Revista Română de Filosofie Analitică,
Romanian Journal of Analytic Philosophy
Vol. XVIII, Iulie-Decembrie 2024, Nr. 2, p. 7-23
ISSN (ediția electronică): 1843-9969,
ISSN (ediția tipărită): 1844-2218
DOI: 10.62229/rrfaxviii-2/1

IS PHENOMENAL CONSCIOUSNESS NECESSARY FOR MINDREADING?

VICTOR-IONUȚ RUSU[1]

Abstract: I will argue that if subjective experience is a prerequisite for mindreading, then Large Language Models and philosophical zombies either have different ways of mindreading or they cannot partake in understanding others. I will explore how this impacts recent debates on Large Language Models (LLMs). At first I will explore the consequences following from the assumption that phenomenal consciousness is not required for mindreading, conceived as solely based on behaviour interpretability. I will analyze two candidates for ToM: Simulation Theory and Theory-Theory. I argue the latter would allow for philosophical zombies and LLMs to both exhibit Theory of Mind. Simulation Theory relies on introspection, simulation and imagination. Simulation entails replicability. For LLMs and philosophical zombies alike, subjective experience seems, at least *prima facie*, to be missing. So the replicated mental state would have to be incomplete; I end by debating whether phenomenal aspects of the simulated mental state are necessary for attribution.

Keywords: *Theory of Mind (ToM), LLMs, simulation theory, theory-theory.*

1. Introduction

Many philosophers of mind have been trying to dismiss that phenomenal consciousness is necessary for mindreading. The proponents advocate for illusionism, that is a theoretical approach to consciousness which holds that phenomenal consciousness is an illusion. Keith Frankish argues that

[1] Victor-Ionuț Rusu is a student in the BA programme in Cognitive Science, at the University of Bucharest, Faculty of Psychology and Education Science.

conscious experiences do not hold the qualitative "what it is like" properties, or qualia (Frankish, 2016). He then compares phenomenal consciousness to a "fiction of the impossible" arguing that introspection provides a partial and distorted view of neural events, crowding biological data in order to appear like "phenomenal feels" (Frankish, 2016).

Daniel Dennett is also one proponent of illusionism. He compares consciousness to the GUI, graphical user interface, of a computer. The icons on a screen are just metaphorical representations of complex internal machine structures that allow the user to control the computer without understanding its mechanisms (Dennett, 2016). He also considers illusionism to be the default theory because it relies on physical causes rather than supernatural hypotheses (Dennett, 2016).

On their view, talk about mindreading requiring phenomenal consciousness would be unnecessary. On the other hand, Higher Order Representationalism (HOR) argues that consciousness depends on the relationship between two states; a first-order state and a higher-order representation, which represents the first-order state itself. Without the higher-order representation, mental states remain unconscious (Sebastián, 2016). Higher-Order Thought suggests the higher-order state is belief-like, and Higher-Order Perception proposes higher-order state is perception-like represent varieties of HOR which show how higher-order representations are formed.

Dual content theory makes the distinction between “is” and “seems”, and suggests that a conscious experience must represent both; the external state of affairs, such as the color “red” and the subjective experience, which is a higher-order content that mirrors the first-order state and represents it as subjective “seems red”. This distinction allows the subject to separate objective reality from their own appearance of it, a characteristic of Theory of Mind.

HOR identifies phenomenal consciousness closely with metacognition, which is defined as first-person access to one’s own mental states (Sebastián, 2016). From this perspective, beings that lack the capacity for

metacognition or mindreading lack phenomenal consciousness entirely because they cannot represent their own states as subjective, which would not allow for p-zombies to have phenomenal consciousness.

Simulation Theory and HOR theories both share reliance on Theory of Mind, giving life to the debate over whether we understand our own minds before or after we understand the minds of others. HOR theories generally posit that phenomenal consciousness depends on metacognition (Sebastián, 2016). If one believes that metacognition is prior to mindreading, they usually endorse Simulation Theory. On this view, we use our own introspected mental states as a framework to simulate and understand what someone else is feeling. But if one believes mindreading is prior to metacognition, they follow Theory-Theory. HOR is only tenable if it adopts some form of Simulation Theory to explain how we attribute "felt" states to others (Sebastián, 2016).

