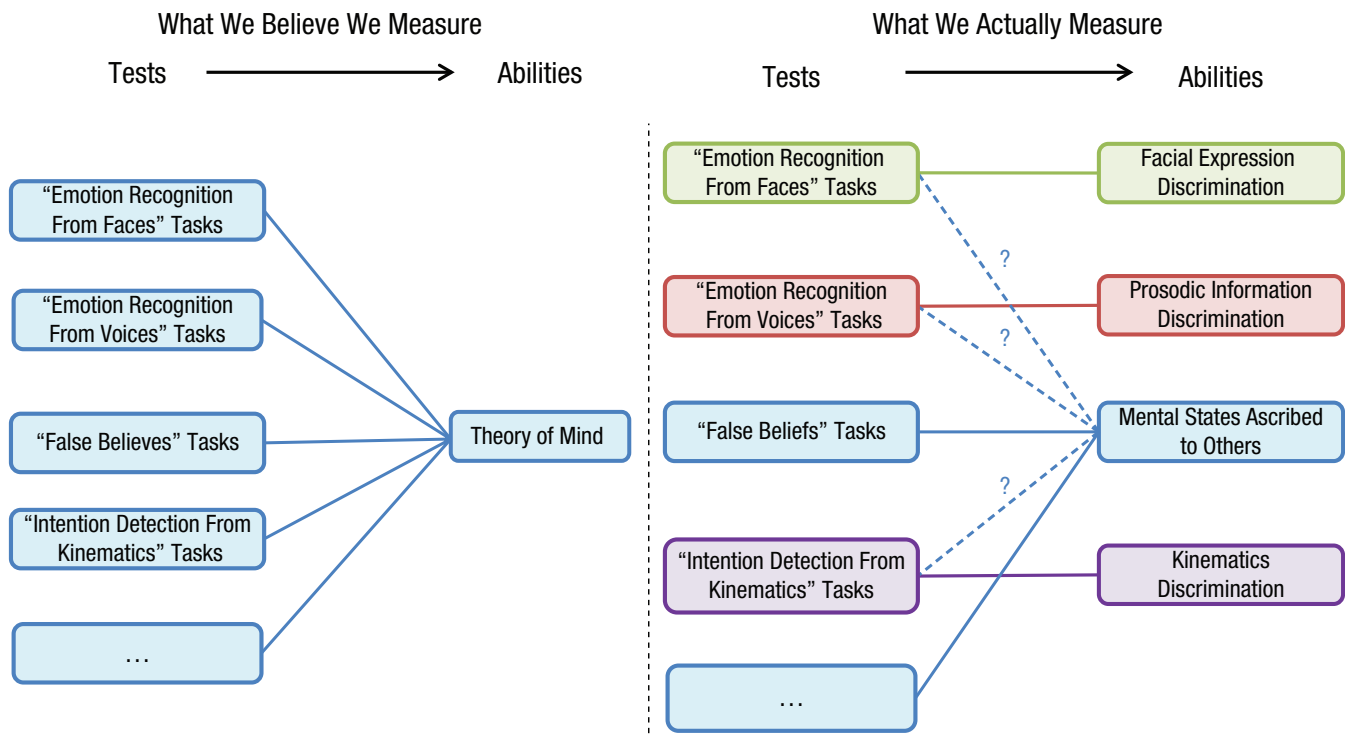
What We Believe We Measure

Tests ⟶ Abilities

What We Actually Measure

Tests ⟶ Abilities

Fig. 2. Illustration of the fact that most classic tasks used to measure theory of mind actually quantify lower-level cognitive processes.

