REBUILDING THE LANDSCAPE OF PSYCHOLOGICAL UNDERSTANDING AFTER THE MINDREADING WAR

abstract

‘Mindreading war’ refers here to the intricate net of connected debates both in the philosophy and the cognitive sciences concerning the onset, the development, and the nature of the cognitive mechanisms underlying mindreading – i.e., the alleged ability to attribute mental states to predict and explain others’ behavior. The mindreading war has lasted for almost forty years by now with apparently no winners or losers. This article argues that the present stalemate results from the lack of initial theoretical discussion about foundational issues that led to the conflict. Recovering the dialogue between psychologists and philosophers is necessary if we are to start rebuilding the landscape of psychological understanding once this long war is over.

keywords

folk psychology, mindreading, social cognition, theory of mind
1. Introduction

Mindreading, or Theory of Mind (ToM), is usually defined as the ability to attribute mental states (e.g., beliefs, desires, and intentions) to others to predict and explain their behavior. By ‘mindreading war’, I refer here to the intricate network of connected debates both in the philosophy and the cognitive sciences concerning the onset and the development of this capacity in infancy and early childhood as well as the description of the cognitive mechanisms underlying it.

The mindreading war has lasted for almost forty years by now with apparently no winners or losers. In psychology, the confrontation between the proponents of nativist and constructivist accounts persists to the present day, while dual-system theorists of social cognition have also joined the discussion more recently. In philosophy, the debate has dissolved in a plethora of proposals addressing a range of different issues with unclear consequences on the empirical debate (see Apperly, 2008, 2009; Stich & Nichols, 1997).

This article surveys the outcomes of the conflict but first clarifies its onset and development. Accordingly, it describes the start of the war in section 2, the empirical debate in psychology in section 3, and the first and second theoretical confrontation in philosophy in sections 4 and 5, respectively. Finally, section 6 tries to clarify the relevance of each of these phases to the wider context of the mindreading war. A clear understanding of the foundational issues that led to the mindreading war is necessary if we are to start rebuilding the landscape of psychological understanding once this long war is over.

2. The Casus Belli and the Start of the Mindreading War

In 1978, two comparative psychologists devised a procedure to assess whether a subject would understand the intention behind an observed action (Premack & Woodruff, 1978). They found that a chimpanzee was extremely reliable in passing the task, and concluded that chimpanzees have a Theory of Mind (ToM) – i.e., a theoretical capacity to infer the mental states of other agents, and to exploit these attributions to predict others’ actions.

Philosophers, who at that time were debating about the meaning and the scientific relevance of mental state terms, acknowledged the validity of the procedure to assess a subject’s understanding of mental states but argued that the proposed experiment did not test the possession of the concept most distinctive of human psychological understanding, that is, belief (Bennett, 1978; Dennett, 1978; Harman, 1978).

---

1 I wish to thank Jay Garfield for useful comments on a previous version of this article.
Developmental psychologists rose to the challenge, and responded a few years later by devising a new procedure, the false belief test (Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983), which requires subjects to predict an actor’s behavior that crucially depends on her possession of a false belief. Strikingly, they found that children below age four could not pass the task, which suggested to them that younger children lacked a proper concept of belief as well as full-fledged ToM (Wellman, Cross, & Watson, 2001).

The earliest results from the false belief test started a debate among developmental psychologists, which has developed exponentially. On the one hand, advocates of modularism interpreted the results as demonstrating that children are endowed with a ToM module (ToMM) – i.e., a cognitive mechanism specific for social cognition – that has been shaped through natural selection, and is underpinned by dedicated neural processes. Some of them initially proposed that ToMM reaches full maturation right when children pass the false belief test (Baron-Cohen, 1995). Over time, however, modularists have converged on the claim that ToMM matures in early infancy (Baillargeon, Scott, & He, 2010; Leslie, 2005), and that children younger than four struggle with the false belief test because of limited computational resources (Bloom & German, 2000).

On the other hand, advocates of constructivism claimed that children progressively construct a ToM through a process of hypothesis testing and revision that is akin to the construction of a scientific theory (Gopnik & Meltzoff, 1996; Gopnik & Wellman, 1994; Perner, 1991; Wellman, 1990; see also Perner, Huemer, & Leahy, 2015 for a very recent proposal). Accordingly, they argued that younger children fail the false belief test because they lack a full-fledged concept of belief, which is acquired only after age four.

Such a vibrant confrontation has become even more intense after the findings that even infants manifest a sensitivity to others’ (false) beliefs when assessed by measuring different behavioral indices (e.g., anticipatory and preferential gazing, fixation time) rather than by direct questioning (see Baillargeon, Scott, & Bian, 2016 for a review). According to modularists, these results from “spontaneous-response” – as opposed to traditional “elicited-response” – false belief tasks demonstrate the very early onset of the ToM mechanism while constructivists have replied by explaining the same data in the terms of basic capacities to form an expectation about observed actions requiring no attribution of mental states (Perner & Ruffman, 2005; Rakoczy, 2012; Ruffman, 2014; see also Banovsky, 2016; Fenici, 2014; Fenici & Zawidzki, 2016).