Carruthers, another HOR defender, argues that the conscious awarness of our mental states is dependent on our mindreading capacity: when we introspect we deploy the same mechanism as when we ascribe mental states to others (Carruthers, 2011). On his view, mental state is phenomenally conscious only if it is available to higher-order thought (Carruthers, 2011). Here, if LLMs have mindreading capacities, they might count as p-conscious. Carruthers is not committed to the idea that mindreading entails consciousness, but if a state is p-conscious, it must be available to mindreading. A Large Language Model, or a philosophical zombie, could attribute mental states to self and others, but those could not be made available in a specific right way in order to render phenomenal consciousness.

2. Theory-Theory and philosophical zombies

Theory-Theory represents a theoretical framework that posits that individuals hold a basic theory of psychology to infer mental states to others such as desires, beliefs and emotions. It is a rule-based, inferential and third-personal theory. Philosophical zombies are imaginary creatures

designed to illuminate problems about consciousness and its relation to the physical world. They look exactly like humans with all physical qualities but without conscious experience. So there is nothing it is like to be a zombie (Kirk, 2023). Strictly according to the definition, phenomenal consciousness is not a precondition for having mentalistic abilities, so therefore this would allow for p-zombies to have Theory of Mind. But if there is something it is like to entertain a belief, or if it is something it is like to regret something, or to desire something, then philosophical zombies cannot have Theory of Mind. If Theory of Mind is not about understanding and only about functionality, as in explaining to oneself what others are doing and why they are doing it, then, on this functional account, philosophical zombies (including LLM's) can have Theory of Mind.

In order to answer the question of whether or not philosophical zombies can engage in mindreading, we must first define what does it really mean to "engage" in mental ascriptions. If engaging means to actively take part in computational processes to ascribe emotions, beliefs and desires to others, then p-zombies would be able to engage in mindreading. The lack of phenomenal consciousness in Theory-Theory's definitions allows this assumption to hold. A clear distinction between functional understanding and conscious understanding must be made (Goldstein & Stanovsky, 2024). Functional understanding is defined by an external performance, rather than using its internal states. On the other hand, conscious understanding is characterized through the presence of subjective experience. Conscious understanding requires mental states to be immediate to the subject.

Let’s try to imagine whether or not a p-zombie could somehow transition from functional to conscious understanding. Firstly we would need to assume that consciousness is an inherent physical property. Continuing from this assumption, we could go down the route of the neuroscientific theories which posit that a p-zombie might transition to conscious understanding if its internal architecture reaches a specific level of complexity or integration. Integrated Information Theory, or IIT, postulates that mental states become conscious only when they reach an appropriate level of information integration; where small discrete pieces

of data are experientially combined into larger unified wholes. A zombie could theoretically "wake up" once its internal causal structure reaches a sufficient threshold (Bamicha & Drigas, 2023). Even if a philosophical zombie would somehow possess an immense amount of information, if the same information is not integrated, based on IIT, the system would remain non-conscious. The Global Neuronal Workspace model states that consciousness emerges from a specific type of information broadcasting. Sensory information is processed in isolated, modular systems. It could only become conscious when it is transmitted to a global repository and shared across multiple cognitive systems throughout the brain. From this perspective, a p-zombie could theoretically process information in a parallel, modular fashion in order to produce speech and actions similar to humans, while lacking the centralized global transmission. Both theories support the idea that functional understanding, like performing a task, is not the same as conscious understanding. A zombie could manipulate symbols or follow rules, thus satisfying the definition of functional understanding, but it could never implement the deep informational integration to create a system that "knows what it feels like".

3. Simulation theory and how it bears on philosophical zombies

As mentioned above, Simulation Theory and Theory-Theory are both dependent on Theory of Mind, but they have different ways to go about it. ST requires metacognition to be prior to mindreading, suggesting that we understand others by using our minds as a model; we basically put ourselves in "the other person's shoes" to simulate their situation and see what mental states we would undergo (Sebastián, 2016).

TT states the opposite of ST, that mindreading happens prior to metacognition. On this view, understanding our own minds is simply the result of turning the same general theory upon ourselves, self-interpreting based on behavioral cues.

