Эпистемология и философия науки
2026. Т. 63. № 1. С. 56–62
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 1, pp. 56–62
DOI: 17

Commentary on P. Perconti, “Rethinking Subjectivity: The Social Roots of Consciousness”

Massimo Marraffa –
PhD, Professor.
Department of Philosophy, Communication and Performing Arts.
Roma Tre University.
Via Ostiense, 234-236,
00146 Rome RM, Italy;
e-mail: massimo.marraffa@uniroma3.it

Cristina Meini –
PhD, Professor.
Department of Sustainable Development and Ecological Transition, University of Eastern Piedmont.
Piazza S. Eusebio 5,
13100 Vercelli, Italy;
e-mail: cristina.meini@uniupo.it

The authors concur with Pietro Perconti’s thesis that conceptual self-awareness is a by-product of social cognition. However, the authors question some aspects of his theory, particularly the dual-process account dividing mindreading into System 1 (low-level simulation) and System 2 (high-level simulation). This distinction is problematic because it simplifies the complex nature of mindreading, which involves both Theory Theory (TT) and Simulation Theory (ST). High-level mindreading, which includes cognitive processes like perspective-taking and counterfactual imagination, contrasts sharply with low-level mindreading processes, which are automatic and involve simple mirroring. Furthermore, the Simulationist model’s neo-Cartesian view on simulation as projection is inconsistent with a “mindreading first” approach, which suggests that self-knowledge develops in parallel with knowledge of others. Additionally, the authors argue that conceptual self-consciousness cannot solely be explained through social interactions and that other mechanisms, like adaptive traits, play significant roles. The social cognition framework must integrate these perspectives and acknowledge the developmental origins and varied functions of consciousness.

Keywords: self-consciousness, consciousness, dual-process theory, simulation theory, social cognition

Комментарий к статье П. Перконти «К переосмыслению субъективности: социальные корни сознания»

Массимо Марраффа –
доктор философии, профессор.
Департамент философии, коммуникации

Авторы согласны с тезисом Пьетро Перконти о том, что концептуальное самосознание является побочным продуктом социального познания. Однако они ставят под сомнение некоторые аспекты его теории, в частности представление о двух процессах, разделяющее чтение разума (mind-reading)

и исполнительского искусства,
Третий университет Рима.
Via Ostiense, 234-236, 00146 Рим, Италия.
e-mail: massimo.marraffa@uniroma3.it

Кристина Мейни –
доктор философии, профессор.
Департамент устойчивого развития и экологического перехода, Университет Восточного Пьемонта.
Piazza S. Eusebio 5, 13100, Верчелли, Италия.
e-mail: cristina.meini@uniupo.it

на Систему 1 (низкоуровневая симуляция) и Систему 2 (высокоуровневая симуляция). Это разделение является проблематичным, так как упрощает сложный характер чтения разума, который включает как теорию теории (TT), так и теорию симуляции (ST). Высокоуровневое чтение разума, включающее когнитивные процессы, такие как принятие чужой точки зрения и контрфактическое воображение, резко контрастирует с низкоуровневыми процессами, которые автоматичны и включают простое отражение. Кроме того, неокартизианский взгляд симуляционистской модели на симуляцию как проекцию не согласуется с подходом "сначала чтение разума", который предполагает, что самопознание развивается параллельно с познанием других. Дополнительно, авторы утверждают, что концептуальное самосознание не может объясняться исключительно через социальные взаимодействия и что другие механизмы, такие как адаптивные черты, играют значительную роль. Структура социального познания должна учитывать эти перспективы и учитывать эволюционные истоки и разнообразные функции сознания.

Ключевые слова: самосознание, сознание, теория дуального процесса, теория симуляции, социальное познание

We fully agree with Pietro Perconti's main thesis that conceptual self-awareness is a by-product of social cognition. However, we believe that some of the ingredients in his recipe can be questioned.

A distinctive feature of Perconti's theory of self-consciousness is that "it is based on a dual-process account about the overall architecture of brain processing". Accordingly, the author distinguishes between a System1 (S1) Mindreading and a System2 (S2) Mindreading. The first is "a low-level simulation process concerning the understanding of the aim of an action. This kind of implicit understanding is largely governed by the mirror system in the brain". The second is "a high-level simulation activity of projection consisting in cognitive processes such as the taking of a different point of view from one's own, and the counterfactual imagination". However, we believe that the distinction between low-level and high-level simulation processes is the wrong basis to build a "mindreading priority account" of introspective self-consciousness on. Let us see why.

