

Inge Wertwijn, S1033944

# Mindreading

Truth or Dare?

Date: 25/08/2020

Wordcount: 10391 ex notes, glossary and biography

Caspar David Friedrich, Wanderer über dem Nebelmeer

# CONTENTS

<b>1</b>	<b>EXPLANANDUM: WHAT IS MINDREADING?</b>	<b>3</b>
<b>2</b>	<b>EXPLANANS: THEORETICAL LANDSCAPE</b>	<b>4</b>
2.1	The traditional account	4
2.1.1	Theory-theory	5
2.1.2	Simulation theory	6
2.1.3	Hybrid accounts	8
2.2	The phenomenology-inspired account	9
2.2.1	Interaction theory	10
2.2.2	Enactive theories	10
<b>3</b>	<b>WHY MINDREAD? COMMON ASSUMPTIONS</b>	<b>11</b>
3.1	Actions are belief-motivated	12
3.2	Humans compete for success	13
3.3	Mindreading is in the brain	14
3.4	Mindreading is in our genes	16
3.5	Taking stock	17
<b>4</b>	<b>MINDREADING: BUILDING ON TRUST</b>	<b>18</b>
4.1	Mindreading and sharing	18
4.2	Trust and cognitive offloading	19
4.3	Trust, reliability and justified belief	20
4.4	Traits and stereotypes	21
4.5	Sensemaking and feedback mechanisms	21
4.6	Sketch for a modified approach	22
4.6.1	General principles	22
4.6.2	Implications	23
<b>5</b>	<b>CONCLUSION</b>	<b>24</b>
<b>6</b>	<b>GLOSSARY</b>	<b>25</b>
<b>7</b>	<b>BIOGRAPHY</b>	<b>28</b>

## ABSTRACT

*Mindreading* allows us to attribute mental states such as thoughts, feelings and intentions to each other and to ourselves. Without this ability, we think ourselves incapable of social interaction. The philosophical debate around *mindreading* started up in the last century and has largely concentrated on how *mindreading* is supposed to work. In recent decades, more and more data from experiments in neurology and psychology have become available. This is an exciting development, because it allows for the empirical testing of philosophical ideas. Some philosophical accounts of *mindreading* are not borne out or are even contradicted by empirical evidence.

Why do we mindread? This is the central question of this paper. Philosophers have largely neglected this question by concentrating on other questions which Goldman (2006) helpfully put together as wish-list for a theory of *mindreading*. In the first half of this paper, the main theories of *mindreading* are plotted against this desideratum to show to what extent these questions have been considered; and by implication, to show what has not been considered, namely why we mindread. From this overview emerges the second part of this paper. It seems that philosophical accounts of how we mindread are built on questionable foundations or intuitive assumptions which have not been proven. This causes definitional problems which turn *mindreading* into a so-called *wicked problem*.

The last section is devoted to a sketch of a possible answer to the “why” question. Perhaps we can capture the essence of *mindreading* by looking at the smallest common denominator of mindreading theories. It is proposed that this essence is the establishment of *trust* between trustor and trustee, in combination with some kind of monitoring mechanism. *Trust* is fallible and therefore potentially risky for the trustor. Yet it also allows for cognitive off-loading. *Trust* allows us automate our responses, saving precious time when in danger and freeing up cognitive resources for other tasks.

## 1 EXPLANANDUM: WHAT IS MINDREADING?

---

We regard *mindreading* as a practice available to all of us, whatever our intelligence, age, health or education. We all mindread, all of the time, and it comes as naturally as breathing. That is why *mindreading* and *folk psychology* are often used interchangeably.

Yet the topic of *mindreading* has also been enmeshed with fundamental philosophical questions<sup>1</sup>. From a metaphysical perspective, we ask what are minds and mental states are, what they are made of, and whether mental states and physical things are made of the same stuff. From an epistemological perspective, we wonder how we recognise other minds, or how we obtain privileged access to our own mental states. We may even wonder to what extent our knowledge of others or self is true knowledge, in the sense of a justified or rational belief. This leads us into further questions of normativity and regulation. In spite of this, Goldman (2006)<sup>2</sup> has listed seven questions which a comprehensive theory of *mindreading* should have an answer to.

1. How do we attribute mental states to each other?
2. How do people mindread themselves?
3. How is the *mindreading* capacity, or skill, acquired?

---

<sup>1</sup> Goldman, 2006, pp. 4–10

<sup>2</sup> Goldman, 2006, p. 21. The original questions are abbreviated here. Goldman himself, like most philosophers at that time, concentrated on the first four. Nichols & Stich, 2003, p. 77 also list criteria, but these are less comprehensive, focusing on *mindreading* accuracy in adults, children en persons with autism.

4. What is the contents of mental states? do people conceive the difference between belief and desire, anger and disgust?
5. What is the cognitive architecture of mindreading and how does it fit with other cognitive domains?
6. What is the relationship between mindreading and other forms of social cognition, such as empathy and imitation? How is it related to clinical problems such as autism?
7. What is the evolutionary story behind mindreading?

Let's regard these questions as a desideratum and use it to map out theories that purport to explain mindreading. The idea is not to provide a complete or detailed overview of theories, nor to evaluate or criticise, but to chart the extent to which various accounts of mindreading present a comprehensive theory of mindreading. We will see that they do, but we will also see that philosophers have been focussed on *what* and *when* and *where* and *how* but have not been so hot on *why*<sup>3</sup>.

## 2 EXPLANANS: THEORETICAL LANDSCAPE

---

The interest in mindreading started to emerge around the 1970s but did not appear out of the blue<sup>4</sup>. In *Philosophia*, these were the heydays of [Functionalism](#), the doctrine that holds that a mental state is defined not by its internal constitution but by its function. Hence, "folk psychology is the theory that gives ordinary mental state terms their meaning"<sup>5</sup>. Across other disciplines, researchers worked towards an empirically verifiable, scientific theory. Evolutionary biologists were already claiming that mindreading would enhance evolutionary fitness. Evolutionary psychologists researched how the ability to mindread evolved. Anthropologists and cross-cultural psychologists investigated the trade-off between mindreading practices and cultural upbringing. Clinical psychologists tried to improve the mind-reading abilities of people with autism. Cognitive psychologists traced the development of mindreading in childhood.

Two families of accounts have emerged from decades of philosophical debate: the traditional account and what I will call the the phenomenology-inspired account.

### 2.1 THE TRADITIONAL ACCOUNT

On the traditional account, one person ascribes mental states to another person in a kind of observational stance. Social interaction or context does not play much of a role. Both Theory-Theory (TT) and Simulation-Theory (ST) belong to this family. The claim is that we do this to explain and predict the behaviour of other people. We are generally quite successful at this, which may be taken as an indication that our folk psychology is both important and true<sup>6</sup>. Yet there are also many social-psychological experiments showing how often we are wrong about others and ourselves. As Morton<sup>7</sup> put it: "we seem to make many and systemic mistakes".

---

<sup>3</sup> Inspired by the lyrics of Gethsemane by Rice & Lloyd Webber, 1970

<sup>4</sup> Nichols & Stich, 2003, pp. 2–5; Morton, 2009, p. 3. In analytic circles, folk psychology did not immediately become popular. In 1984, when I took my finals at Oxford, the subject did not figure in the curriculum at all.

<sup>5</sup> Nichols & Stich, 2003, p. 7

<sup>6</sup> Dennett, 1987 chapter 3, Folk Psychology as a Source of Theory

<sup>7</sup> Morton, 2009, p. 14. "We think that people are more likely to repeat patterns of behaviour than in fact they are (Nisbett and Ross 1980); we underestimate how much people's opinions are affected by their social situations (Festinger 1964); we are often completely wrong about the reasons for which we have made choices

On the traditional account, it is generally assumed that folk psychology ‘happens’ in the brain. This is because of its functionalist roots. [Functionalism](#) allows for multiple realizability, which means that every mental state is identical with some brain state - without claiming that any specific mental state is to be equated with a specific brain state. On the representational theory of mind, mental processes like reasoning, imagining and thinking are thought of as sequences of intentional mental states. The computational theory of mind take a further step: mental processes are computations, rule governed sequences that can be evaluated semantically<sup>8</sup>.

Equating mental states with brain states introduces some new questions<sup>9</sup> which correspond to [question 5](#) of Goldman’s desideratum on page 4:

- How is folk psychology represented in the brain? As a language, or a network or something else?
- What brain areas are involved in mindreading?
- Does the brain have a special module for mindreading?

### 2.1.1 Theory-theory

Theory-theory (TT) assumes that we have a working hypothesis of how the mind works. An early formulation is by Premack and Woodruff<sup>10</sup>, who famously wrote: “An individual has a theory of mind if he imputes mental states to himself and others. A system of inferences of this kind is properly viewed as a theory because such states are not directly observable, and the system can be used to make predictions about the behavior of others”. The system of inferences is a layman’s psychological framework, consisting of concepts, laws and rules; in short, a theory. We interpret the behaviour of other people by applying this theory to our observations. This corresponds to [question 1](#) of Goldman’s desideratum.

Philosophers have also theorised how, by what method, we do this interpretation. Often, the focus has been on the attribution of [propositional attitudes](#). Davidson and Dennett both assumed that propositional attitude ascription is a [rational](#)<sup>11</sup> process. Davidson called it [sense-making](#)<sup>12</sup> and postulated that the contents of mental states is determined through a triangular structure requiring interaction between two individuals and the world. Dennett proposes his theory of the [intentional stance](#)<sup>13</sup>. To be rational, which we strive for according to Dennett, is to take the intentional stance.

Feelings and emotions are not easily accounted for in TT. [Privileged access](#) corresponds to [question 2](#) of Goldman’s desideratum and poses a problem for TT. It seems unlikely that we interpret our own mental states in a third-person inference process. Feeling pain simply does not feel like making an

---

(Nisbett and Wilson 1977; Kornblith 1989; Gopnik 1993); and we overestimate the accuracy of people's memory for details of events they have experienced (Loftus 1979; Conway 1997)”.

<sup>8</sup> Pitt, 2000

<sup>9</sup> See: Ravenscroft, 2019. He lists a number of questions, most of which overlap with Goldman (2006), see paragraph 1, Explanandum: what is mindreading? The remaining questions are about the connection between the brain and mindreading, here rephrased.

<sup>10</sup> Premack & Woodruff, 1978

<sup>11</sup> Goldman (2006, page 4, 54) regards rational explanation of mindreading as a separate branch from theory-theory accounts, possibly because its normative character and because it is restricted to propositional attitudes.

<sup>12</sup> Davidson, 2001, pp. 182–183. His theory involves the “semantic stance” and is highly dependent on understanding language.

<sup>13</sup> Dennett, 1987 chapter 2, paragraph ‘The Intentional Strategy and How It Works’

inference. Goldman<sup>14</sup> summarises other, more diluted, formulations of [functionalism](#) which do allow for immediate self-knowledge, as a kind of reporting, rather than inferencing.