Participants' first-person engagement is not the only strategy one can rely on, as long as the test involves distinguishing between the participant's and the character's mental states. Other tests in which participants are mere observers of a social scene also meet the mentalizing and nonmerging criteria (e.g., the false-belief task; Wimmer & Perner, 1983). As previously underlined, not all types of mental-state inferences benefit from such tests, but the same logic can be virtually transferred to any type of mental-state inference. This was, for example, the case of the MASC (movie for the assessment of social cognition; Dziobek et al., 2006), in which participants have to infer the mental states (both emotional and cognitive) that drive a character's actions within a complex social interaction, in movie scenes displaying multiple agents. To our knowledge, the MASC seems to represent the only available test that allows an assessment of the inference of others' emotions, excluding alternative lower-level accounts (such as visual or auditory categorization). Careful attention should be paid to address this issue in future test development.

Regardless of the precise strategies chosen to address the presently discussed criteria, we argue that an important theoretical shift is needed for the designers of clinical and experimental measures of theory of mind. The crucial point is that tasks aimed at estimating any aspect of theory of mind should minimally ensure that participants distinguish between their own and others'

mental states. This point is especially true when participants experience a mental state similar to that of the stimulus character (e.g., when facing a big spider with my partner, I know that both of us are scared, but I also know that each of us has our own qualitative and quantitative experience of fear).

It is likely that numerous classic tests of theory of mind measure lower-level social-cognitive processes such as kinematics processing (see Obhi, 2012), social attention (see Heyes, 2014), emotion recognition (e.g., Oakley et al., 2016), or even prosodic information discriminations rather than theory of mind abilities (see Fig. 2). Although this route may turn out to be challenging, especially for tests with a long-standing tradition of being associated with theory of mind, tasks that do not meet the mentalizing and the nonmerging criteria should no longer be considered valid assessments for theory of mind (see Table 1 for an evaluation of each task regarding the mentalizing and nonmerging criteria). A long-term consequence of this change will be whether the concept of "theory of mind" will survive in its current operational fuzziness.

The suggested paradigm shift is in line with the urgent need for conceptual clarification in the field. As we emphasized earlier, two main efforts will be required to develop a general model of the structure of social cognition, which may be necessary for this field to be considered a unitary domain of science. The first level

is terminological, and the second is methodological. Clarity and consensus in the field of social cognition cannot arise without pruning ambiguities and confusion at both the theoretical and the practical levels of this scientific area. Specific hierarchical organizations can be postulated (e.g., "theory of mind" involves "emotion categorization," which relies on "face processing," which requires "social attention"), but in the absence of sufficiently specific evaluations, no conclusive argument should be drawn. By determining more strictly which tests actually measure theory of mind and which tests do not, a clearer outline of theory of mind will be delineated. Therefore, the paradigmatic and conceptual levels of clarification inherently and dialectically depend on each other.

An old (Ford, 1979; Kurdek, 1978; Underwood & Moore, 1982) but still unsolved question is whether there is a general mechanism supporting the different types of theory-of-mind judgments (e.g., "beliefs ascription," "emotion ascription") or whether different independent constructs coexist and support each type of inference. Current experimental evidence is available in support of both hypotheses (Bons et al., 2013; Cook, Brewer, Shah, & Bird, 2013; Erle & Topolinski, 2015, 2017; Hamilton et al., 2009; Kanske et al., 2016; Mattan et al., 2016; Maurage et al., 2016; Shamay-Tsoory & Aharon-Peretz, 2007). Unfortunately, the arguments collected by a variety of authors are based on the use of a heterogeneous set of evaluations, which is responsible for the incommensurability and confusion. Careful selection within the existing tests for theory of mind, associated with high-levels of caution concerning mentalizing and nonmerging criteria in the development of new tasks, would allow the reconciliation of findings that may appear contradictory (e.g., evidence supporting both the presence and the absence of theory of mind in a given animal species). It has been recently argued that the involvement of a common mechanism for all types of theory-of-mind judgments could be consistent with the existence of apparent double dissociations between different types of inferences (Quesque & Rossetti, 2019).

Finally, from an ontogenetic point of view, refining the tasks that provide actual measures of theory of mind will also help clarify the extensive developmental variability across the different types of mental states' inferences (Quesque & Rossetti, 2019). This preliminary step will enable a more accurate view of the actual development of theory-of-mind abilities and the definition of more precise stages in this development.