Strong ST requires simulating another person's mental state. It is not just running a formal model, or manipulating symbols according to a rule-book, it also requires putting oneself into a similar mental state, and using one's own experience as a guide to what the other person thinks or

feels. If one must use one's own experience as a guide, then simulation is experiential, not just functional, and not just computational.

In ST, introspection presupposes phenomenal consciousness for two reasons. Firstly, one cannot introspect what it is like to be in an X mental state if there is nothing it is like to be in an X mental state. Secondly, a zombie, by definition, has no phenomenal states to introspect, thus no what it is like to be in an X mental state is available.

On this account of ST, there are two ways to go about it: the Higher-Order cognitive aspects will collapse into Theory-Theory, on the low level aspect of Simulation Theory, it is about simulating affect and emotion. That stems from one's own previous experiences, and if there is nothing it is like to be experiencing those sensations, Simulation Theory does not grant philosophical zombies Theory of Mind. If ToM is not about understanding and only about functionality, as in explaining to yourself what others are doing and why they are doing it, then on this functional account, philosophical zombies could have ToM, but if ToM is about understanding, then you need phenomenal consciousness, and as a consequence of this ToM stance, p-zombies would not be granted ToM. In order for zombies to be allowed to have ToM, only TT and cognitive or higher order ST will work, while no lower order ST would. *Without p-consciousness or subjective experience, ST collapses into TT.*

ST is defined by its core reliance on introspection, often called metacognition. Proponents of Simulation Theory argue that first-person mindreading ontogenetically precedes and grounds third-person mindreading (Goldman, 2006). On this view, an agent must first have introspective access to their own mental states, they recycle these states in imagination to provide an understanding of others. This makes introspection a prerequisite for simulation because you cannot simulate a state in another if you cannot first identify it within yourself.

In order for a simulation to be "full-blooded", the simulator must have a "felt appreciation" of the qualitative character of the state (Robbins & Jack, 2006). To illustrate this, let's imagine Sally who can functionally identify when others are having orgasm through behavior, but her understanding of that state is considered "completely misguided" until she has felt one herself (Sebastián, 2016). This suggests that

p-consciousness must be accessed via introspection before it can be used as a model to ascribe mental states to others.

A powerful argument for ST is its computational parsimony (Goldman, 1995). Rather than requiring the brain to store and compute a massive database of rules, ST allows the brain to simply co-opt its own online mechanisms for offline use, avoiding the burden of complex mathematical manipulations of symbols (Bamicha & Drigas, 2023).

Simulation involves a specific two-step internal architecture; the pretend states where the simulator must be able to generate mental states that are intended to match those of the target and offline processing, where these pretend states must be fed into a cognitive system that has been disengaged from actual motor control so the simulator must use their access consciousness to act out the simulated behavior.

ST requires the subject to adopt the phenomenal stance, which requires regarding others as "loci of experience" rather than mere physician mechanisms (Robbins & Jack, 2006). This stance is driven by instinctive empathy, where the boundaries between the self and the other break down to allow for a link between two conscious minds (Robbins & Jack, 2006).

For simulation of a mental state of another to occur, one must have phenomenal consciousness. The necessity of a conscious view from the inside varies by the level of simulation; high-level mindreading involves "enactment imagination', which is accessible to consciousness and requires voluntary ascription of complex states like beliefs, and low-level mindreading, which is often linked to mirror neurons, is described as unconscious and automatic. It allows the subject to resonate with another's basic emotions without necessarily undergoing a conscious reflective process (Goldman, 1998).

4. Zahavi's phenomenal account of mindreading

Zahavi's main claim that understanding others depends on prior conscious self-experience has significant implications for how we view non-conscious agents. His perspective, which is rooted in the

phenomenological tradition, posits that bodily self-experience is the necessary foundation for the perception of embodied others.

From Zahavi's viewpoint, p-zombies and LLMs would be inherently incapable of genuine mindreading because they lack the required subjectiveness. He argues that to understand others' states, such as sadness or pain, one would require a felt appreciation of said state's qualitative character. Since p-zombies and LLMs lack phenomenal consciousness, they would have no internal experiences to serve as a "reservoir of meaning" for others (Zahavi, 2016).