How does TT develop? This is [question 3](#) of Goldman's desideratum. One idea is the theory of the *child scientist*, which holds that children build up this impressive capacity by working on their theory of mind "virtually all his waking hours"<sup>15</sup>. This capacity is assumed to be domain-general, dependent on general capacities for learning and reasoning. A very different possibility is the postulation of an *inborn capacity*. This family of theories was developed by out of the observation that autistic children's ability to mindread is impaired. It provides an answer to both [question 6](#) and [question 7](#) of Goldman's desideratum because it postulates<sup>16</sup> an innate module that is part of our genetic constitution, dedicated to a domain-specific mindreading function. Apperly and Butterfill<sup>17</sup> have yet another kind of account which combines the advantages of both explanations. They postulate a an early developing, fast and relatively inflexible system-1 and a later developing, flexible and slow system-2 to account for the stages in which young children pass false-belief tasks. Their account is reminiscent of Goldman's low-level and high-level mindreading in paragraph 2.1.2. It is disliked by Carruthers<sup>18</sup> who prefers the simpler option of postulating one system that is enriched over time.

TT assumes that the attribution of mental states is dependent on internally represented knowledge, which provides a partial answer to [question 4](#) of Goldman's desideratum. But where does this content come from? Some philosophers assume a kind of *folk- or analytic functionalism* i.e. beliefs, desires and intentions can be understood in terms of their causal role between input from the environment and the output of observable behaviour. Goldman<sup>19</sup> dubs this *representational functionalism* because it is not an ontological theory, but a psychological theory of how mental-state attribution actually works. It does not matter whether the mindreading person is competent, i.e. whether these causal relations actually exist. What matters is that this knowledge of causal relations is used in ascribing mental states.

The structural organisation of the theory-theory itself relates to both [question 4](#) and [question 5](#) of Goldman's desideratum. As Stich and Nichols<sup>20</sup> point out, there are a number of ways to fill the "information" box central to TT. It may be sentence-like and rule based. It may be like the mental models championed by Johnson-Lard. Or a connectionist network. It may contain law-like generalisations or just rules of thumb, It may even resemble a scientific theory. Botterill<sup>21</sup> considers TT to be a fruitful framework for considering the relations between folk and scientific psychologies. He conceives of TT as consisting of a Lakatos-style core based on a small number of principles. This is the information-bearing system which is expandable to specific situations through auxiliary hypotheses<sup>22</sup>.

### **2.1.2 Simulation theory**

Simulation theory says that we do not rely on a theory of how the mind works. We understand the mental states of other people directly, by putting ourselves in other people's shoes, and imagining how we would feel and behave in their situation. The ST answer to [question 1](#) of Goldman's

---

<sup>14</sup> Goldman, 2006, pp. 25–26

<sup>15</sup> Gopnik & Wellman, 1992

<sup>16</sup> Scholl & Leslie, 1999

<sup>17</sup> Apperly & Butterfill, 2009

<sup>18</sup> Carruthers, 2016

<sup>19</sup> Goldman, 2013, pp. 139–144

<sup>20</sup> Stich & Nichols, 1995

<sup>21</sup> Botterill & Carruthers, 1999, p. 12

<sup>22</sup> Botterill, 1996

desideratum is therefore: when we perceive a mental state in someone else, we generate an equivalent state in ourselves, and “understand” the other from our own simulation.

On the introspective view, ST is like TT. It assumes a third-person perspective from which inferences are made about the mental states of the other person, even though no theory is required. Mental concepts are understood prior to simulation; introspection and self-monitoring supply the rest. This creates the same problem with [privileged access](#) as with TT, posing a problem in answering [question 2](#) of Goldman’s desideratum. We simply do not feel that we interpret when ascribing a mental state to ourselves. Gordon<sup>23</sup> has offered the notion of an *ascent routine* to solve this problem. The idea is this: when we self-ascribe a mental state, we might say either that that ‘p’ or that we ‘believe that p’. It makes no difference. There is no inference, no rationalisation, no introspection. A similar argument holds for other kinds of mental states, Gordon<sup>24</sup> argues. If, at the ice cream parlour, I express my liking for a particular flavour of ice-cream, I report my liking. I don’t think about my likes or dislikes, I simply choose from what is available.

Simulation is often associated with [Empathy](#), because it is by inner imitation that we gain knowledge about other minds. Quine even thought them equivalent<sup>25</sup>. The discovery of mirror neurons is often cited as support for ST, thereby providing an answer to [question 5](#) of Goldman’s desideratum. When one monkey observes another to grasp something, the same set of neurons is activated in both the acting and the observing monkey’s brain. Mirror neurons exist in human brains too. However, there appears to be no connection between imitation by non-human primates and mirror-neurons, which is why Gallese and Goldman<sup>26</sup> assign mirror neurons only a precursory role in mindreading. Goldman<sup>27</sup> expands this point of view in later writings, explaining that mirroring processes in themselves do not constitute mindreading. He distinguishes<sup>28</sup> between two levels of mindreading. High-level mindreading is driven by imagination or pretence; low-level mindreading is largely driven by automatic and unconscious mental mimicry. He argues that mirror neurons play a part in some low-level mindreading, but that the bulk of mindreading cannot be explained by mirroring.

Goldman’s distinction between low- and high level mindreading also feature in what Barlassina and Gordon<sup>29</sup> have dubbed his *three-stage model*, which basically says that mental simulation does not constitute a mindreading event, but only causally contributes: the *causation view*. It is opposed to the *constitution view*, of which Gordon and his notion of ascent routines, is a proponent. The constitution view is that mindreading is constituted, not caused, by simulation. If, for example, I experience distaste when I see from my son’s expression that he dislikes his courgettes, my mental state represents his distaste. This causation-constitution controversy in the ST camp represents a wide gulf when we try to determine the contents of mental states, i.e. answer [question 4](#) of Goldman’s desideratum. On the constitution view, the contents of mental states comes from the experience of mindreading; on the causal view, it is not clear what constitutes a mental state. Is it a mental representation? Shea<sup>30</sup> points out that it is a mistake to confuse the causal role of mental representations with their meaning, just like honeybees do not need to understand what they are doing; they only need a causal disposition to

---

<sup>23</sup> Gordon, 1996

<sup>24</sup> Gordon, 2007

<sup>25</sup> Quine, 1992, p. 46

<sup>26</sup> Gallese & Goldman, 1998

<sup>27</sup> Goldman, 2009

<sup>28</sup> Goldman, 2006, chapters 6 and 7

<sup>29</sup> Barlassina & Gordon, 2017

<sup>30</sup> Shea, 2018, p. 37

fly off to a specific location. Goldman<sup>31</sup> himself speaks of a type of embodied enactment, or enactment imagination, a *pretend* mental state.

ST has its own answer to [question 3](#) of Goldman's desideratum. As we saw earlier, the *child scientist* version of TT derived support from the false-belief experiments showing children becoming proficient at attributing false beliefs by the age of four. However, later research by Baillargeon et al<sup>32</sup> showed that a non-verbal version of the false-belief test can be passed by 7-month old infants. Goldman and Jordon account for this early onset by arguing that these infants are sensitive to the false beliefs of others but do not represent those beliefs. They simply imagine the world from the other point of view without forming a kind of judgement – a kind of primitive psychological competence not yet developed into fully-fledged mindreading ability.

On the ST account, it is to be expected that learning to mimic, imitate and role-playing, contribute to the development of simulation abilities. This answers [question 6](#) of Goldman's desideratum. Autism is regarded as involving an impaired simulation ability, which is supported by evidence that autistic children are deficient in role playing from an early age<sup>33</sup>.

The distinction that Goldman has drawn between high level and low level mindreading calls for two separate evolutionary accounts in answer to [question 7](#) of his desideratum. High-level mindreading depends on perspective taking which Goldman equates with enactment imagination. Intrapersonal mindreading may well have developed out of visual and motor imagination<sup>34</sup>, but the leap from intra- to interpersonal mindreading is not easily accounted for. Goldman briefly speculates that the capacity for perspective taking may have evolved from mirror-neuron activity in primates, when animals who observe each other happen to share the same goal. However, on his own theory, mirror-neurons are only supposed to be involved in low-level automatic processes. He has much more to say about the evolution of low-level mindreading, based on adaptive value of emotion-recognition<sup>35</sup>. He discusses two emotions: disgust and anger. Disgust is a part of a general gut defence mechanisms that is found across all mammals. Showing disgust has obvious adaptive value because recognition of this emotion may save others from also ingesting toxic food. Similarly for anger. Anger is associated with defense and fight responses. Recognising anger has evolutionary adaptive value because it allows for the recognition of danger. Goldman claims that we humans recognise emotions through facial expressions. This is an evolutionary adaptation in itself that allows for emotions to become *contagious*. The contagion-mechanism is then repurposed, made into an *exaptation*, because it is there. This then becomes low-level mindreading.

### 2.1.3 Hybrid accounts

Over the years, TT and ST accounts have reconfigured themselves into hybrid accounts. The general consensus seems to be that there are several kinds of mindreading, some more ST-like, others more TT like.

In the TT camp, Nichols and Stich<sup>36</sup> famously produced their hybrid account which includes a simulation process for inference prediction. They assume the existence of an inference mechanism which

---

<sup>31</sup> Goldman, 2006, pp. 48; 220; Goldman & Jordan, 2013

<sup>32</sup> Baillargeon, Scott, & Bian, 2016

<sup>33</sup> Goldman, 2006, pp. 192–197; 200–206

<sup>34</sup> Goldman, 2006, p. 220, quoting Currie (1995a) and Dennett (1978c)

<sup>35</sup> Goldman, 2006, p. 219, quoting Sripada & Goldman (2005) and Gould & Vrba (1982)

<sup>36</sup> Nichols & Stich, 2003, p. 135



runs on representations in the *Possible World Box* (PWB). This PWB is filled with both ‘pretend’ beliefs about what the other person believe, and beliefs about what the other desires. Then another part of the proposed cognitive architecture comes into play: the Planner, a mechanism for determining how to achieve a goal. This Planner is used to predict what the other person might do, based not on the mindreader’s own beliefs, but on the content of the PWB.

As Goldman puts it: “Simulation and mirroring are not equivalent; mirroring is just one species of simulation. Hence, if a type of mental state is not readable by mirroring, it is still possible it can be read by simulating, just a different form of simulating. It is also possible, of course, that it can be read by theorizing, and I do not wish to deny that some acts of mindreading, partly or wholly, consist of theorizing”<sup>37</sup>. In particular, Goldman stipulates the need for a quarantine mechanism which helps to keeps real and simulated mental states apart. Comparing different mental states requires “some theorising”<sup>38</sup>, he says, that is, not simulation.

More hybrid accounts exist, just as there are more variations on [Theory-Theory](#) and [Simulation-Theory](#) accounts than discussed in this paper. The point is that combining elements of both accounts into one does not increase the explanatory value of the [traditional account](#) per se. Both TT and ST already supply answers to Goldman’s desideratum of seven questions.

## 2.2 THE PHENOMENOLOGY-INSPIRED ACCOUNT

In the 21<sup>st</sup> century a different kind of approach emerged which can be traced back to Gallagher’s proposal for [social cognition: interaction theory](#) (IT). IT opposes the traditional account of TT and ST and instead offers a new account. Gallagher<sup>39</sup> challenges four suppositions that underlie traditional accounts of mindreading:

- Hidden minds: we have no direct access to the mental states of other people. Therefore we need some kind of extra cognitive perception process, such as mindreading, so we can understand the mental states of others by inference or simulation.
- Mindreading as default: our mindreading practices are ubiquitous. We do it all the time and in every social situation, and it is our number 1 method for understanding others.
- Observational stance: we observe others from a third-person point of view.
- Methodological individualism: a mindreading process is not shared, it is confined to the individual who does the mindreading.