(1) Until the mid-1980s the debate on the nature of mindreading was a debate between the different variants of Theory Theory (TT). But in 1986, Robert Gordon and, independently, Jane Heal gave life to an alternative, the Simulation Theory (ST).

Gradually, however, a consensus emerged to the effect that mindreading involves both TT and ST. For example, Goldman [Goldman, 2006] grants a variety of possible roles for theorizing in the context of what he calls 'high-level mindreading'. This is the imaginative simulation that is subject to voluntary control, is accessible to consciousness, and involves the ascription of complex mental states such as propositional attitudes.

High-level mindreading is contrasted with low-level mindreading, which is unconscious, hard-wired, involves the attribution of structurally simple mental states such as face-based emotions (e.g., joy, fear, disgust), and relies on simple imitative or mirroring processes (e.g., [Goldman and Sripada, 2005]).

In the same simulationist vein, Stueber [Stueber, 2019, §2.1] proposed that 'primary empathy' or 'mirroring' (a low-level form of empathy regarded as a precursor to Goldman's notion of low-level mindreading) would find its neural realisation in the activity of mirror neurons. In our opinion, however, the activity of mirror neurons can at best constitute the implementation of an automatic low-level mechanism that generates a response of *emotional contagion* [Jacob and de Vignemont, 2016]. It is only when mirroring is loaded with cognition that it acquires the capacity to provide a genuinely empathic response; but then it has become a component in a high-level cognitive phenomenon, resulting in the collapse of primary empathy to a 're-enactive' or 'reconstructive' form of empathy, based on high-level mindreading.

(2) Added to this difficulty is another, in our eyes an even more serious one: the simulationist model of reconstructive empathy proposed by Goldman is based on a neo-Cartesian theory of mental concepts and conceived of simulation as a form of projection. In short, it is antithetical to a 'mindreading first' account. This is the case both during the development of mindreading in childhood and in the context of adult mentalistic performance – first-person mentalisation both precedes and serves as a foundation for third-person mentalisation. However, we believe that this approach is not confirmed in the literature of developmental psychology, social psychology or cognitive psychopathology; rather, the data delivered from these fields of research are best explained by a theory of introspection that accepts the Rylean perspective of a *parity* between the first and third person [Schwitzgebel, 2024].

The most robust version of a self/other parity account of self-knowledge was elaborated by an author cited in the target article. In contrast to the project of developing a theory of introspection as an *inner sense*, [Carruthers 2011, 2015] has developed two theories: the Interpretive Sensory-Access theory of self-knowledge (ISA theory), a theory of the nature and sources of self-knowledge that postulates both sensory access and a process of self-interpretation; and the sensory-based theory of conscious thinking, a theory of the causes and contents of reflective thinking according to which everything we are conscious of is a set of sensory-based contents present in working memory.

According to the ISA theory, only a very limited repertoire of sensory or quasi-sensory states (i.e., states that include sensory components) can be accessed non-interpretatively. The self-attribution of occurrent thoughts always occurs through *self-interpretation*. We turn "our evolved mindreading capacities on ourselves" and engage in the unconscious

interpretation of behavioural data, contextual data, and perceptual events such as visual and auditory images.

To explain the possibility of conscious access to sensory or quasi-sensory states, the ISA theory adopts the *Global Workspace Theory* (e.g., [Mashour et al., 2020]). Within the Global Workspace framework, Carruthers can formulate the thesis that only sensory or quasi-sensory states are conscious mental states.

Firstly, occurrent thoughts cannot be part of *first-order* access consciousness. Being amodal, thoughts cannot be globally broadcast, and thus cannot be accessible to consciousness. The reason is that global broadcasting depends on *top-down attention*, which has exclusively sensory focus. Thus, only mental events with a sensory-based format can become first-order access-conscious.

Secondly, occurrent thoughts cannot even become *higher-order* access-conscious. When one of the modular subsystems accesses the global workspace, its outputs are broadcast to an array of executive, conceptual, and affective systems. These systems 'consume' sensory information to draw inferences, form memories, generate affective reactions, form judgments, plan and make decisions, and verbally report. Among these conceptual systems, there is a multi-componential mindreading system, whose task is to generate higher-order, metarepresentational beliefs about the mental states of others and of oneself.