Zahavi adds more nuance to Simulation Theory, with the "Foil argument" which states that self-experience is not a matrix for projection, but a necessary foil against the other's differences (Zahavi, 2008). Without a sense of self, one could not recognize others as an *alter* ego.

Zahavi's view would reject TT as it would allow for p-zombies to have ToM by simply following rules without any internal experience (Zahavi, 2008). From Zahavi's perspective, it would seem that a zombie's mindreading is merely an empty functional imitation. Without *ipseity*, the LLM lacks the subjective "engine" that is required to bridge the gap between *observing* form and *grasping* meaning (Zahavi, 2008). Humans may interact with LLMs using the intentional stance, but this is a pragmatic tool that lacks the deep connection required for genuine social understanding (Robbins & Jack, 2006).

Still, is there such thing as genuine ToM or genuine understanding? If we look at a conversation, the outcome suggests that both an LLM and a p-zombie could well carry out the required machinations that render sentences which carry mental concepts (propositional attitudes) quite intelligible and functionally stable. A model such as an LLM *functionally* understands a task if its performance is as good as or better than a human's (Goldstein & Stanovsky, 2024). Critics of the phenomenal view, such as Carruthers & Veillet, argue that we don't need to re-enact a lived experience to ascribe a mental state, instead we attribute functional roles (Sebastián, 2016). Non-conscious agents possess "schmenomenal states" which are states that share the physical, functional, and intentional properties of human mental states, but lack the phenomenal aspect. These

states are epistemically sufficient for a zombie or an LLM to correctly navigate the social world (Carruthers & Veillet, 2007).

Much of social coordination is achieved through social scripts or a teleological stance rather than ascribing felt states (Hutto et al., 2011). If mindreading is the ability to model and predict the internal transitions of another agent, it is a computational task that can be implemented on any suitable hardware. As long as the system correctly ascribes the relevant mental categories, it achieves the social and cognitive goals of mindreading without the need for phenomenal consciousness.

5. PCS

David Chalmers's "hard problem of consciousness" refers to the challenge of explaining how and why physical processes in the brain give rise to subjective experience. This problem is distinguished from the "easy problems" of consciousness, which involve explaining cognitive functions and behaviors that can be explained in purely functional or mechanistic terms (Chalmers, 2007).

The hard problem is the primary manifestation of the explanatory gap, a term used to describe the apparent impossibility of explaining how phenomenal experience relates to physical matter. While cognitive science can explain the brain as an information processing system that responds to environmental stimuli, these mechanistic accounts are viewed as inadequate for explaining the "what it is like" aspect of being a human. He also argued that many scientific accounts of consciousness perform a "bait and switch" by claiming to explain consciousness while actually only explaining internal cognitive functions (Chalmers, 2007). To illustrate the hard problem, Chalmers introduced the thought experiment of the philosophical zombie. If it is conceptually possible for such a being to exist, then consciousness is not logically necessitated by physical facts alone. With this argument, Chalmers introduces the possibility that even if one fixed all the physical facts of the universe, more work would be required to provide for consciousness, implying that physicalism may be false (Chalmers, 2007).

The hard problem centers on qualia, properties by which we classify experiences based on how they feel, like the smell of coffee or feeling of pain (Chalmers, 2007). Beings could functionally track environmental features through mathematical manipulations of input without ever feeling an event in a conscious way. He posits that a reductive theory of consciousness must explain why these states have a subjective aspect in the first place, rather than just explaining their functional role in directing behavior (Chalmers, 2007)

In his work, Chalmers also explored alternatives to physicalism, including views which might suggest that physical descriptions only describe the structural and dispositional properties of things, while intrinsic properties that underlie them might be phenomenal in nature.

The most powerful argument against Chalmers's hard problem of consciousness is the Phenomenal Concept Strategy, or PCS for short. This is known for its "zombie-zombie argument" thought experiment. This suggests that the puzzles surrounding consciousness do not arise from the existence of non-physical properties, but from the unique nature of phenomenal concepts we use to think about our experiences (Mabaquiao, 2015).