In the above paragraphs, we have seen that the systematic use of mentalizing and nonmerging criteria to determine whether a task is a valid measure of theory of mind would provide many benefits. At the conceptual level, this paradigm shift is consistent with the need for terminological clarification, whereas at the theoretical level, this pruning would allow us to clarify a currently divided body of scientific literature. These considerations prompt scientists in the field, both authors and reviewers, to systematically assess whether methodological choices allow us to elaborate on acceptable discussions of theory of mind abilities. Most disagreements in the field are likely to stem from the insufficient attention given to this methodological dimension, resulting in overgeneralized interpretations. It is at the level of interpretation, rather than fact, that these disagreements take place, and reunifying experimental findings with legitimate interpretations is an open door to unifying the field.

ORCID iD

Quesque François ![iD] https://orcid.org/0000-0001-7954-8316

Note

1. It is worth noting that all of these studies focused on visuospatial perspective-taking judgments. We, however, have no reason to believe that inferences for other sensory modalities would not rely on the same core components (for a discussion, see Quesque & Rossetti, 2019).

References

Achim, A. M., Guitton, M., Jackson, P. L., Boutin, A., & Monetta, L. (2013). On what ground do we mentalize? Characteristics of current tasks and sources of information that contribute to mentalizing judgments. *Psychological Assessment*, *25*, 117–126.

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

Apperly, I. A. (2012). What is "theory of mind"? Concepts, cognitive processes and individual differences. *The Quarterly Journal of Experimental Psychology*, *65*, 825–839.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry*, *38*, 813–822.

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders*, *29*, 407–418.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*, 163–175.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*, 241–251.

Barsalou, L. W. (2013). Mirroring as pattern completion inferences within situated conceptualizations. *Cortex*, *49*, 2951–2953.

Batson, C. D. (2009). These things called empathy: Eight related but distinct phenomena. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3–15). Cambridge, MA: The MIT Press.

Bons, D., van den Broek, E., Scheepers, F., Herpers, P., Rommelse, N., & Buitelaaar, J. K. (2013). Motor, emotional, and cognitive empathy in children and adolescents with autism spectrum disorder and conduct disorder. *Journal of Abnormal Child Psychology*, *41*, 425–443.

Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 2359–2367.

Brunet, E., Sarfati, Y., Hardy-Baylé, M. C., & Decety, J. (2000). A PET investigation of the attribution of intentions with a nonverbal task. *NeuroImage*, *11*, 157–166.

Carkhuff, R. R., & Truax, C. B. (1965). Training in counseling and psychotherapy: An evaluation of an integrated didactic and experiential approach. *Journal of Consulting Psychology*, *29*, 333–336. doi:10.1037/h0022187.

Chapman, E., Baron-Cohen, S., Auyeung, B., Knickmeyer, R., Taylor, K., & Hackett, G. (2006). Fetal testosterone and empathy: Evidence from the empathy quotient (EQ) and the "reading the mind in the eyes" test. *Social Neuroscience*, *1*, 135–148.

Coello, Y., Quesque, F., Gigliotti, M. F., Ott, L., & Bruyelle, J. L. (2018). Idiosyncratic representation of peripersonal space depends on the success of one's own motor actions, but also the successful actions of others! *PLOS ONE*, *13*(5), Article e0196874. doi:10.1371/journal.pone.0196874

Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, not autism, predicts poor recognition of emotional facial expressions. *Psychological Science*, *24*, 723–732.

Cuff, B. M., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review*, *8*, 144–153. doi:10.1177/1754073914558466

Declerck, C. H., & Bogaert, S. (2008). Social value orientation: Related to empathy and the ability to read the mind in the eyes. *The Journal of Social Psychology*, *148*, 711–726.

de Waal, F. B. M. (1996). *Good natured: The origins of right and wrong in humans and other animals*. London, England: Harvard University Press.

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., . . . Convit, A. (2006). Introducing MASC: A movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders*, *36*, 623–636.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*, 124–129.