Ratcliffe<sup>40</sup> takes a further step. If mindreading really were central to our social life, he says, the traditional account leaves much out that is central to that social life, namely: the role of situational understanding, how we perceive intentionality, the effects of interaction on interpersonal understanding and our bodily responsiveness to others. In short, where Gallagher accuses *mindreading* of claiming too much, Ratcliffe adds that *mindreading* also explains too little.

The phenomenology-inspired account focusses on personal experience rather than on third person observation. The idea is that we do not normally ascribe mental states such as emotions of intentions through inference. Rather, we perceive them directly in our social contact with others, through their actions, gestures and expression. Underlying this approach is the [4E conception of the mind](#): mental

---

<sup>37</sup> Goldman, 2009, p. 324

<sup>38</sup> Goldman, 2006, p. 170

<sup>39</sup> Gallagher, 2012, p. 194

<sup>40</sup> M. M. Ratcliffe, 2007, p. 223

processes are seen as [embodied](#), [embedded](#), [enacted](#) and [extended](#). This conception is shared by Gallagher's *Interaction Theory* (IT) and enactive theories<sup>41</sup> briefly described below.

- Phenomenology-inspired accounts draw their inspiration from direct experience, and hence provide new answers to [question 1](#) and [question 2](#) of Goldman's desideratum. We do not attribute mental states, we experience them directly. And we don't need a separate account of self-experience, since self-experience is as direct as other-experience.
- Mental states also do not have any representational content, which is the simple answer to [question 4](#) of Goldman's desideratum. The content of direct experience is the experience itself.

### 2.2.1 Interaction theory

Gallagher's IT is based on notions of primary and secondary intersubjectivity originally set out by Travençolo<sup>42</sup>. *Primary intersubjectivity* is what infants have developed by the end of their first year: the ability to recognise emotions and intentions in the faces, voices and postures of others. It is a "non-mentalistic, perceptually-based embodied understanding". After the first year, *secondary intersubjectivity* builds on *primary intersubjectivity*. Shared or joint attention develops, because the infant starts to recognise relevant context and starts to recognise intentions as relations between agents or between agents and objects. This is a "perceptual capacity that is fast, automatic, irresistible and highly stimulus-driven"<sup>43</sup>. Gallagher's claim is that primary and secondary intersubjectivity is all we normally need for social interaction. He does not deny that we theorise about the beliefs and intentions of others – but such theorising does not constitute the bulk of social interaction. He allows that it is possible that other theories, perhaps even ST or TT, may account for how we – on occasion – "more deeply interpret"<sup>44</sup> another person. An important example of such a theory is Hutto's *Narrative Practice Hypothesis*<sup>45</sup>, which explains how we create stories (narratives) to provide reasons for actions.

Because interaction theory is so firmly rooted in developmental psychology, IT constitutes its own answer to [question 3](#) of Goldman's desideratum.

### 2.2.2 Enactive theories

Enactive theories are the New Kids on the Block. They are firmly anchored in the philosophical phenomenalist tradition, but Gallagher also sees important pragmatic<sup>46</sup> roots. There are three main strands in enactive theories<sup>47</sup>:

*Autopoietic enactivism*<sup>48</sup> is grounded in the biodynamic of living systems. It consists of a framework of 6 levels. At the bottom we find autonomous cells with intrinsic teleology. Through the levels of adaptivity, agency, mentality and sociality, we arrive at the top level, culture, which includes culture and language. An important concept is [autonomy](#), which allows the organism to actively modify its relations with the environment.

---

<sup>41</sup> Gallagher, 2012, pp. 210–211; also see Michael, Christensen, & Overgaard, 2014

<sup>42</sup> Gallagher, 2012, p. 195; Gallagher, 2008

<sup>43</sup> Gallagher, 2008, quoting Scholl & Tremulet (2000)

<sup>44</sup> Gallagher, 2008, p. 168

<sup>45</sup> Hutto, 2008; Gallagher & Hutto, 2008

<sup>46</sup> Gallagher, 2017, pp. 50–51. This is an interesting idea. It may allow for a connection between the enactive approach and pragmatist philosophers such as Robert Brandom and his 'deontic scorekeeping'. Gallagher (2012, p. 74) also makes this connection.

<sup>47</sup> Ward, Silverman, & Villalobos, 2017

<sup>48</sup> Di Paolo, 2005

- Because autopoietic enactivism is based on biodynamics, it tells its own evolutionary story in answer to [question 7](#) of Goldman’s desideratum Through the concepts of *precariousness*<sup>49</sup>, *operational closure*<sup>50</sup>, *robustness*<sup>51</sup>, *active homeostasis*<sup>52</sup> en *freedom*<sup>53</sup>, it explains how living organism survive<sup>54</sup> and thrive through interaction with their environment.
- Autopoietic enactivism provides an interesting answer to [question 6](#) of Goldman’s desideratum. Because [sense-making](#) and embodied perception are fundamentally interwoven, we should be looking at how the perceptions of people with autism differ from non-autistic people. De Jaegher<sup>55</sup> proposes that research into autistic embodiment start to look at sensory and perceptual differences and discover why they are different. As she puts it: “When a person with autism moves, perceives, or emotes differently, this relates inextricably to how he understands the world”. This fact, she says, is under-recognised.

*Sensorimotor enactivism*<sup>56</sup> views perception, not as not as a passive sense-think-act” event, but as an exploratory process. It postulates sensorimotor contingencies, which are regularities in the relationship between sensory stimulation and the actions of the perceiver. The theory is not committed to [autopoiesis](#); indeed, Degenaar and O’Regan<sup>57</sup> state that there may be consciousness without life.

- Sensorimotor enactivism is usually regarded as a research program intent to discover how perception works, and in that way supplies its own answer to [question 3](#) of Goldman’s desideratum

*Radical enactivism*(REC)<sup>58</sup> attempts to clean up both autopoietic and sensorimotor enactivism. It draws a sharp line between basic contentless cognition and linguistically mediated thought. Sensorimotor enactivism is to be cleared of any traces of representationalism. [Autopoiesis](#) and [sense-making](#) are deemed too vague and liberal<sup>59</sup>. Instead REC offers a Developmental-Explanatory Thesis which says that “nothing other than its history of active engaging structures or explains an organism’s current interactive tendencies”, essentially a process of “laying down a path in walking”. This is the REC answer to [question 3](#) of Goldman’s desideratum.

### 3 WHY MINDREAD? COMMON ASSUMPTIONS

---

In outlining the landscape of mindreading-related theories, some issues stand out.

- The general consensus seems to be that there are different kinds of mindreading which cannot be explained by one theory. No theory “does it all”.
- In spite of ostensibly fierce competition between viewpoints and indeed between philosophers, both the traditional and the phenomenologically inspired accounts may be viewed as

---

<sup>49</sup> Di Paolo, 2018, p. 12

<sup>50</sup> Di Paolo, 2009 ; illustrated by Di Paolo & Thompson, 2014

<sup>51</sup> Di Paolo, 2005, p. 434

<sup>52</sup> Di Paolo, 2009, p. 9

<sup>53</sup> Di Paolo, 2009, p. 17

<sup>54</sup> Froese & Di Paolo, 2011, p. 9

<sup>55</sup> De Jaegher, 2013

<sup>56</sup> O’Regan & Noë, 2001

<sup>57</sup> Degenaar & O’Regan, 2017

<sup>58</sup> Hutto, 2005

<sup>59</sup> Hutto & Myin, 2013, pp. 34–35;

serious theories of cognition. They all attempt to answer Goldman's 7-question desideratum for such a theory.

- Each theory seems to be based on a set of assumptions which are not questioned.

Let's examine some of the assumptions which seem to underlie mindreading accounts. There are specific reasons why we should not be so sure of them. There is also a general reason. What these assumptions have in common, is that they flatter us. In the absence of serious proof, sometimes even in the face of conflicting evidence, we still like to see humankind as special. Intelligent, rational, and creative. Predestined through genes or evolution to be the master of creation. In short, these assumptions feed our feelings of superiority. We are gripped by a picture, as Hutto puts it<sup>60</sup>, which we cannot resist.

### 3.1 ACTIONS ARE BELIEF-MOTIVATED

The idea that beliefs and desires are the cause of human action goes as far back as Aristotle<sup>61</sup> and in more recent history, Hume<sup>62</sup>. In the twentieth century it became a popular, mainstream belief. In 1987 Fodor<sup>63</sup> wrote: "there is, so far, no human group that doesn't explain behaviour by imputing beliefs and desires to the behaviour (And if an anthropologist claimed to have found such a group, I wouldn't believe him)". Yet he was already well aware of that the belief-desire model of behaviour could not count on universal acceptance amongst philosophers. In 1984 he wrote: "there are, of course, two kinds of philosophers. One kind of philosopher takes it as a working hypothesis that belief/desire psychology (or anyhow some variety of propositional attitude psychology is the best theory of the cognitive mind we can now envision [...]) The other kind of philosopher takes it that the entire apparatus of propositional attitude psychology is conceptually flawed in irremediable ways".

Turner<sup>64</sup> notes that on top of philosophical disapproval, findings from the field of linguistics show that our Western way of expressing beliefs, desires, reasons and intention, cannot always be translated into other languages. Things have not improved for the popularity of the belief-desire model<sup>65</sup>.

Many philosophers and scientists now argue that our understanding of others is mostly immediate, not a conscious interpretation. The TT/ST hybrid proposal by Nichols & Stich<sup>66</sup> makes room for this by stating that "most of the processes of behaviour prediction that we have described require little or no conscious access". On this proposal, we have two kinds of mindreading, conscious and unconscious, which interestingly, is a general possibility accepted the traditional and the phenomenologically-inspired accounts accept, be it for different reasons (we saw earlier: Apperly & Butterfill<sup>67</sup>, Goldman<sup>68</sup> and Gallagher<sup>69</sup>). Slors<sup>70</sup> points out that the move to go "sub-personal", i.e. to allow that most kinds of mind-reading happen unconsciously, does not explain why we constantly talk as if the belief-desire model were true. He proposes that we look at TT, not as a social-cognitive mechanism with psycho-

---

<sup>60</sup> Hutto, 2009

<sup>61</sup> Aristotle, 1984, bk. 3, paragraph 10 (433a9-433a12)

<sup>62</sup> Hume, 1888

<sup>63</sup> Fodor, 1987, p. 132

<sup>64</sup> Turner, 2018, quoting Bittner (2001), Needham (1972 and Mercier & Sperber (2011,2017)

<sup>65</sup> Strijbos & de Bruin, 2012, quoting Gallagher & Zahavi (2008), Ratcliffe (2007) en Gallese (2005)

<sup>66</sup> Nichols & Stich, 2003, p. 94

<sup>67</sup> Apperly & Butterfill, 2009

<sup>68</sup> Goldman, 2006, chapters 6 and 7

<sup>69</sup> Gallagher, 2008, p. 168

<sup>70</sup> Slors, 2012

neural reality, but as talk generated by a TT model<sup>71</sup> we adopt in the [intentional stance](#). What is generated by this model may be false<sup>72</sup>, but it is our cultural, rational way to talk about ourselves to each other.

Strijbos & de Bruin<sup>73</sup> take this idea even further, and connect it with Brandom's notion of deontic scorekeeping, the game of giving and asking for reasons. Explaining this theory is beyond the scope of this paper. Briefly, people keep constant tabs on each other. Interpreting and explaining actions in terms of their reasons is a normative requirement. We constantly call each other out and ask for justification. This practice seems ubiquitous, in fact, there is evidence, says Andrews<sup>74</sup>, that it also occurs amongst animals. It is, however, not dependent on the belief-desire model to be physically real. It may be just a cultural practice, something we like to do.