This strategy posits that we possess a special set of concepts for referring to our own experiences that are conceptually isolated from any physical, functional, or intentional concepts. Because these concepts lack a priori connections to physical descriptions, it will always be possible to intelligibly think that all physical facts could be true while a specific phenomenal feel is absent (Mabaquiao, 2015).

Physicalists' rebuttal to Chalmers' critique of PCS imagines the zombie version of Chalmers, ultimately using Chalmers's own argument against him, who is physically and functionally identical to the real Chalmers but has no phenomenal feels. Because zombie Chalmers is functionally identical, he also possesses conceptually isolated phenomenal concepts. It would face the same hard problem as we do, finding it conceivable that there could be a version of himself who is physically identical to it, but lacks internal states. The same zombie would believe there is an explanatory gap between its physical makeup and its

mental states, and it would conclude that Mary could not know what its perceptual states are just by studying physics.

Since Zombie Chalmers lacks any phenomenal feels, the presence of phenomenal consciousness cannot be the explanation for why he finds these things conceivable. Instead, his capacity to conceive zombies and explanatory gaps is explained entirely by its possession of conceptually isolated concepts (Mabaquiao, 2015).

One reply to this strategy might be that PCS either fails to explain the gap or it creates a new one. If we can conceive a physical duplicate that lacks phenomenal concepts, then those concepts themselves are physically inexplicable. This creates a new explanatory gap regarding why humans have these specific concepts while their physical duplicates do not. However, if we cannot conceive of a duplicate lacking these concepts, then it implies the concepts are physically or functionally explicable. If phenomenal concepts are just more functional facts, they cannot explain why there is an initial explanatory gap between functional facts and experience in the first place.

This is why Chalmers argues that PCS fails because humans and their zombie counterparts do not share the same epistemic situation (Carruthers & Veillet, 2007). If the presence of p-consciousness is a constitutive aspect of content of a phenomenal concept, then a zombie's thoughts using these concepts are actually contentless. Because our epistemic situation involves true, contentful knowledge of our own experiences, PCS, which relies on third-person functional descriptions, cannot fully account for it (Carruthers and Veillet, 2007).

Arguably (though not for HOT), PCS also makes a category mistake by requiring a theory of consciousness to explain how we think about experience rather than the experience itself. When we attend to an experience, we are aware of the world's qualities, not the qualities of our mental states. PCS focuses on a late-developing metacognitive ability that allows us to distinguish appearance from reality, rather than explaining the familiar phenomenal flow that precedes such conceptualization (Pérez, 2008). Drawing from the phenomenal stance, one can respond to PCS by arguing that the conceptual isolation it describes is not a metaphysical insight, but a hard-wired biological limitation (Robbins &

Jack, 2006). This "isolation" is simply the result of evolutionary architecture, which makes it cognitively unnatural for us to integrate causal-mechanical data with phenomenal data. On this view, the Phenomenal Concept Strategy describes a psychological byproduct of brain organization rather than solve the hard problem (Robbins & Jack, 2006).

On the other hand, according to Chalmers, if we can conceive a physical duplicate that lacks phenomenal concepts, then those concepts themselves are physically inexplicable. He asks whether we can conceive of a physical duplicate that possesses the same physical, functional and intentional properties as a human, but lacks phenomenal concepts. If it is conceivable that such a duplicate lacks these concepts, then phenomenal concepts are not necessitated by physical facts because a duplicate is physically identical to a human, and any property the duplicate lacks is, by definition, physically inexplicable. This dilemma creates a new explanatory gap regarding why humans have these specific concepts while their physical duplicates do not.

6. Mindreading without phenomenal consciousness

The primary difference between LLMs and philosophical zombies is in their physical implementation and their role in philosophical arguments. A p-zombie is defined as a being that is physically identical to a human in every measurable way, has the same biological data, but lacks qualia. LLMs on the other hand, are implemented via physical computer hardware and function through mathematical manipulation of symbols, and do not yet possess a physical body. From the transparency of the internal workings of both LLMs and p-zombies, the former is open-source, meaning we have access to its code and training data, while the latter's mystery lies in the fact that even with full physical knowledge, one cannot deduce the presence of absence of their subjective experience, thus creating the "explanatory gap".