The debate is still far from being settled. It has been given new impetus by the appearance of dual-system theorists on the battle field (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). In agreement with modularists, they contend that infants are endowed with socio-cognitive abilities that are specific to the processing of others’ mental states – although they initially track what another agent has seen – i.e., her perceptual states – rather than what she believes. In agreement with constructivists, however, they claim that these basic abilities are significantly limited, and that the mature possession of ToM depends on the later emergence of distinct forms of sophisticated mindreading after age four.

3. The First Phase of the Conflict: The Theory-of-Mind Debate in Psychology
4. The Second Phase of the Conflict: The Mindreading Debate in Philosophy

While psychologists argued about the onset and the development of ToM through the ontogeny, some philosophers noticed that both modularists and constructivists had each assumed that the capacity to attribute mental states relies on either tacit (modularism) or explicit (constructivism) knowledge of the principles of folk psychology, and thereby had endorsed a theoretical account about ToM – what they called the Theory Theory (TT) of the mind. In contrast, these partisans of the Simulation Theory (ST) argued that the attribution of mental states – or mindreading, as they re-labelled it – has a practical nature, and depends on the capacity to simulate another agent by putting oneself “in her shoes”.

Some of these scholars focused on the personal-level features of the simulatory mechanism (Heal, 1986), others on the subpersonal mechanism itself (Gallese & Goldman, 1998; Goldman, 2006; Gordon, 1986). Simulation theorists also disagreed about whether simulation recruits the simulator’s introspection capacities. In general, however, they all agreed that attributing mental states to an agent to predict her behavior proceeds in three steps: (i) the simulator projects herself in the place of the agent, and simulates the mental states she would have if she were there; (ii) the simulator practically decides what she would do if she had the mental states she pretends to have; (iii) the simulator interrupts the simulation, and projects on the agent the intention to act according to what she planned to do during the process.

A novel conflict then arose (see Carruthers & Smith, 1996; Davies & Stone, 1995 for a review), with some philosophers taking up the arms against their colleagues, and in defence of some of the traditional “theoretical” solutions to the ToM debate in psychology (Botterill, 1996). Progressively, the two sides acknowledged that there are compelling cases favoring each of the contenders, and that people may attribute mental states through both theorizing and simulation depending on the context – although each argued that their proposed strategy is the default (e.g., Carruthers, 1996; Goldman, 2006). In this context, proponents of “hybrid” accounts also introduced combined cognitive architectures implementing mindreading through a mix of theorizing and simulation (Currie & Ravenscroft, 2004; Meltzoff, 2002, 2005; Mitchell, Currie, & Ziegler, 2009; Nichols & Stich, 2003). The contenders seemed to having agreed to a truce but the spark for a novel conflict was smouldering under the ashes, and peace was far from definitive.

5. The Third Phase of the Conflict: The Debate About Social Cognition in Philosophy

The memory of the clash between the supporters of TT and ST was still alive when a novel movement of opposition rose among philosophers. Several dissenters rebelled against the traditional parties arguing that both TT and ST misconstrued the role of mindreading in social interaction. These Interaction Theorists (ITs), as we may call them, did not regroup around a central claim but all gave a prominent role to social interaction (possibly with no attribution of mental states) in the explanation of traditional cases of mindreading.

Some supporters of IT criticized the idea – shared by TT and ST – that mental states can be only inferred indirectly from the observation of behavior (Gallagher, 2008, 2011; Krueger & Overgaard, 2012; Overgaard, 2015; Zahavi, 2008, 2011). Inspired by phenomenology (Scheler, 1954, p. 260; Husserl, 1982, p. 51), or by Wittgenstein (see Overgaard & Zahavi, 2009 for an analysis), they claimed instead that mental states can be the object of direct perception so that mindreading requires neither theorizing nor simulation.

Other ITs argued that action understanding often depends on the practical knowledge of the social context in which it occurs. Accordingly, many cases of everyday social interaction do not really require mindreading (Gallagher, 2001, 2012; Gallagher & Hutto, 2008; Hutto, 2017). Some also endorsed the complementary narrative practice hypothesis (NPH), according to which genuine mindreading (whenever it occurs) involves a narrative capacity to evaluate and discuss one’s reasons for acting but significantly does not require either simulation or theorizing (Hutto, 2004, 2008).
Finally, some ITs affiliated to enactivism argued that social cognition – as every case of cognitive activity – is essentially constituted by the dynamic processes of interaction between the body and the environment (both physical and social) (Fuchs & De Jaegher, 2009; De Jaegher, 2009; McGann & De Jaegher, 2009). Consequently, they criticized both TT and ST for defining mindreading as thespectatorial capacity of an observer that looks at other agents from a detached (“third-person”) perspective. In contrast, they claimed, our engagement with others is mostly embedded in a social (“second-person”) interactional context.

After decades of trench warfare with increasingly sophisticated theoretical weaponry deployed, it seems that no winners or losers have emerged. In psychology, modularists, constructivists, and dual-system theorists still battle to gain supremacy in the field. Their debate is far from being settled, however, with each party bringing new evidence to bear favoring one view or another. Importantly, little attention is given to how much the dispute is actually empirical and how much it depends on the theoretical framework that psychologists implicitly assume to account for human development (Overton, 2015).