### 3.2 HUMANS COMPETE FOR SUCCESS

According to the *Machiavellian hypothesis of Social intelligence*, we mindread in order to be successful. To being able to mind-read is advantageous in socially complex situations that require competition, cooperation, or coalition. Because humans are thought to be much better at mind-reading than animals, this hypothesis helps to explain the success of our species. The hypothesis was formulated by Byrne & Whiten in 1988 and 1997<sup>75</sup> in response to a description of Frans de Waal's book *Chimpanzee Politics* (1982)<sup>76</sup>. Since then, it has been much quoted, elaborated, extended and somewhat revised, but still very much alive. In 2018, the *Journal of Comparative Psychology* issued a special edition on Machiavellian Intelligence.<sup>77</sup>

The Machiavellian hypothesis is built on Humphrey's (1976) notion of primates having large brains. Large brains are expensive, in terms of energy expenditure and vulnerability, but they also allow us to interact socially and to remember those interactions. This idea developed into the Social intelligence hypothesis<sup>78</sup>, which says that large brains and intelligence are necessary for solving social problems; social intelligence can be used for other things (like toolmaking); social complexity drives selection for enlarged brains and intelligence and complex social groups select for greater social intelligence.

On this conception, the socially intelligent primate understands, predicts and manipulates conspecifics because this is socially advantageous. The word *manipulation* is used here in a neutral sense, to include the idea of influencing and cooperating as well as coercion and deception, including mimicry and camouflage. However, the Social Intelligence is not the only hypothesis attempting to explain the evolution of large brains. Dunbar & Schultz<sup>79</sup> provide welcome guidance. They list 6 core observations to be explained. They also define two broad categories of hypotheses: the instrumental hypothesis (food) and the social hypotheses. The latter category includes five further hypotheses: (i)

---

<sup>71</sup> This is on the same lines as Ravenscroft, 2019, quoting Maibom (2003) and Godfrey-Smith (2005) who propose that folk psychology is a model that can be used for different purposes in different circumstances.

<sup>72</sup> Dennett, 2009

<sup>73</sup> Strijbos & de Bruin, 2012

<sup>74</sup> Andrews, 2020. This paper is part of a cluster of papers: Vincent et al (2019) and Monsó & Andrews (2020). In another (unpublished) paper I have reviewed the empirical evidence that Andrews cites in these three articles. Most of it does not hold up (too vague or irrelevant), but there remains some compelling evidence for conventions, sanctions and cognitive dissonance.

<sup>75</sup> Whiten & Byrne, 1997

<sup>76</sup> Byrne, 1996

<sup>77</sup> Hopper, Waal, van der, & Caldwell, 2018

<sup>78</sup> Johnson-Ulrich, 2017

<sup>79</sup> Dunbar & Schultz, 2017

Machiavellian intelligence ii) Cultural intelligence, (iii) Vygotskian intelligence, (iv) Scheherazade and (v) Social Brain. These hypotheses differ in what factor is deemed most important (food versus social processes), whether food, mating or predation is most important to evolutionary fitness and whether this benefit is directly beneficial to individuals. Dunbar & Schultz conclude that the Social Brain Hypothesis provides a good overall framework to explain all findings, but that none of the others come even close. The Machiavellian hypothesis comes out as purely descriptive, but without explanatory value. In fact, Machiavellian behaviour is more likely to be a consequence of living in large groups, rather than its cause. It also offers no explanation of why group size (and hence, according to Dunbar & Schultz, brain size) should vary across primates.

Andrews<sup>80</sup> directly challenges the Machiavellian hypothesis on other grounds. On her view, which she dubs *pluralistic folk psychology*, mindreading is not a causal theory but a social competence<sup>81</sup>: the ability to identify, predict, explain, justify, normalise and coordinate behaviour. These abilities are supported not by one, but by different cognitive mechanisms – which can be shown to exist in animals and humans alike. Recognising intentional behaviour (and agents) is a core competence which all primates, perhaps even all animals, have. Intentional behaviour includes not just propositional attitudes (which animals may or may not have), but also moods, traits, dispositions, emotions and enabling conditions.

### 3.3 MINDREADING IS IN THE BRAIN

The relationship between mental processes and the brain is a huge topic which cannot be done justice here. The claim is that thinking is the processing of meaningful physical entities, that is, mental representations. What we think about each other may not be true, but it is real, in a physical, scientific sense. Shea<sup>82</sup> calls it the “most important insight of the 20<sup>th</sup> century, the answer to the mystery of thinking. The general idea is something like this:

- a. mental processes are brain processes which handle representational content
- b. representational content has a physical reality which we can or should be able to establish empirically.
- c. representational content is causally or otherwise related to an object, property or condition with physical reality

Statements a) and b) are generally assumed to be true, although there are huge differences between theorists about what a representation is. Rowlands<sup>83</sup> points out that the classic idea of representation was initially modelled on language and has five characteristics:

- is *internal* (identical to some neural configuration inside the representing subject) and may take several forms (images, prototypes, symbols, etc)
- has *duration* when the representation is activated, brought online, so to speak. Duration is distinguished from possession.
- has *content* which *refers to something external* to itself
- requires *interpretation*
- is *passive*, in the sense that it is produced, is the end result of a chain of events.

---

<sup>80</sup> Andrews, 2012; Andrews, 2018

<sup>81</sup> She calls this competence “Mengzian” as opposed to “Machiavellian”.

<sup>82</sup> Shea, 2018, p. 6

<sup>83</sup> Rowlands, 2006, pp. 2–5

This general model of representation has been filled in, extended and modified in myriad ways, made the backbone of a number of important theories, and still we don't have a good idea of how representations get their content. We are, as Shea<sup>84</sup> puts it, like the academic in the cartoon, musing: "Well it works in practice, Bob, but I'm not sure it's really gonna work in theory".

Some philosophers are simply not interested. If one belongs to that group of philosophers who - on Fodor's<sup>85</sup> account - do not believe in "the entire apparatus of propositional attitude psychology", then the representation of beliefs and intentions is simply irrelevant.

From the application of the theory to Artificial intelligence, we know that that theorising about representations has its limits: it cannot, by itself, tell us how a system recognises relevant features in a changing environment. This is known as the *frame problem*<sup>86</sup>.

Andrews<sup>87</sup> notes we have two questions to answer about mindreading, namely how do we 1) attribute mental content to ourselves and others and 2) predict and explain behaviour. She argues that the answer to the first question is not necessarily an answer to the second. The mechanisms and methods we use to predict others' behaviour need not be the same methods we use to attribute mental states.

Ratcliffe<sup>88</sup> says that important philosophical work has not yet been done. The whole notion of a brain process presupposes a particular take on personhood. When I mindread, the mental process that is going on in my brain, belongs to me. I use it to interpret another person whose mental processes do not belong to me but to that other person, and which are invisible to me. Still, somehow I know the difference between my own mental states and those that I assume to belong to someone else. Ratcliffe provides an overview of responses from the traditional accounts of mindreading to this objection. The important point, however, is that none of the mindreading accounts, traditional or phenomenological, account for how or why we recognise other as persons. On those accounts, there does not seem to be an essential difference between intra- and interpersonal mindreading. Another issue is that our typically western conception of individual thinking person may not be universally true. In African and South American communities<sup>89</sup>, *personhood* is developed through social connections and standing in society and in relation to life-events. Sometimes referred to as 'dividual' (as opposed to individual), *personhood* is not static nor determined by cognitive characteristics. By analogy, there may exist mental processes which run through groups, rather than through individual group members.

The existence of representations is not settled, not even in the eyes of its defenders. Cecelia Heyes<sup>90</sup>, in a reply to criticisms to Cognitive Gadgets<sup>91</sup>, suggests that philosophy, like the natural science, focusses on *understanding* and *meaning* whereas cognitive scientists focus on *explanation* and *information*. She agrees that we are still a long way off from integrating the two approaches, but that there is very little point in throwing away the achievements of either side. Only an integrated approach will lead to a better understanding.

---

<sup>84</sup> Shea, 2018, p. 6..

<sup>85</sup> Fodor, 1987, p. 132

<sup>86</sup> Shanahan, 2016

<sup>87</sup> Andrews, 2012, p. 10

<sup>88</sup> M. Ratcliffe, 2009..

<sup>89</sup> Motsamai Molefe, 2019; Menkiti, 1984; Degnen, 2018

<sup>90</sup> Heyes, 2019a

<sup>91</sup> Heyes, 2018

Gallagher<sup>92</sup> agrees. He points out that although some phenomenologists think that representations are an unnecessary theoretical construct, the enactive program does make claims about how the dynamics of brain-body-environment work. These claims can be tested at the same time but need not be taken literally by scientists: “Although to work out a philosophy of nature is not to do science, it can still offer clarifications relevant to doing science, and it can inform empirical investigations”.

As Pitt<sup>93</sup> summarises: “Contemporary philosophers of mind have typically supposed (or at least hoped) that the mind can be naturalized – i.e. that all mental facts have explanations in terms of natural science. This assumption is shared with cognitive science, which attempts to provide accounts of mental states and processes in terms (ultimately) of features of the brain and central nervous system. In the course of doing so, the various subdisciplines of cognitive science (including cognitive and computational psychology and cognitive and computational neuroscience) postulate a number of different kinds of structures and processes, many of which are not directly implicated by mental states and processes as commonsensically conceived. There remains, however, a shared commitment to the idea that mental states and processes are to be explained in terms of mental representations”.

### 3.4 MINDREADING IS IN OUR GENES

Fodor’s<sup>94</sup> account mental modules has been highly influential<sup>95</sup>. It is on the extreme end of what Carruthers<sup>96</sup> calls the modularity spectrum. A *Fodor module* is a processing system that is dedicated, specific, fast, neurologically fixed and innate. Its conception may be traced back, amongst others, to the work of Gall, the founding father of phrenology. Fodor was also inspired by Noam Chomsky’s language acquisition device. He proposed a mental module for a compositionally structured language of thought (Mentalese) which underlies all natural languages. It presupposes the existence of propositional attitudes and representations. It also lends credence to the idea that we have two kinds of processing systems: one innate or early developing, automatic, dedicated fast-and-efficient system-1 for implicit mentalising, and another, later developing, controlled, slow-and-flexible system-2 for explicit mentalising.

The idea that the mind consists of multiple modules has also dominated Evolutionary Psychology<sup>97</sup>. Specialised components have evolved to help us deal with specific problems our Stone Age ancestors had to face. However, alternative views have arisen. Cecilia Heyes, for instance, claims that mindreading, like text reading, is a *cognitive gadget* rather than a *cognitive instinct* – grist, not mill. She explains this by analogy with text reading. It has been empirically established that the brain area involved in text reading is created as the ability develops. In much the same way, mindreading develops from domain general learning mechanisms. In evolutionary terms, the cognitive abilities for selective social learning and imitation are important, because these are “gifts that go on giving”<sup>98</sup>, allowing for the inheritance of more and more skills. She likens<sup>99</sup> the mind to a hand as opposed to a swiss-army knife. A hand can perform many tasks, including tasks that are presently unknown. A swiss army knife

---

<sup>92</sup>Gallagher, 2017, p. 23

<sup>93</sup> Pitt, 2000

<sup>94</sup> Fodor, 1983

<sup>95</sup> It is not quite clear how influential Fodor’s account is now. The entry in the Stanford Encyclopedia by Rescorla, 2019 suggests that it is still very influential; whereas Marc Slors has stated on several recent occasions “that nobody believes in a Language of Thought anymore”.