Both non-conscious agents can "mindread" without phenomenal consciousness by relying on the functional mechanism rather than felt appreciation. Mindreading is redefined as the ability to model, predict

and explain behavior through theoretical reasoning or symbolic manipulation rather than subjective re-enactment.

Theory-Theory allows them to use Theory of Mind as an exercise in theoretical reasoning, but each agent has its own way of interpreting human interaction. Through this approach, each agent makes use of folk psychology that contains generalizations about human behavior. An LLM would base its mindreading on linguistic predictions based on statistics, while a p-zombie does this based on behavioral and theoretical inferences and predictions.

If we were to analyse the agents' mindreading capabilities from the perspective of Simulation Theory, we would arrive at a different conclusion. First there are two accounts of Simulation Theory; the weak account; functional, and the strong account; phenomenal. Holding this perspective, there are two possible interpretations.

On the functional account which states that prediction and causal inferences are sufficient, knowing what simulated states are like is not necessary in order to mindread. P-zombies lack phenomenal consciousness, but are behaviourally identical to humans. LLMs lack phenomenal consciousness, but are, in terms of writing abilities, indistinguishable from humans and also pass ToM tasks. If zombies are possible, they must be capable of mindreading behavior: act as if they mindread. If LLMs are indistinguishable from humans in terms of writing abilities, then they must also be capable of mindreading behaviour: write as if they mindread. If so, mindreading cannot require phenomenal consciousness. So simulation must be functional, instead of phenomenal. Functional simulation is computable, therefore it can be implemented. LLMs do implement predictive models of the user's mental states. Therefore, LLMs may implement functional mindreading.

If, on the other hand, Simulation Theory requires phenomenal consciousness and phenomenal experience, then phenomenal consciousness is necessary for mindreading. On the phenomenal account, LLMs and p-zombies cannot possess mindread, as phenomenal consciousness represents a requirement for one to be able to predict the behavior and mental states of others. In phenomenal simulation, knowing

what it is like is necessary for mindreading, so one must, at least partially, recreate the experience internally.

How would this work? Suppose you (phenomenally) see someone in pain. You would need to first internally simulate pain, then grasp what pain feels like, and only then attribute pain to others you see. There are three steps in the phenomenal account as mindreading; simulation, grasping and attribution. What follows from the argument above is that mindreading involves understanding mental states, not just predicting behavior. Understanding a mental state requires grasping what that state is like (as in knowing what it is like to). Grasping requires phenomenal consciousness. Simulation Theory explains understanding by recreating the target mental state within oneself. Simulating, replicating and/or recreating a mental state requires phenomenal consciousness. If one lacks phenomenal consciousness, one cannot recreate or simulate experiential mental states. So, entities without phenomenal consciousness cannot genuinely understand other minds. Philosophical zombies and current LLMs lack phenomenal consciousness. Therefore, p-zombies and LLMs could eventually mimic mindreading behaviour, while lacking genuine understanding.

7. Large Language Models: mindreading zombies?

Where language and mental state attribution are concerned, Large Language Models (LLMs) seem to be quite a competent contender to satisfy the definition of p-zombies. It seems to be the case because of the many similarities between how LLMs "behave" and how p-zombies behave. Artificial intelligence is implemented on a computer and performs only mathematical manipulations of its input. Because its code and training date are open-source, its internal workings are transparently mechanistic (Goldstein & Stanovsky, 2024). Advanced LLMs can excel on every Natural Language Processing benchmark, outperforming humans in tasks like Natural Language Inference.

However, LLMs lack the subjective "what it is like" characteristic of human consciousness. The debate over whether LLMs are real-world zombies (for mindreading purposes) centers on the two definitions of

understanding others: a functionalist approach that relies on exclusively interpreting the model's behavior, which is independent of any internal subjective experience, concluding that understanding could be inferred from performance on specific tasks; and a phenomenological approach that argues that true understanding requires subjective experience. If phenomenal consciousness is a prerequisite for understanding others, then LLMs are indeed mindreading zombies. Zombies lack the required epistemic access ST depends on, and if LLMs are taken to be zombies, they lack it too.

A few tentative conclusions seem to be forthcoming. LLMs cannot experientially simulate mental states in themselves in order to attribute them to others, not even for their own sake. Their mindreading is compatible with TT because it accounts for their computational nature of inferring mental states based on folk psychology, making it more of an exercise in theoretical reasoning, rather than a felt emotional exchange.