Now, it is plausible to hypothesise that the mindreading system evolved as the ability to understand the minds of others in order to provide an adaptive advantage in pursuing the aims of two motivational systems: the first committed to self-assertiveness and competition (e.g., [Byrne, 2022]), the second aimed to pro-sociality and cooperation (e.g., [Hrdy, 2009]). Only at a later stage did ancestral mindreaders begin to apply this capacity to themselves, forming beliefs around their own mental states as well as the mental states of others. Since the mindreading system evolved to understand the minds of others, it *looks outwards*: it only has access to the information transmitted by perceptual systems and must draw its conclusions solely from sensory information. However, since the system does not allow direct access to our own thoughts, we must infer these from observing our own behaviour and the situation in which we behave, interpreting ourselves exactly as we interpret others. The only difference between knowing thoughts in the first and third person is that in one's own case, the mindreading system has more information on which to base the interpretation. In one's own case, in addition to using overt behaviour, one can also rely on sensory and quasi-sensory states such as visual and auditory mental images that are globally broadcast. Thus, the ISA theory restricts self/other parity to a particular subclass of mental states, namely propositional attitudinal events (vs. sensory mental events).

At the heart of Carruthers's case for the ISA theory is the data from social psychology regarding frequent and pervasive confabulation relative

to one's own current or very recent thoughts. If the self-knowledge of thoughts is not direct, but results from self-directed mindreading, then there should be distinctive patterns of error in our claims about our thoughts, mirroring the ways in which we can be misled about the thoughts of others – e.g., because the theory-driven interpretive process is fed by misleading sensory and behavioural data, or the theories that we use to interpret ourselves are inadequate.

(3) The disappearance of conscious thought still leaves room for a distinction between unconscious *intuitive* processes and conscious *reflective* processes. The latter are forms of mental activity that are directed, e.g., toward solving a problem, arriving at a judgment, or reaching a decision. Since working memory is the system – a sensory-based system – that underlies these reflective processes, our conscious reflections will be composed exclusively of sensory-like events. Working memory uses top-down attention to activate and sustain imagistic representations in conscious form; there is no place within it for purely abstract non-sensory states such as beliefs, goals, or decisions.

Note that the distinction between unconscious intuitive processes and conscious reflective processes *does not* map onto the distinction between S1 and S2. It is difficult to see the evolutionary plausibility of two cognitive systems implemented in distinct neural subsystems: Why on earth would evolution start anew with S2 rather than modifying, expanding, or integrating the architecture of the pre-existing S1? This sort of objection led Frankish [Frankish, 2009] to put forward the hypothesis that S2 is realised within S1, i.e., that there are not two separate systems, but two levels or layers of cognitive processes, one dependent on the other.

Let us therefore assume the reality of the distinction between I tuitive and reflective processes of reasoning. This is not, however, a vindication of the S1/S2 distinction, since the properties typically used to distinguish S1 from S2 *crosscut* one another, and hence the distinction does not map onto the distinction between intuitive and reflective reasoning. Carruthers [Carruthers, 2014] offers several cases of crosscutting. Let us consider reflective reasoning. It is easy to show that, in some contexts, reflection does not improve but rather impairs performance, that there are some tasks where reliance on intuitive reasoning is best, and that reflective reasoning can also employ heuristics. As for intuitive systems, some can be slow, some can be controlled, and some can approach the highest normative standards.

As a result of this, Carruthers [Carruthers, 2014, 2015] suggests a new two-system account, according to which there is an intuitive system and a reflective system. Not too dissimilarly, Evans and Stanovich [Evans and Stanovich, 2013] opt for a dual-process theory, according to which Type-1 processes are autonomous and Type-2 processes use working memory and involve cognitive decoupling.

To conclude, basing the defence of the 'mindreading first' thesis on the distinction between S1 and S2 – and specifically on the distinction

between two systems of mindreading – does not seem to us a good solution. Arguably, there is a single mindreading system, whose computations vary in cost depending on the cognitive processes it interacts with (e.g., working memory or pragmatic processes, see [Jacob, 2019]). Importantly, this is not the only system that 'looks outward': observing the formation of self-awareness from a developmental perspective (the *selfing process* discussed in [McAdams, 1996], inspired by William James), we realise that all dimensions of the self arise from a social perspective. For instance, Fonagy and al.'s [Fonagy et al., 2002] social biofeedback model is an influential model of the social construction of emotional introspection based on the pre-existing ability to detect others' emotions; and even earlier, attention to the other is primary in the formation of the bodily self.