<sup>96</sup> Carruthers, 2006, p. 3

<sup>97</sup> Carruthers, 2006, referring to Tooby & Cosmides (1992); Heyes, 2012, *ibidem*

<sup>98</sup> Heyes, 2018, p. 3

<sup>99</sup> Heyes, 2012



is a set of clever instruments designed for specific purposes. Traditional philosophers and scientists had things upside down, she says: "the very idea of mental states was invented by smart people using general-purpose cognitive mechanisms"<sup>100</sup>.

Cecilia Heyes has also examined the empirical evidence for the notion of two separate cognitive systems for mentalising<sup>101</sup> and found it wanting. Her statement is that contextual cues are processed by a generic cognitive capacity that is not specific to thinking or to language. This she calls [submentalising](#). Moreover, she says, most of the time apes and human do not need to mentalise: "unless one needs to discuss behaviour, or to catch a Hollywood spy, submentalising may be the smart option"<sup>102</sup>. In Heyes' view, mentalising is culturally inherited. It is not an inherited cognitive capacity; we learn it through social interaction just as we learn to read, only earlier<sup>103</sup>.

### 3.5 TAKING STOCK

We have been looking at some of the important assumptions about why we mindread:

- belief-desire psychology: we need mindreading for social interaction
- human competition: when resources are scarce, one has to be clever
- mindreading as a brain function: our brain is wired that way
- mindreading as a genetic ability: we mindread because we can

These underlie the theoretical landscape that attempts to explain the phenomenon. Trying to create that theory amounts to tackling a so-called 'wicked' problem – for *mindreading* ticks many of the criteria listed by Rittel and Weber<sup>104</sup>:

- There is no definite formulation of the problem.
- The information needed to understand the problem depends on one's idea of solving it.
- There are different explanations but no way to choose between them.
- There is no finite set of solutions, so we may never know when we are done examining.

How to proceed? The remainder of this paper contains a sketch of an answer to the question that should be answered first: why do we mindread?

---

<sup>100</sup> Heyes, 2019b

<sup>101</sup> Heyes prefers the term 'mentalising' to 'mind-reading'; I have followed her use of that term when discussing her views.

<sup>102</sup> Heyes, 2017, p. 2

<sup>103</sup> Heyes, 2018, p. 147

<sup>104</sup> Rittel & Webber, 1973

## 4 MINDREADING: BUILDING ON TRUST

---

In a recent article, Spaulding<sup>105</sup> brings up the disconnect that still exists between research on mindreading and research on social cognition and social psychology. There is a narrow-focus research agenda that is shared between disciplines, focused on belief-attributions for the purpose of behaviour prediction and explanation. Unfortunately, Spaulding says, other topics from social cognition and psychology, such as in-out grouping, stereotyping, character trait attribution, and the role of personal goals and situational context which are clearly relevant to mindreading, tend to be left out of the philosophical discussion. A scan of the [Stanford Encyclopedia of Philosophy](#) suggests that she is right.

Spaulding also says that mindreading has different goals: (i) when we want to be as accurate as possible, for instance, in a job interview, we tend to think carefully about the behaviour and mental state of the other person; (ii) when we don't care that much, we will select a low-effort mindreading strategy; (iii) or we may just want to influence someone, rather than explain or predict their behaviour; (iv) or we may be moved by an instinctive goal such as in-group preservation, and mindread the behaviour of out-group members with a particular bias. What kind of goals are these? Let's take a leaf from Aristotle<sup>106</sup> and his four causes (material, formal, efficient and final), necessary and sufficient for an adequate explanation. With Spaulding's first three goals one might say that the final cause is to mindread the other person(s) but that the result may vary in accuracy. That is a *material* cause: the choice to use a lesser or better quality process. The last example, that of out-group bias, is a *formal* cause: the mindreading is done (designed) differently with in- or outgroups. None of these examples are about *final* causes of mindreading.

In previous paragraphs it has already become apparent that there is not much consensus on what mindreading is or how it works. In academic work on the topic of mindreading, there tends to be an endless pitting of arguments against each other. Yet there are also common denominators to be found. These are discussed in the next paragraphs, followed by a sketch of how these may be connected in a modified mindreading framework.

### 4.1 MINDREADING AND SHARING

Using textual analysis<sup>107</sup> on a sample of 20 influential articles<sup>108</sup> from all sides of the debate, the word *shared* is frequently used as an adjective – *shared* representations, practice, viewpoint, state, knowledge, understanding, reality, experience, intentions, context: all *shared*. The noun that is qualified by *shared*, is usually defined, but not the word *shared* itself. The same is true for words that imply a type of sharing: empathy, coordination, joint attention, common ground and trust – these tend to be used liberally but also tend to escape definition.

Extrapolating from this sample, that *sharing* – however defined – may be a common denominator in philosophical texts on mindreading, the question is why *sharing* might be important? A possible

---

<sup>105</sup> Spaulding, 2020, citing other philosophers who have said the same: Apperly (2012), Rakoczy (2014) and von Eckhart (1997); see also

<sup>106</sup> Shields, 2016

<sup>107</sup> Using computational literary analysis tools. Analysis is available but not included with this paper because of its length

<sup>108</sup> In fact, the reading list for this seminar.

answer is suggested by Fonagy & Luyten<sup>109</sup>. They regard mentalising<sup>110</sup> as multidimensional, the dimensions being 1) automatic versus controlled mentalising, (2) mentalising with regard to self or others, (3) mentalising based on external or internal features of self and others, and (4) cognitive versus affective mentalising. Normally, these dimensions are balanced, but if one dimension gains persistence dominance over another, mentalising ability may fail. According to Fonagy & Luyten, this is what happens in personality disorders. When mentalising ability is impaired, the patient becomes incapable of taking in relevant social data. This in turn causes interpersonal difficulties and the loss of a stable idea of *self*. The patient has then become unable to activate her capacity to think differently about herself and her surroundings.

Of particular interest to our question of why *sharing* would be so important to mindreading, is the role of the therapist. Fonagy & Luyten say that the therapist must promote curiosity in the patient about the way mental states motivate and explain the actions of the self and others: “in effect the therapist is modelling how he or she engages in mentalizing in relation to the patient”. The therapist can show, for example, that his mind has been changed by the patient. That way the therapist gives agency to the patient and increases her faith in the value of social understanding<sup>111</sup>. The aim is to regenerate the patient’s own capacity to mentalise, but it is not the ultimate aim. The ultimate aim is for the patient to be able to learn from social experience, “to update and build on knowledge about the self and others in social situations”: “the essence of all effective psychotherapy”. The key ingredient, they say, is [epistemic trust](#), which the patient derives from being able to mentalise. This allows for the recovery of the capacity of social exchange. Although Fonagy & Luyten do not suggest it, it may also be the case that the therapist-patient (therapeutic) trust, once established, allows the patient to outsource (part of) her own impaired mindreading process to the therapist. If the restoration of [epistemic trust](#) is key to being able to mentalise, the role of the therapist might be to take on a task that the patient temporarily cannot perform for herself, in order to kick-start the continuous cycle of social learning.

Fonagy & Luyten mention different kinds of *trust*: patient’s trust in oneself; patient’s trust in the mentalising process; patient’s trust in own mentalising ability, patient’s epistemic trust in the informant as a source of relevant social information, trust between therapist and patient, and the history of trust and mistrust the patient carries with her. Trust begets trust: once the patient has mustered enough trust to re-join the social game, she can start to rebuild her trust of self and others, and re-establish a coherent and flexible sense of self.

Let’s have a further look at trust in philosophical literature.

## 4.2 TRUST AND COGNITIVE OFFLOADING

The received view is that almost all knowledge depends for its acquisition on trust in others. The argument for this is that no person has the time, intellect and experience necessary to learn all the facts about the world that are collectively known<sup>112</sup>. We might say that trust allows for [cognitive offloading](#): if I can use your knowledge without having to question it, I don’t have to verify it myself. [Cognitive](#)

---

<sup>109</sup> P. Fonagy & Luyten, 2018

<sup>110</sup> Fonagy & Luyten (2018) use the term mentalising rather than mindreading. I have followed their use when discussing their article.

<sup>111</sup> Peter Fonagy, Luyten, Allison, & Campbell, 2019

<sup>112</sup> McLeod, 2015

[offloading](#) has not received much philosophical attention<sup>113</sup>, but it is an important and truly ubiquitous cognitive strategy.

Risko & Gilbert<sup>114</sup> recognise roughly two forms: offloading onto the body (e.g. using our fingers to point, or our eyes to index locations in space), and offloading into-the-world. Both strategies save energy and help us think. Without it, we would be stuck with our unaided mental abilities (e.g. see only a small part of the visual field accurately, remember only a few things actively). Findings are consistent with this, according to Risko & Gilbert, for cognitive offloading improves performance in perception, memory, arithmetic, counting, spatial reasoning and other domains. They also say that there is much that we do not know yet about the mechanics of cognitive offloading, and they list a number of research questions. For present purposes, however, the notion of cognitive offloading as ubiquitous means of outsourcing complex cognitive processes to things or agents outside one's immediate control, is sufficient.

### 4.3 TRUST, RELIABILITY AND JUSTIFIED BELIEF

There are various conceptions of trust, oscillating between *mere reliability*, *trust*, and *justified trust*. We might view this as a scale of trustworthiness:

*Mere reliability*: I can rely on something or someone without trust. For instance, I can trust my employer to respect my privacy rights, not because my employer has any interest in my privacy, but I know my employer to be concerned, out of self-interest with their public image. If something goes wrong with respect to my privacy, I will be disappointed with my employer, but I will not feel betrayed. Reliance is also not necessarily about a justified belief. I may decide to rely on something or someone simply because I have no choice, or no better alternative.

*Trust*: Assuming a will-based account<sup>115</sup> of trustworthiness, a trustee is deemed trustworthy by the trustor when she is assumed to act out of *goodwill* towards the trustor or whatever the trustee is entrusted with, or both. In other words, the trustor must attribute motives to the trusted person. Again, as with reliability, this secondary belief about the trustee's trustworthiness does not have to be justified. We do misplace our trust.

*Justified trust*: Requires the fulfilment of extra conditions. McLeod<sup>116</sup> tells us that at present there is no comprehensive philosophical account of when trust is warranted. She recommends an integrated approach, including the fields of philosophy of mind, epistemology and value theory. Indeed, there are many elements to consider:

- trustworthiness of the trustee
- ability: can the trustee deliver?
- plausibility: do circumstances allow?
- instrumental trust value: affording opportunity for cooperation, autonomy, self-respect, knowledge, and meaningful relationship

---

<sup>113</sup> There is no description or discussion of cognitive offloading in the Stanford Encyclopedia of Philosophy. However, Marc Slors' forthcoming book contains an overview and discussion.

<sup>114</sup> Risko & Gilbert, 2016

<sup>115</sup> McLeod, 2015, quoting Jones (1999) and Baier (1986,1991,1995, 2004). There are other accounts: virtue-based and risk-based, but on those views violation of trust not cause betrayal, only disappointment.