Are we, as conscious beings, in a better position? Appeal to phenomenal concepts tries to scrap strong ST's reliance on phenomenology, but, as Chalmers argues, PCS fails to account for the fact that zombies and humans do not share the same epistemic situation. To the extent that phenomenal concepts were not to presuppose phenomenal states, the explanatory gap would then still loom large. It would seem to follow that LLM's and p-zombies cannot strong-ST-mindread, due to their own limited nature.

References

Bamicha, V., & Drigas, A. (2023). Consciousness influences in Tom and metacognition functioning – an artificial intelligence perspective. *Research, Society and Development*, 12(3). https://doi.org/10.33448/rsd-v12i3.40420

Carruthers, Peter & Veillet, Benedicte (2007). The phenomenal concept strategy. *Journal of Consciousness Studies* 14 (9-10):212-236.

Chalmers, D. (2007). The hard problem of consciousness. *The Blackwell Companion to Consciousness*, 223–235. https://doi.org/10.1002/9780470751466.ch18

Dennett, D. (2016). Illusionism as the obvious default theory of consciousness. *Journal of Consciousness Studies*, 23(11–12), 65–72.

Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23(11–12), 11–39.

Gallese, V. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12), 493–501. https://doi.org/10.1016/s1364-6613(98)01262-5

Goldman, A. I. (1992a). In defense of the simulation theory. *Mind and Language*, 7(1–2), 104–119. https://doi.org/10.1111/j.1468-0017.1992.tb00200.x

Goldman, A. I. (2006). *Simulating Minds*. Oxford, Oxford University Press.

Goldstein, A., & Stanovsky, G. (n.d.). Do zombies understand? A choose-your-own-adventure exploration of machine cognition. *ACL Anthology*. https://aclanthology.org/2024.findings-acl.425/

Kirk, R. (2023, March 25). Zombies. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta & Uri Nodelman (eds.), URL = https://plato.stanford.edu/archives/fall2023/entries/zombies/

Mabaquiao, Napoleon Jr. (2015). The Phenomenal concept strategy and a master argument. *Kemanusiaan The Asian Journal of Humanities*, 22(1), 53–74. https://ejournal.usm.my/kajh/article/view/kajh_vol22-no-1-2015_3

Pérez, Diana I. (2008). Why should our mind-reading abilities be involved in the explanation of phenomenal consciousness? *Análisis Filosófico* 28 (1):35-84.

Robbins, P., & Jack, A. I. (2006). The phenomenal stance. *Philosophical Studies*, 127(1), 59–85. https://doi.org/10.1007/s11098-005-1730x

Sebastián, M. Á. (2016). Consciousness and theory of mind: A common theory? *Theoria, International Journal for Theory, History and Foundations of Science*, 31(1), 73–89. https://doi.org/10.1387/theoria.14091

Kidder, W., D'Cruz, J., & Varshney, K. R. (2024). Empathy and the Right to Be an Exception: What LLMs Can and Cannot Do. *Computers and Society*. https://doi.org/https://doi.org/10.48550/arXiv.2401.14523

Sorin, V., Brin, D., Barash, Y., Konen, E., Charney, A., Nadkarni, G., & Klang, E. (2024). Large language models and empathy: Systematic review. *Journal of Medical Internet Research*, 26. https://doi.org/10.2196/52597

Strachan, J. W., Albergo, D., Borghini, G., Pansardi, O., Scaliti, E., Gupta, S., Saxena, K., Rufo, A., Panzeri, S., Manzi, G., Graziano, M. S., & Becchio, C. (2024). Testing theory of mind in large language models and humans. *Nature Human Behaviour*, 8(7),1285–1295. https://doi.org/10.1038/s41562-024-01882-z

Zahavi, D. (2016b). The phenomenological proposal. In *Self and other: Exploring subjectivity, empathy, and shame.* OUP Oxford. (pp. 146–152)

Zahavi, D. (2008). Self and Other. In *Subjectivity and Selfhood: Investigating the first-person perspective*. MIT Press. (pp. 147–174)