Epley, N., & Caruso, E. M. (2009). Perspective taking: Misstepping into others' shoes. In K. D. Markman, W. M. P. Klein, & J. A. Suhr (Eds.), *Handbook of imagination and mental simulation* (pp. 295–309). New York, NY: Psychology Press.

Erle, T. M., & Topolinski, S. (2015). Spatial and empathic perspective-taking correlate on a dispositional level. *Social Cognition*, *33*, 187–210. doi:0.1521/soco.2015.33.3.187

Erle, T. M., & Topolinski, S. (2017). The grounded nature of psychological perspective-taking. *Journal of Personality and Social Psychology*, *112*, 683–695. doi:10.1037/pspa0000081

Fiske, S. T., & Taylor, S. E. (2013). *Social cognition: From brains to culture*. London, England: Sage.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology*, *17*, 99–103. doi:10.1037/0012-1649.17.1.99

Ford, M. E. (1979). The construct validity of egocentrism. *Psychological Bulletin*, *86*, 1169–1188.

Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2016). When do humans spontaneously adopt another's visuospatial perspective? *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 401–412. doi:10.1037/xhp0000153.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*, 531–534.

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, *63*, 287–313.

Galinsky, A. D., Ku, G., & Wang, C. S. (2005). Perspective-taking and self-other overlap: Fostering social bonds and facilitating social coordination. *Group Processes & Intergroup Relations*, *8*, 109–124.

Gallese, V., & Goldman, A. I. (1998). Mirror neurons and the simulation theory of mindreading. *Trends in Cognitive Sciences*, *2*, 493–551.

Gallese, V., & Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends in Cognitive Sciences*, *15*, 512–519.

Gallotti, M., & Frith, C. D. (2013). Social cognition in the we-mode. *Trends in Cognitive Sciences*, *17*, 160–165.

Gangopadhyay, N., & Schilbach, L. (2012). Seeing minds: A neurophilosophical investigation of the role of perception-action coupling in social perception. *Social Neuroscience*, *7*, 410–423.

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. D. (2007). The 'Reading the Mind in the Voice' test-revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders*, *37*, 1096–1106.

Goldman, A., & de Vignemont, F. (2009). Is social cognition embodied? *Trends in Cognitive Sciences*, *13*, 154–159.

Hamilton, A. F. C. D., Brindley, R., & Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition*, *113*, 37–44.

Happé, F., Cook, J. L., & Bird, G. (2017). The structure of social cognition: In (ter) dependence of sociocognitive processes. *Annual Review of Psychology*, *68*, 243–267.

Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*, 129–154.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, *32*, 175–191.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*, 243–259.

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science*, *9*, 131–143.

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, *33*, 307–316.

Kanske, P., Böckler, A., Trautwein, F. M., Parianen Lesemann, F. H., & Singer, T. (2016). Are strong empathizers better mentalizers? Evidence for independence and interaction between the routes of social cognition. *Social Cognitive and Affective Neuroscience*, *11*, 1383–1392.

Kessler, K., & Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology*, *1*, Article 213. doi:10.3389/fpsyg.2019.02996

Keysar, B. (1994). The illusory transparency of intention: Linguistic perspective taking in text. *Cognitive Psychology*, *26*, 165–208.

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

Kulke, L., Johannsen, J., & Rakoczy, H. (2019). Why can some implicit Theory of Mind tasks be replicated and others cannot? A test of mentalizing versus submentalizing accounts. *PLOS ONE*, *14*, Article e0213772. doi:10.1371/journal.pone.0213772

Kulke, L., Reiß, M., Krist, H., & Rakoczy, H. (2018). How robust are anticipatory looking measures of theory of mind? Replication attempts across the life span. *Cognitive Development*, *46*, 97–111.

Kulke, L., von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. *Psychological Science*, *29*, 888–900.

Kurdek, L. A. (1978). Relationship between cognitive perspective taking and teachers' ratings of children's classroom behavior in grades one through four. *The Journal of Genetic Psychology*, *132*, 21–27.