<sup>116</sup> McLeod, 2015

- *trust* versus distrust: according to Hawley<sup>117</sup>, an account of trust should also explain when trust is not appropriate, for instance when distrust is appropriate, or because neither trust nor distrust is appropriate.
- *justified belief* in trust versus *prediction*; as Heyes<sup>118</sup> asks: to what extent and under what conditions are these the same?
- *trust* cannot be willed, it can only be encouraged and developed. This implies the involvement of affect (trust “piggybacks” evolutionary affective development, according to Fontagy and Luyten<sup>119</sup>)

The issue of *epistemic individualism*, is raised by Goldberg in response to Goldman’s theory of *process reliabilism*. This theory holds that my belief in *x* is justified if and only if that belief has been formed by a reliable process. Goldberg notes that such a view relies on *process individualism*, i.e. “the belief-forming processes relevant to the epistemic assessment of a given belief all take place within the mind/brain of the believing subject herself<sup>120</sup>”. He argues that at least some of the beliefs that we form through epistemic reliance also require our assessment of the reliability of the cognitive processes themselves<sup>121</sup>. For instance, if we assess the reliability of a memory, we not only assess the reliability of the memory itself, but we also assess the reliability of our memorising ability: we assess ourselves as a source of information. Goldman’s claim is that the extent to which we assess ourselves as a reliable source of information in forming and sustain a particular belief, is determined by the extent to which we rely on that belief. We do the same in our assessment of others, and of society. In focussing on the individual as the locus of an epistemic belief, we assess the way that individual manages epistemic reliance on others, i.e. we assess *competence*.

#### 4.4 TRAITS AND STEREOTYPES

Spaulding<sup>122</sup> draw our attention to traits and stereotypes, citing quite a large number of empirical evidence. When we come across other people, we immediately sort them into categories – age, gender, race, nationality, religion, class and many others. Most of these categories are modulated by context, or the way we feel, or by what we want to achieve. We tend to attach character traits to stereotypes, such as competence, dominance, trustworthiness and aggressiveness. Often, Spaulding says, we use trait attribution in the same way that we use mindreading: to explain and predict behaviour.

It may be that traits and stereotypes should be regarded as the historic products of cultural learning, that is, a consolidation of attribution of stereotypes and traits by others. If that is correct, traits and stereotypes, once fixed, represent a source of epistemic knowledge. The same may be true of habits and conventions. The extent to which such cultural knowledge is trusted, would have to fit the general account of how we go about trusting knowledge, as discussed in the previous paragraph.

#### 4.5 SENSEMAKING AND FEEDBACK MECHANISMS

The topic of sense-making has come up several times already. There is the general, common-sense notion of mindreading as attempting to make sense of others. On the traditional account (see paragraph 2.1.1), there is Davidson’s notion of content-determination of mental states as a triangulation

---

<sup>117</sup> Hawley, 2014

<sup>118</sup> C. Heyes, personal communication, 21 August 2020

<sup>119</sup> P. Fonagy & Luyten, 2018

<sup>120</sup> S. Goldberg, 2010, p. 1

<sup>121</sup> S. C. Goldberg, 2010, chapter 3

<sup>122</sup> Spaulding, 2020

involving two individuals and the world. On the phenomenology-inspired account, an extensive description of what it means to make sense of the world is offered by philosophers such as Di Paolo, De Jaegher and others in terms of *autopoietic enactivism* (see paragraph 2.2.2). This is essentially a biological theory which says that an organism does not just survive, but is taking active steps so it will still be there tomorrow. It does this by anticipating on its environment. As a phenomenological function sense-making is made possible by physical regulation and by self-monitoring processes. Di Paolo<sup>123</sup> points at the underlying mechanism with reference to Ashby<sup>124</sup>'s model of adaptive behaviour. The essence of this model is that to maintain homeostasis actively, you need two processes: one which is the primary process, and another one which is monitoring the primary process, to make sure that it can keep going. That secondary process, the monitoring process that keeps tabs on the primary through a *feed-back mechanism*, is what allows the organism to stay alive and adapt to the environment.

The notion of a *feed-back mechanism* is of interest here, because it keeps reappearing in connection with mindreading:

- We seem to adjust the quality of our mindreading to the goals of the moment<sup>125</sup>.
- We assess the extent to which we can or must rely a situation, an object or a person before running the risk of outsourcing something we are dependent on (see paragraphs 4.2 and 4.3).
- Regulatory accounts of mindreading, such as those by McGeer<sup>126</sup>, are centred around the notion of regulation. Similarly for language-related theories, such Brandom's deontic scorekeeping and Geurts'<sup>127</sup> view of shared (reflexive) commitments as essential for action-coordination. Andrews'<sup>128</sup> theory of naïve normativity is also based on a feedback mechanism, as the violation of naïve norms lead to cognitive dissonance, even in animals.
- A cultural view of mindreading, such as that by Fenici & Zawidzki's<sup>129</sup> account also draws attention to the mindshaping aspect of mindreading. They say that mental state ascription may have come about "to track the practical commitments cognitive agents assume when creating, sharing, and negotiating shared projects and goals". They stress that making these practical commitments have an important regulatory function in that it supports action coordination and alignment on joint goals".

## 4.6 SKETCH FOR A MODIFIED APPROACH

Having collected some of the essential ingredients, it is now time to envisage how these elements might feature in answer to the "why" question of mindreading.

### 4.6.1 General principles

1. Mindreading affords us some sense of what others think or feel.
  - a. This 'sense' take various forms: we may experience this as inferring knowledge, or as experience the other's mental state, or our awareness of it may be indirect.
  - b. Whatever we sense need not be true.

---

<sup>123</sup> Di Paolo, 2009, p. 9

<sup>124</sup> Ashby, 1960, p. 83

<sup>125</sup> Spaulding, 2020, also see beginning of chapter 4

<sup>126</sup> McGeer, 2007

<sup>127</sup> Geurts, 2019

<sup>128</sup> Andrews, 2019

<sup>129</sup> Fenici & Zawidzki, 2020

- c. These various forms of sensing what others think or feel may or may not correspond to differences in underlying cognitive architecture.
  - d. We may or may not express what we sense about others in language.
  - e. Mindreading does not require consent, i.e. one cannot avoid being mindread
2. What others think or feel is of interest to us for many reasons. Reasons may be specific to the situation. Generally speaking, there are two main reasons:
    - a. We may wish to rely on others for performing a task we cannot do (by) ourselves, such as providing food or shelter. We may also wish to use the other person as a source of knowledge, social or otherwise. Both are forms of *cognitive offloading*, which we may voluntarily choose, but it may also be the case that we have no choice in the matter.
    - b. When we rely on other people, we take a risk. In outsourcing tasks or in knowledge acquisition, we become dependent on the other person(s). We normally require some assurance, that is, we want to know to what extent we can *rely on* and even *trust* the other person. This is another reason for mindreading. We might also mindread ourselves to find out to what extent we can bear the negative consequences of a misplaced trust. In assessing the extent to which we can rely or trust the other person, we assess relevant aspects of that person's reliability or trustworthiness. We also assess our trust in our own mindreading abilities.
  3. Mindreading supports a regulatory feedback mechanism (active homeostasis). If we express our sense of what others think or feel through observable behaviour, those others can respond to affirm or correct, or we can correct ourselves.
    - a. Over time, we collectively build up a library of successful and unsuccessful mindreading outcomes. This may be in our own lives, but the same mechanism affords *cultural learning* through conventions, habits, traits, customs and even language.
    - b. Cultural learning includes *mindshaping* as well as mindreading, as we collectively build on our knowledge to further it. This includes knowledge about determination of trustworthiness for cognitive offloading, and about what tasks can be offloaded and to what effect.
    - c. Regulation implies controls, i.e. the setting of *thresholds*<sup>130</sup>. If a threshold is not crossed, no action is taken. Not to take action when things start to change, implies a form of *trust* in the control settings. For instance, when I drive I assume that the other cars will keep sufficient distance even if those distances keep changing. Therefore I do not brake every time another car comes near. But I will brake if I think there is going to be a collision.

#### 4.6.2 Implications

##### *No human-animal dichotomy*

On the above approach, there is no reason to suspect a sharp dichotomy between humans and other animals. The question becomes: what do animals want to rely on each other for, and how do they determine to what extent the other may be relied on or trusted? It may be important to investigate the nature of group-centred as opposed to individual-centred feedback mechanisms.

---

<sup>130</sup> Corresponding to Di Paolo's notion of robustness, see Di Paolo, 2005, p. 434

### *Impaired mindreading distorts social feedback mechanism*

It appears that mindreading is connected to (some) personality disorders. We need an account that distinguishes between a) not mindreading (for instance, when I am alone, writing this paper), b) flawed mindreading (causing a misfit with the social environment) and c) successful mindreading. On this proposal, only situation b) should lead to significant problems with PD patients because the feedback mechanism inherent in mindreading, breaks down. It does not matter where the error originates, inside the patient or with the outside world; when the feedback mechanism becomes distorted, the patient is in danger of spiralling down into distrust, incapacitating him or her to be part of normal social interaction. Importantly, not being able to mindread should lead to less problems than faulty mindreading.

### *Differences but also overlap between reliance on agents and non-agents*

If the analysis presented in paragraph 4.3 is correct, we would expect to find a difference but also overlap between trusting or relying on another agent (including ourselves), and relying on a non-agent. It would be interesting to work out exactly what that difference consists in, and how it relates to the regulatory feedback mechanism and its thresholds. The first thing to do would be to establish a working definition of agency that is not restricted to intentional human behaviour.

## **5 CONCLUSION**

---

In the first part of this paper, I have sketched mindreading as an explanandum that has been approached from several perspectives: the traditional approach (theory-theory, simulation theory or hybrid) and the phenomenology-inspired approach. Obviously it has not been possible in these few pages to do justice to the many intricacies of that theoretical landscape. The point of the exercise was to show that the explanans offered have not been concerned overmuch with the “why” question. However, some underlying assumptions about “why” are implicit in these accounts. These include belief-desire psychology, Machiavellian competitiveness, mindreading is in the head, and mindreading as a genetic hard-wired capability. These assumptions have been examined and found questionable if not unsupportable.

In the remainder of this paper, a number of notions have been examined which should feature in a modified approach to mindreading: cognitive offloading, trust and reliability, cultural knowledge such as stereotypes, traits and conventions, sense-making and regulatory feedback.

This paper concludes with a modified sketch of mindreading. First, as a primary goal oriented process, aimed at cognitive offloading. Secondly, as a risk-management process, regulating reliance and trust according to pre-established thresholds. This risk management may take place inside all living creatures and at individual or at group level, thereby allowing for continuous improvement, culminating in cultural learning.



## 6 GLOSSARY

---

**4E cognition:** the view that mental processes are embodied, embedded, enacted and extended. The term was coined by Gallagher. See Rowlands, 2010, p. 3.

**Autopoiesis** has several meanings:

- The original definition of autopoiesis was based on mechanical assumptions and explicitly reject teleology. See Di Paolo, 2005, p. 433
- On the re-reading by Weber & Varela, Kant's notion of natural purposes was added, allowing autopoiesis to become a self-producing process. See Weber & Varela, 2002
- A new reading by Di Paolo defined autopoiesis as consisting of two processes: *self-producing* and *self-distinguishing*. These processes allow the organism to exist and survive itself, which is the outcome of autonomy. Autopoiesis is influenced by two factors: *robustness*, which is a positive factor; and *precariousness* which is a negative factor. Di Paolo, 2005, p. 434

**Autonomous system:** A network of co-dependent, precarious processes able to sustain itself and define an identity as a self-determined system. The same systemic relation can be found on many different levels. Examples include living cells, immune networks, sensorimotor flows of neural and bodily activity, habits, social institutions and so on. See: De Jaegher, Di Paolo, & Gallagher, 2010

**Cognitive offloading:** the use of physical action to alter the information processing requirements of a task so as to reduce cognitive demand. See: Risko & Gilbert, 2016

**Coupling:** The influence between a system's variables and another system's parameters. It can be mutual, for instance a person walking a dog held by a leash. See: De Jaegher et al., 2010

**Embodied cognition:** Many features of cognition are embodied in that they are deeply dependent upon characteristics of the physical body of an agent, such that the agent's beyond-the-brain body plays a significant causal role, or a physically constitutive role, in that agent's cognitive processing. See: Wilson & Foglia, 2017. Also see the [4E cognition](#) entry.