In philosophy, the confrontation between TT and ST have ended without apparent resolution while IT have also occupied an area in the debate. These factions seem to have divided the field in different zones of influence, each patrolling her own area without challenging their rivals. Importantly, little of the philosophical discussion between TT, ST, and IT seems to influence the empirical debate in psychology.

I argue that the present stalemate results from the lack of initial theoretical discussion (both in psychology and philosophy) about foundational issues at the origin of the debate. The earliest studies employing the false belief test paradigm implicitly assumed that there must be a ToM competence that is possible to test. By acknowledging the validity of these procedures in assessing one’s understanding of belief, and moving to discuss its most proper description, most of philosophers in the initial TT-ST debate corroborated the idea that we mostly make sense of others’ actions by tacitly attributing mental states. Accordingly, ToM or mindreading quickly turned from an operational construct related to an experimental procedure into a reified cognitive capacity responsible for most of our social interaction. As such, it was then immediately available as an object for (diverging) psychological and philosophical inquiries. The more recent discussion in philosophy has however shown that there are important reasons to question the initial setting of the ToM debate. If, as ITs have argued, many cases of everyday social interaction may not require at all the attribution of mental states, we should reconsider the **explanandum** – not only the **explanans** – of (either spontaneous- or elicited-response) false belief tests. Because mindreading may be not so pervasive in everyday social interaction, developmental psychologists may have got it wrong in assuming that infants and children must attribute mental states in order to succeed in these tasks (see, e.g., Andrews, 2012, pp. 22-34; de Bruin & Newen, 2012, p. 254).

This consideration also suggests that we reconsider the plausibility of mindreading (or ToM) as a unified cognitive capacity. If, as ITs have also argued, action prediction depends on a variety of cognitive mechanisms, mindreading (or ToM) may not exist as a unified cognitive capacity (Fenici, 2012, 2017; Garfield, Peterson, & Perry, 2001; Hutto, 2008). In contrast, we may attribute mental states to others more as a way to explain and rationalize their (past) actions rather than to predict their (future) ones (McGeer, 2007).

If this analysis is at least partially correct, psychologists and philosophers would greatly benefit from an armistice conference – a collective reassessment of the foundational issues that led to the mindreading war. I conjecture that recovering the dialogue between the two disciplines will help restructuring the present landscape of psychological understanding, and enter a new era of peaceful successful collaboration between them.
I conclude offering reflection about four different issues that, I believe, require additional theoretical clarification, and should attract the new generations of philosophers for the more general benefit of the scientific community.

1) **The role of philosophical theories of content within the empirical debate.** As I noted in section 3, psychologists are still debating whether the infant’s selective response to an agent’s (false) beliefs in spontaneous-response false belief tasks really demonstrates that, in their second year, infants understand the concept of belief – with proponents of modularist accounts of ToM claiming that it does, and advocates of constructivist accounts claiming that it does not (see Fenici, 2015 for discussion). Reaching two opposite conclusions, Buckner (2014) and Hutto (2017) have argued that the issue is partially decided by what theory of mental content one assumes. Philosophers should make explicit how some differences among the opposed accounts in the ToM debate among psychologists might actually depend on their different philosophical assumptions about representational concepts.

2) **The traditional debate between TT vs. ST.** It is surprising that neither TT nor ST have taken advantage of the vast amount of new evidence brought forward by psychologists to decide the empirical ToM debate. To some, this might indicate that, despite its appearance, the TT–ST debate is underdetermined by the experimental data because it aims to provide a high-level, functional description of mindreading capacities that can be accommodated with any empirical evidence (Heal, 1994; Stich & Nichols, 1997). To others, it might indicate that even more data are required before settling the issue. Whatever is the explanation, philosophers should clarify why the traditional debate about mindreading is so silent when its empirical counterpart looks so flourishing.

3) **The silent invasion of IT.** The clash of IT on the traditional TT–ST debate lacked significant consequences. In some cases, IT has been criticized to be generic or implausible (e.g., Bohl, 2015; Spaulding, 2015). In many other cases, it has advanced its claims without outcry. In general, IT is by now perceived as another proposal in the field aside TT and ST. Still, IT purports to be a radical alternative to both of them. Philosophers should indicate what among the three is the most plausible candidate, or whether some forms of co-existence are also possible.

4) **The significance of the pluralist solution.** The latest discussion in the philosophical debate saw the appearance of a wide notion of pluralism about folk psychology. On this view, being a folk psychologist depends on having the capacity to deploy a variety of strategies to make sense of others’ actions that include (but are not limited to) the abilities to attribute mental states, behavioral dispositions and traits, social roles and stereotypes as well as the capacity to embed observed actions within social norms and scripts (Andrews, 2012; Fiebich & Coltheart, 2015; Maibom, 2007). It remains unclear whether these are all specific forms of psychological understanding or whether they are part of some more general understanding of human action. Philosophers should clarify the boundaries (if any) of our psychological comprehension, and whether they constitute a domain different from the general comprehension of our own form of life.
REFERENCES