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review*, *94*, 412–426. doi:0.1037/0033-295X.94.4.412

Lewkowicz, D., Quesque, F., Coello, Y., & Delevoye-Turrell, Y. N. (2015). Individual differences in reading social

intentions from motor deviants. *Frontiers in Psychology*, *6*, Article 1175. doi:10.3389/fpsyg.2015.01175.

Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., & Flavell, J. H. (1974). The early development of inferences about the visual percepts of others. *Child Development*, *45*(2), 357–366.

Mattan, B. D., Rotshtein, P., & Quinn, K. A. (2016). Empathy and visual perspective-taking performance. *Cognitive Neuroscience*, *7*, 170–181.

Maurage, P., D'hondt, F., Timary, P., Mary, C., Franck, N., & Peyroux, E. (2016). Dissociating affective and cognitive theory of mind in recently detoxified alcohol-dependent individuals. *Alcoholism: Clinical and Experimental Research*, *40*, 1926–1934.

Maurage, P., Grynberg, D., Noël, X., Joassin, F., Hanak, C., Verbanck, P., . . . Philippot, P. (2011). The "Reading the Mind in the Eyes" test as a new way to explore complex emotions decoding in alcohol dependence. *Psychiatry Research*, *190*, 375–378.

Mehrabian, A. (1996). Manual for the Balanced Emotional Empathy Scale (BEES). Monterey, CA: Author.

Müller, C. A., Schmitt, K., Barber, A. L., & Huber, L. (2015). Dogs can discriminate emotional expressions of human faces. *Current Biology*, *25*, 601–605.

Oakley, B. F., Brewer, R., Bird, G., & Catmur, C. (2016). Theory of mind is not theory of emotion: A cautionary note on the Reading the Mind in the Eyes Test. *Journal of Abnormal Psychology*, *125*, 818–823. doi:10.1037/abn0000182

Obhi, S. (2012). The amazing capacity to read intentions from movement kinematics. *Frontiers in Human Neuroscience*, *6*, Article 162. doi:10.3389/fnhum.2012.00162

Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London, England: Routledge & Kegan Paul.

Povinelli, D. J., Nelson, K. E., & Boysen, S. T. (1990). Inferences about guessing and knowing by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *104*(3), 203–210.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*, 515–526.

Preston, S. D., & de Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral & Brain Sciences*, *25*, 1–20.

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

Quesque, F., & Rossetti, Y. (2019). Unifying perspectives on the ability to represent others' mental states. Manuscript submitted for publication.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192.

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

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral & Brain Sciences*, *36*, 393–414.

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*, Article 712. doi:10.3389/fnhum.2013.00712

Schuwerk, T., Priewasser, B., Sodian, B., & Perner, J. (2018). The robustness and generalizability of findings on spontaneous false belief sensitivity: A replication attempt. *Royal Society Open Science*, *5*(5), Article 172273. doi:10.1098/rsos.172273

Sebanz, N., & Shiffrar, M. (2009). Detecting deception in a bluffing body: The role of expertise. *Psychonomic Bulletin & Review*, *16*, 170–175.

Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia, 45*, 3054–3067.

Surian, L., & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *British Journal of Developmental Psychology*, *30*, 30–44.

Underwood, B., & Moore, B. (1982). Perspective-taking and altruism. *Psychological Bulletin*, *91*, 143–173. doi:10.1037/0033-2909.91.1.143

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science*, *18*, 600–606. doi:10.1111/j.1467-9280.2007.01946.x

Yaniv, I., & Shatz, M. (1990). Heuristics of reasoning and analogy in children's visual perspective taking. *Child Development*, *61*, 1491–1501.

Zaitchik, D., Walker, C., Miller, S., LaViolette, P., Feczko, E., & Dickerson, B. C. (2010). Mental state attribution and the temporoparietal junction: An fMRI study comparing belief, emotion, and perception. *Neuropsychologia*, *48*, 2528–2536.