**Embedded cognition:** Mental processes have been designed to function only in tandem with a certain environment that lies outside the brain of the subject. In the absence of the right environmental scaffolding, mental processes cannot do what they are supposed to do, or can only do what they are supposed to do less than optimally. See: Rowlands, 2010, p. 3. Also see the [4E cognition](#) entry.

**Enacted cognition:** Mental processes are made up not just of neural processes but also of things that the organism does more generally—that they are constituted in part by the ways in which an organism acts on the world and the ways in which world, as a result, acts back on that organism. Rowlands, 2010, p. 3. Also see the [4E cognition](#) entry.

**Epistemic trust** is defined in terms of an individual's experience of communication from others, specifically, the ability to receive and treat new knowledge from others as personally relevant and therefore to be capable of modifying durable representational structures pertaining to self, others, and interpersonal relationships. Underpinning this capability is the consideration of the informant as a "trustworthy" source likely to communicate information that is generalizable and relevant to the self. See: P. Fonagy & Luyten, 2018

**Extended cognition:** Cognitive systems themselves extend beyond the boundary of the individual organism. On this view, features of an agent's physical, social, and cultural environment can do more than distribute cognitive processing: they may well partially constitute that agent's cognitive system. See: Wilson & Foglia, 2017. Also see the [4E cognition](#) entry.

**Empathy:** There are several definitions. Here the word *empathy* is used to refer to inner imitation as a means to gain knowledge about other minds. See: Stueber, 2019, paragraph 2.1

**Functionalism:** The doctrine that what makes something a thought, desire, pain (or any other type of mental state) depends not on its internal constitution, but solely on its function, or the role it plays, in the cognitive system of which it is a part. More precisely, functionalist theories take the identity of a mental state to be determined by its causal relations to sensory stimulations, other mental states, and behavior. There are several varieties of this theory: machine functionalism, psycho-functionalism and analytic functionalism. See: Levin, 2018

**Intentional stance:** The strategy of interpreting the behavior of an entity (person, animal, artifact, whatever) by treating it as if it were a rational agent who governed its 'choice' of 'action' by a 'consideration' of its 'beliefs' and 'desires.' See: Dennett, 2009

**Mirror neurons:** A particular class of neurons, originally discovered in the ventral premotor cortex of macaque monkeys (area F5) and then observed in the inferior parietal lobule (PF/PFG), which code goal-related motor acts like grasping. They become activated both when the subject performs a particular action, e.g. a precision grip, and when the subject observes another individual performing a similar goal-related action. There is evidence for a similar system in the human brain. See: Binder, Hirokawa, & Windhorst, 2009

**Propositional attitude:** the mental state of having some attitude, stance, take, or opinion about a proposition or about the potential state of affairs in which that proposition is true—a mental state of the sort canonically expressible in the form "S A that P", where S picks out the individual possessing the mental state, A picks out the attitude, and P is a sentence expressing a proposition. See: Schwitzgebel, 2019

**Rationality:** a system that does what it ought to do given its beliefs and desires. See Dennett, 2009. Not everyone agrees, particularly not on issues of agency and normativity. See: Rovane, 2004

**Regulated coupling:** Motivated changes that an agent makes to the constraints and parametrical conditions that influence the coupling between the agent and another system. The other system can be an agent that could itself be regulating the coupling, in which case we speak of a 'co-regulated' coupling. A simple example: moving closer to someone speaking in a low voice to hear him better. See: De Jaegher et al., 2010

**Privileged access:** The state of a person is mental if and only if that person's knowledge that he has the state is in principle superior, in some specified way, to the knowledge of that fact that is available to anyone else. See: Alston, 1971, citing G.E. Moore (1964) and F. Brentano (1964)

**Sense-making** has several meaning:

- Triangular externalism, involving two people and a shared public environment. Triangulation is a term from navigation. Davidson (1991) argues that if "two people note each other's reactions (in the case of language, verbal reactions), each can correlate these observed reactions with his or her stimuli from the world. The common cause can now determine the contents of an

utterance and a thought. The triangle which gives content to thought and speech is complete. But it takes two to triangulate. Two or, of course, more". See Davidson, 1991.

- Sense-making, in autopoietic enactivism, is described both as the meaningful outcome and the phenomenological function of a domain. As the meaningful outcome, it is constituted of both *life* and *action*. The organism does not just survive, but is taking active steps so it will still be there tomorrow. As phenomenological function it is constituted by physical regulation and self-monitoring processes. See Di Paolo, 2009, p. 9 and Froese & Di Paolo, 2011, p. 31

**Social cognition:** General term used to describe different forms of cognition about, or actions in regard to, agents or groups of agents, their intentions, emotions, actions and so on, particularly in terms of their relations to other agents and the self. See: De Jaegher et al., 2010

**Submentalising:** Submentalizing behavior looks as if it is controlled by thinking about mental states, but it is not. Submentalizing processes are domain-general cognitive processes that do not involve thinking about mental states but can produce in social contexts behavior that looks as if it is controlled by thinking about mental states. See: Heyes, 2014

## 7 BIOGRAPHY

---

- Alston, W. (1971). Varieties of Privileged Access. *American Philosophical Quarterly*, 83, 20. Retrieved from <http://www.jstor.com/stable/20009401>
- Andrews, K. (2012). *Do Apes Read Minds?: Toward a New Folk Psychology*. <https://doi.org/10.7551/mitpress/9283.001.0001>
- Andrews, K. (2018). Do Apes Attribute Beliefs to Predict Behavior? A Mengzian Social Intelligence Hypothesis. *The Harvard Review of Philosophy*, 25, 89–110. <https://doi.org/10/gd8gs9>
- Andrews, K. (2019, July 30). Naive Normativity. Retrieved 31 December 2019, from Kristin Andrews website: <https://kristinandrews.org/naive-normativity/>
- Andrews, K. (2020). Naïve Normativity: The Social Foundation of Moral Cognition. *Journal of the American Philosophical Association*, 61, 36–56. <https://doi.org/10/gg785g>
- Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 1164, 953–970. <https://doi.org/10/dphxn>
- Aristotle. (1984). *The Complete Works of Aristotle: The Revised Oxford Translation* J. Barnes, Ed. Princeton, New Jersey: Princeton University Press.
- Ashby, W. (1960). *Design for a Brain: The Origin of Adaptive Behaviour*. Springer Science & Business Media.
- Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological Reasoning in Infancy. *Annual Review of Psychology*, 671, 159–186. <https://doi.org/10/gfzfd3>
- Barlassina, L., & Gordon, R. M. (2017). Folk Psychology as Mental Simulation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* Summer 2017. Retrieved from <https://plato.stanford.edu/archives/sum2017/entries/folkpsych-simulation/>
- Binder, M. D., Hirokawa, N., & Windhorst, U. (Eds.). (2009). Mirror Neuron. In *Encyclopedia of Neuroscience* pp. 2378–2378. [https://doi.org/10.1007/978-3-540-29678-2\\_3514](https://doi.org/10.1007/978-3-540-29678-2_3514)
- Botterill, G. (1996). Folk Psychology and Theoretical Status. In P. Carruthers & P. K. Smith (Eds.), *Theories of Theories of Mind* pp. 105–118. Cambridge University Press.
- Botterill, G., & Carruthers, P. (1999). *The philosophy of psychology*. Cambridge, U.K. ; New York: Cambridge University Press.
- Byrne, R. W. (1996). Machiavellian intelligence. *Evolutionary Anthropology: Issues, News, and Reviews*, 55, 172–180. <https://doi.org/10/cngs5n>
- Carruthers, P. (2006). *The architecture of the mind: Massive modularity and the flexibility of thought*. Oxford : Oxford ; New York: Clarendon Press ; Oxford University Press.
- Carruthers, P. (2016). Two Systems for Mindreading? *Review of Philosophy and Psychology*, 71, 141–162. <https://doi.org/10/gfzffp>
- Davidson, D. (1991). Three Varieties of Knowledge. *Royal Institute of Philosophy Supplement*, 30, 153–166. <https://doi.org/10/dp4bcp>
- Davidson, D. (2001). *Essays on Actions and Events*. <https://doi.org/10.1093/0199246270.001.0001>

- De Jaegher, H. (2013). Embodiment and sense-making in autism. *Frontiers in Integrative Neuroscience*, 7. <https://doi.org/10/gf7tzm>
- De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, 14(10), 441–447. <https://doi.org/10.1016/j.tics.2010.06.009>
- Degenaar, J., & O'Regan, J. K. (2017). Sensorimotor Theory and Enactivism. *Topoi*, 36(3), 393–407. <https://doi.org/10.1007/s11245-015-9338-z>
- Degenen, C. (2018). *Cross-Cultural Perspectives on Personhood and the Life Course*. Springer.
- Dennett, D. C. (1987). *The intentional stance*. Cambridge, Mass: MIT Press.
- Dennett, D. C. (2009). Intentional Systems Theory. *The Oxford Handbook of Philosophy of Mind*. <https://doi.org/10.1093/oxfordhb/9780199262618.003.0020>
- Di Paolo, E. (2005). Autopoiesis, Adaptivity, Teleology, Agency. *Phenomenology and the Cognitive Sciences*, 4(4), 429–452. <https://doi.org/10/dsnkd6>
- Di Paolo, E. (2009). Extended Life. *Topoi*, 28(1), 9–21. <https://doi.org/10.1007/s11245-008-9042-3>
- Di Paolo, E. (2018). The Enactive Conception of Life. In *The Oxford Handbook of 4E cognition* Vol. 1. <https://doi.org/10.1093/oxfordhb/9780198735410.013.4>
- Di Paolo, E., & Thompson, E. (2014). *The Enactive Approach*. Retrieved from <https://mindrxiv.org/3vraf/>
- Dunbar, R. I. M., & Shultz, S. (2017). Why are there so many explanations for primate brain evolution? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1727), 20160244. <https://doi.org/10/gfzknd>
- Fenici, M., & Zawidzki, T. W. (2020). The origins of mindreading: How interpretive socio-cognitive practices get off the ground. *Synthese*. <https://doi.org/10/gg6mq3>
- Fodor, J. A. (1983). *The modularity of mind*. MIT press.
- Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind* Vol. 2. MIT press.
- Fonagy, P., & Luyten, P. (2018). Attachment, mentalization, and the self. In W. J. Livesley & R. Larstone (Eds.), *Handbook of personality disorders*. Retrieved from <https://discovery.ucl.ac.uk/id/eprint/1535395/>
- Fonagy, Peter, Luyten, P., Allison, E., & Campbell, C. (2019). Mentalizing, Epistemic Trust and the Phenomenology of Psychotherapy. *Psychopathology*, 52(2), 94–103. <https://doi.org/10/gg8sqd>
- Froese, T., & Di Paolo, E. A. (2011). *The enactive approach: Theoretical sketches from cell to society*. <https://doi.org/10.1075/pc.19.1.01fro>
- Gallagher, S. (2008). Inference or interaction: Social cognition without precursors. *Philosophical Explorations*, 1(3), 163–174. <https://doi.org/10/cf8pp4>
- Gallagher, S. (2012). *Phenomenology*. <https://doi.org/10.1057/9781137283801>
- Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind* First edition. Oxford, United Kingdom: Oxford University Press.