Significantly, this primacy of exploring the social world over self-exploration, which manifests itself at every level of development, has an important emotional-motivational dimension that Carruthers himself has underestimated. Instead, appreciating how the turning of social-directed capacities upon ourselves takes place in socio-communicative interactions with caregivers and other social partners opens up the psychodynamic topic of defences [Marraffa and Meini, 2024]. This is a crucial subject that, introduced by Freud one century ago, should now be written into the agenda of all of us who are interested in investigating the pinnacle of the human spirit.

References

Byrne, 2022 – Byrne, R. W. "Machiavellian Intelligence," in: Vonk, J., Shackelford, T. K. (eds.) *Encyclopedia of Animal Cognition and Behavior*. Cham: Springer International Publishing, 2022, pp. 4033–4038.

Carruthers, 2011 – Carruthers, P. *The Opacity of Mind*. New York: Oxford University Press, 2011.

Carruthers, 2014 – Carruthers, P. "The Fragmentation of Reasoning," in: Quintanilla, P., Mantilla, C., Cépeda, P. (eds.) *Cognición Social y Lenguaje*. Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú, 2014, pp. 181–204.

Carruthers, 2015 – Carruthers, P. *The Centered Mind*. New York: Oxford University Press, 2015.

Evans, Stanovich, 2013 – Evans, J. S. B., Stanovich, K. E. "Dual-Process Theories of Higher Cognition: Advancing the Debate," *Perspectives on Psychological Science*, 2013, vol. 8, pp. 223–241.

Fonagy et al., 2002 – Fonagy, P., Gergely, G., Jurist, E., Target, M. *Affect Regulation, Mentalization, and the Development of the Self*. London: Other Press, 2002.

Frankish, 2009 – Frankish, K. "Systems and Levels: Dual-System Theories and the Personal – Subpersonal Distinction," in: Evans, J. S. B., Frankish, K. E. (eds.) *In Two Minds: Dual Processes and Beyond*. New York: Oxford University Press, 2009, pp. 89–107.

Goldman, 2006 – Goldman, A.I. *Simulating Minds*. New York: Oxford University Press, 2006.

Goldman, Sripada, 2005 – Goldman, A. I., Sripada, C. "Simulationist Models of Face-Based Emotion Recognition," *Cognition*, 2005, vol. 94, pp. 193–213.

Hrdy, 2009 – Hrdy, S. *Mothers and Others*. Cambridge, MA: Harvard University Press, 2009.

Jacob, 2019 – Jacob, P. "Challenging the Two-Systems Model of Mindreading," in: Avramides, A., Parrott, M. (eds.) *Knowing Other Minds*. New York: Oxford University Press, 2019, pp. 79–106.

Jacob, de Vignemont, 2016 – Jacob, P., de Vignemont, F. "Vicarious Experiences: Perception, Mirroring or Imagination?" in: Kiverstein, J. (ed.) *The Routledge Handbook of Philosophy of the Social Mind*. London: Routledge, 2016, pp. 498–514.

Marraffa, Meini, 2024 – Marraffa, M., Meini, C. *The Developmental Psychology of Personal Identity: A Philosophical Perspective*. London: Bloomsbury, 2024.

Mashour et al., 2020 – Mashour, G. A., Roelfsema, P., Changeux, J.-P., Dehaene, S. "Conscious Processing and the Global Neuronal Workspace Hypothesis," *Neuron*, 2020, vol. 105, no. 5, pp. 776–798.

McAdams, 1996 – McAdams, D. P. "Personality, Modernity, and the Storied Self: A Contemporary Framework for Studying Persons," *Psychological Inquiry*, 1996, vol. 7, no. 4, pp. 295–321.

Schwitzgebel, 2024 – Schwitzgebel, E. "Introspection," in: Zalta, E. N., Nodelman, U. (eds.) *The Stanford Encyclopedia of Philosophy*. Spring 2024 edition. [https://plato.stanford.edu/archives/spr2024/entries/introspection/, accessed on 23.08.2024].

Stueber, 2019 – Stueber, K. "Empathy," in: Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy*. Fall 2019 edition. [https://plato.stanford.edu/archives/fall2019/entries/empathy/, accessed on 23.08.2024].