- Gallagher, S., & Hutto, D. D. (2008). 2. Understanding others through primary interaction and narrative practice. In J. Zlatev, T. P. Racine, C. Sinha, & E. Itkonen (Eds.), *Converging Evidence in Language and Communication Research* Vol. 12, pp. 17–38.  
<https://doi.org/10.1075/celcr.12.04gal>
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 212, 493–501. <https://doi.org/10/c3kjb6>
- Geurts, B. (2019). Communication as commitment sharing: Speech acts, implicatures, common ground. *Theoretical Linguistics*, 451–2, 1–30. <https://doi.org/10/gf4wzx>
- Goldberg, S. (2010). *Relying on others: An essay in epistemology*. Oxford: Oxford University Press.
- Goldberg, S. C. (2010). *Relying on Others: An Essay in Epistemology*. Oxford University Press.
- Goldman, A. I. (1979). What is justified belief? In *Justification and knowledge* pp. 1–23. Springer.
- Goldman, A. I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press, USA.
- Goldman, A. I. (2009). Mirroring, Mindreading, and Simulation. In J. A. Pineda (Ed.), *Mirror Neuron Systems: The Role of Mirroring Processes in Social Cognition* pp. 311–330.  
[https://doi.org/10.1007/978-1-59745-479-7\\_14](https://doi.org/10.1007/978-1-59745-479-7_14)
- Goldman, A. I. (2013). *Joint ventures: Mindreading, mirroring, and embodied cognition*. Oxford University Press.
- Goldman, A. I., & Jordan, L. (2013). Mindreading by Simulation: The Roles of Imagination and Mirroring. In *Joint Ventures: Mindreading, Mirroring, and Embodied Cognition*. Retrieved from <https://oxford.universitypressscholarship.com/view/10.1093/acprof:osobl/9780199874187.001.0001/acprof-9780199874187-chapter-6>
- Gopnik, A., & Wellman, H. M. (1992). Why the Child’s Theory of Mind Really Is a Theory. *Mind & Language*, 71–2, 145–171. <https://doi.org/10/fng26v>
- Gordon, R. M. (1996). ‘Radical’ Simulationism. In P. Carruthers & P. K. Smith (Eds.), *Theories of Theories of Mind*. Cambridge University Press.
- Gordon, R. M. (2007). Ascent routines for propositional attitudes. *Synthese*, 1592, 151–165.  
<https://doi.org/10/c4dm9w>
- Hawley, K. (2014). Trust, Distrust and Commitment: Trust, Distrust and Commitment. *Noûs*, 481, 1–20. <https://doi.org/10/gfvcrh>
- Heyes, C. (2012). New thinking: The evolution of human cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 3671599, 2091–2096. <https://doi.org/10/f36dtf>
- Heyes, C. (2014). Submentalizing: I Am Not Really Reading Your Mind. *Perspectives on Psychological Science*, 92, 131–143. <https://doi.org/10/f5twqk>
- Heyes, C. (2017). Apes Submentalise. *Trends in Cognitive Sciences*, 211, 1–2.  
<https://doi.org/10/ggfq7h>
- Heyes, C. (2018). *Cognitive Gadgets: The Cultural Evolution of Thinking*. Harvard University press.

- Heyes, C. (2019a). Cognition blindness and cognitive gadgets. *Behavioral and Brain Sciences*, 42, e187. <https://doi.org/10/gg6mrs>
- Heyes, C. (2019b, April 17). How culture works with evolution to produce human cognition. *Aeon / Psyche*. Retrieved from <https://aeon.co/essays/how-culture-works-with-evolution-to-produce-human-cognition>
- Heyes, C. (2020, August 21). *Personal correspondence*.
- Hopper, L. M., Waal, van der, E., & Caldwell, C. A. (2018). Celebrating the continued importance of “Machiavellian Intelligence” 30 years on. <https://doi.org/10.1037/com0000157>
- Hume, D. (1888). Part I: Of virtue and vice in general. Section 1: Moral distinctions not deriv'd from reason. In *Treatise of Human Nature/Book 3: Of morals*. Oxford: Clarendon Press. Wikisource.
- Hutto, D. D. (2005). Knowing What? Radical Versus Conservative Enactivism. *Phenomenology and the Cognitive Sciences*, 44, 389–405. <https://doi.org/10/fshz7g>
- Hutto, D. D. (2008). The Narrative Practice Hypothesis: Clarifications and implications. *Philosophical Explorations*, 113, 175–192. <https://doi.org/10/b244xt>
- Hutto, D. D. (2009). Lessons from Wittgenstein: Elucidating folk psychology. *New Ideas in Psychology*, 272, 197–212. <https://doi.org/10/fj9kbr>
- Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content*. Cambridge, Mass: MIT Press.
- Johnson-Ulrich, L. (2017). The Social Intelligence Hypothesis. In T. K. Shackelford & V. A. Weekes-Shackelford (Eds.), *Encyclopedia of Evolutionary Psychological Science* pp. 1–7. [https://doi.org/10.1007/978-3-319-16999-6\\_3100-1](https://doi.org/10.1007/978-3-319-16999-6_3100-1)
- Levin, J. (2018). Functionalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* Fall 2018. Retrieved from <https://plato.stanford.edu/archives/fall2018/entries/functionalism/>
- McGeer, V. (2007). The Regulative Dimension of Folk Psychology. In D. D. Hutto & M. Ratcliffe (Eds.), *Folk Psychology Re-Assessed* pp. 137–156. [https://doi.org/10.1007/978-1-4020-5558-4\\_8](https://doi.org/10.1007/978-1-4020-5558-4_8)
- McLeod, C. (2015). Trust. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* Fall 2015. Retrieved from <https://plato.stanford.edu/archives/fall2015/entries/rust/>
- Menkiti, I. (1984). Person and community in African Traditional Thought. In E. Wright (Ed.), *African Philosophy, an Introduction* p. 13. Lanham, MD: University Press of America.
- Michael, J., Christensen, W., & Overgaard, S. (2014). Mindreading as social expertise. *Synthese*, 1915, 817–840. <https://doi.org/10/ggxzjs>
- Morton, A. (2009). Folk Psychology. In *The Oxford Handbook of Philosophy of Mind*. Retrieved from <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199262618.001.0001/oxfordhb-9780199262618-e-42>
- Motsamai Molefe. (2019). *An African Philosophy of Personhood, Morality, and Politics*. Springer.

- Nichols, S., & Stich, S. P. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Oxford : New York: Clarendon ; Oxford ; Oxford University Press.
- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *The Behavioral and Brain Sciences*, 245, 939–973; discussion 973-1031.
- Pitt, D. (2000). *Mental Representation*. Retrieved from <https://plato.stanford.edu/archives/spr2020/entries/mental-representation/>
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 14, 515–526. <https://doi.org/10/ddvt4n>
- Quine, W. V. (1992). *Pursuit of truth* Rev. ed. Cambridge, Mass: Harvard University Press.
- Ratcliffe, M. (2009). There Are No Folk Psychological Narratives. *Journal of Consciousness Studies*, 166–8, 29. Retrieved from <https://www.ingentaconnect.com/content/imp/jcs/2009/00000016/F0030006/art00016>
- Ratcliffe, M. M. (2007). *Rethinking Commonsense Psychology: A Critique of Folk Psychology, Theory of Mind and Simulation*. <https://doi.org/10.1057/9780230287006>
- Ravenscroft, I. (2019). Folk Psychology as a Theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* Summer 2019. Retrieved from <https://plato.stanford.edu/archives/sum2019/entries/folkpsych-theory/>
- Rescorla, M. (2019). *The Language of Thought Hypothesis*. Retrieved from <https://plato.stanford.edu/entries/language-thought/#IndiMentExpr>
- Rice, T., & Lloyd Webber, A. (1970). Gethsemane (I Only Want To Say) Lyrics—Jesus Christ Superstar musical. Retrieved 15 August 2020, from All Musicals website: <https://www.allmusicals.com/lyrics/jesuschristsuperstar/gethsemaneionlywanttosay.htm>
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, 209, 676–688. <https://doi.org/10/gf5m3c>
- Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 42, 155–169. <https://doi.org/10/c8mscz>
- Rovane, C. (2004). Rationality and Persons. In A. R. Mele & P. Rawling (Eds.), *Oxford Handbook of Rationality*. Retrieved from <https://oxford.universitypressscholarship.com/view/10.1093/0195145399.001.0001/acprof-9780195145397-chapter-17>
- Rowlands, M. (2006). *Body language: Representation in action*. Cambridge, Mass: MIT Press.
- Rowlands, M. (2010). *The new science of the mind: From extended mind to embodied phenomenology*. Cambridge, Mass: MIT Press.
- Scholl, B. J., & Leslie, A. M. (1999). Modularity, Development and 'Theory of Mind'. *Mind and Language*, 141, 131–153. <https://doi.org/10/d9br6j>
- Schwitzgebel, E. (2019). Belief. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* Fall 2019. Retrieved from <https://plato.stanford.edu/archives/fall2019/entries/belief/>



- Shanahan, M. (2016). *The Frame Problem*. Retrieved from <https://seop.illc.uva.nl/entries/frame-problem/#FraProTod>
- Shea, N. (2018). *Representation in cognitive science* First edition. Oxford, United Kingdom: Oxford University Press.
- Shields, C. (2016). Aristotle. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* Winter 2016. Retrieved from <https://plato.stanford.edu/archives/win2016/entries/aristotle/>
- Slors, M. (2012). The Model-Model of the Theory-Theory. *Inquiry*, 555, 521–542. <https://doi.org/10/ggnr2w>
- Spaulding, S. (2020). What is mindreading? *WIREs Cognitive Science*, 113. <https://doi.org/10/gg8jhd>
- Stich, S., & Nichols, S. (1995). Second thoughts on simulation. *Mental Simulation: Evaluations and Applications*, 87–108.
- Strijbos, D. W., & de Bruin, L. C. (2012). Making Folk Psychology Explicit: The Relevance of Robert Brandom's Philosophy for the Debate on Social Cognition. *Philosophia*, 401, 139–163. <https://doi.org/10/b2gnsq>
- Stueber, K. (2019). Empathy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* Fall 2019. Retrieved from <https://plato.stanford.edu/archives/fall2019/entries/empathy/>
- Turner, S. (2018). The Belief-Desire Model of Action Explanation Reconsidered: Thoughts on Bittner. *Philosophy of the Social Sciences*, 483, 290–308. <https://doi.org/10/gg78xq>
- Ward, D., Silverman, D., & Villalobos, M. (2017). Introduction: The Varieties of Enactivism. *Topoi*, 363, 365–375. <https://doi.org/10.1007/s11245-017-9484-6>
- Weber, A., & Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, 12, 97–125. <https://doi.org/10/b6pxs3>
- Whiten, A., & Byrne, R. W. (1997). *Machiavellian intelligence II: Extensions and evaluations* Vol. 2. Cambridge University Press.
- Wilson, R. A., & Foglia, L. (2017). Embodied Cognition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* Spring 2017. Retrieved from <https://plato.stanford.edu/archives/spr2017/entries/embodied-cognition/>