

AGATA MACHCEWICZ-GRAD *, MACIEJ TARNOWSKI **

UNREAL BELIEFS: EDITORIAL INTRODUCTION

Recent years have brought a revival of the debate on the ontological status of beliefs. The debate was originally shaped in the 1980s by a controversy between representationism, most notably Fodor's psychofunctionalism (1987), on the one hand, and eliminativism, which treated belief as a relic of an outdated theory of mind to be replaced by a mature cognitive science free of the propositional attitudes vocabulary, on the other (Churchland, 1981; Stich 1983). Nowadays, hardly anyone is willing to defend this radical form of antirealism about belief, chiefly due to the apparent indispensability of the central concepts of folk psychology both in everyday social cognition and in the sciences of the mind. For this reason, scepticism toward robustly realist views on belief tends to take one of two main forms: either mental fictionalism (Demeter, 2022; Parent, 2013; Toon, 2023), which treats talk about belief as a useful fiction, or superficialism (Curry, 2021; Mölder, 2010; Schwitzgebel, 2026), which accepts belief's existence but denies its status as a deep, scientifically respectable component of our mental architecture.

The book *Unreal Beliefs: An Anti-Realist Approach in the Metaphysics of Mind* by Krzysztof Pośłajko (2024) offers a new and intriguing variety of superficialism about belief: belief irrealism. The central, though *prima facie* paradoxical, thesis of the book can be stated as follows: beliefs exist, but they are not real. Pośłajko arrives at the distinction between existence and reality by framing the debate on the metaphysics of belief in natural kind terms. He starts from the

* University of Warsaw, Faculty of Philosophy. E-mail: a.machcewicz-grad@uw.edu.pl. ORCID: 0000-0002-6216-4620.

** Jagiellonian University in Kraków, Department of Philosophy. E-mail: maciej.tarnowski@uj.edu.pl. ORCID: 0000-0003-3824-4134.

naturalist assumption that the only properties that are real *sensu stricto* are natural properties, and proceeds to test beliefs against various accounts of natural kinds to determine whether they should be considered to form a natural kind and, hence, qualify as real. According to Pośłajko, the results of the test are negative: beliefs fail the test of naturalness, no matter whether we apply the demanding Lewisian (1983) criterion that requires reducibility of all natural kinds to fundamental physical properties or the now popular, more liberal accounts, such as Boyd's homeostatic property cluster view (1991) or Khalidi's causal account (2013), which emphasise projectibility as crucial for naturalness and admit functional properties as defining properties of natural kinds. Pośłajko does not, however, draw straightforward eliminativist conclusions from these considerations; instead, he claims that beliefs are unreal, i.e., there is no such a psychological kind in the natural world, but they *exist* in a deflationary sense. Beliefs exist in the sense that they can be truly attributed in certain circumstances, just as the properties of being a chair or having a gender can be truly attributed without committing us to the reality of these properties understood as natural kinds. This is the sense in which beliefs exist but are not real according to Pośłajko.

Crucially, belief irrealism involves a call for the revision of the commonsensical concept of belief. A deflationist account denies the central properties attributed to belief by folk psychologists who assume, according to Pośłajko, that beliefs are internally stored, discrete mental representations causally responsible for behaviour in virtue of their content. The category of belief, the irrealist maintains, does not describe any deeper facts having to do with our mental architecture; rather, it refers to superficial facts about our dispositions to think, behave, and experience in particular ways. If our commonsensical notion is to reflect a plausible metaphysics of belief, it should be revised along superficialist lines, which would free it from its problematic commitments. In this respect, Pośłajko's view departs from Schwitzgebel's neo-dispositionalism: the latter is intended to be a theoretical explication of the actual folk psychological concept, which is already implicitly superficialist, while the former holds that folk psychology is guided by robust realism about belief and, hence, prescribes a substantial revision of the commonsense concept of belief (Pośłajko, 2024, pp. 117–123).

Importantly, the proposed revision of our concept of belief does not preclude the possibility of making true belief ascriptions. In this regard, belief irrealism avoids the pitfalls of mental fictionalism, which faces the self-refutation worry in a similar way to eliminativism: after all, to distinguish between utterances about the actual world and the utterances about fiction seems to require adopting different propositional attitudes toward these utterances—and such attitudes are precisely what a fictionalist denies are real (Pośłajko, 2024, pp. 25–28). Belief irrealism, therefore, promises to overcome the limitations of alternative but closely related views on the nature of belief, like neo-dispositionalism and mental fictionalism. Whether this promise is fulfilled is one of the questions extensively discussed in the contributions presented below.

This issue of *Semiotic Studies* is entirely devoted to the discussion of the idea of belief irrealism. Our collection brings together ten critical papers which engage with different aspects of Pośłajko's proposal. They are formulated from diverse philosophical positions, including views ranging from natural kind realism about belief to global intentional antirealism. The collection of critical commentaries is followed by an extensive, collective reply to the critics by Krzysztof Pośłajko.

The issue opens with Tad Zawidzki's paper. Zawidzki challenges the thesis underlying Pośłajko's belief irrealism to the effect that belief is located close to the end of the spectrum of naturalness where non-natural, and hence non-real, properties are located. He sketches two possible ways of developing the view of belief, according to which it not only exists but also forms a natural kind: a Dennettian account of belief as a property attributable to intentional systems and a Brandomian idea of belief as discursive commitment. Zawidzki builds his argument upon the assumption that Boyd's account of natural kinds as homeostatic property clusters sustained by a hidden mechanism is more suited to the discussion of belief than the reductionist Lewisian account. He contends that, pace Pośłajko, both accounts of belief he discusses can be considered to provide the relevant mechanism required by Boyd's account: the mechanism of natural selection in the case of the intentional systems view and the mechanism of mind-shaping secured by deontic attitudes adopted by the members of the community of interpreters. Belief can, therefore, be considered a natural kind in this more liberal sense. Zawidzki is aware that both proposals he points out involve a significant amount of revision with respect to the folk concept of belief. He argues, however, that *any* consistent theory of belief will involve such a revision, given the unsystematic nature of the folk concept.

The second paper by Witold Hensel provides an interesting challenge to the very idea, shared by Pośłajko and many other participants in the debate, of construing the issue of the reality of beliefs in natural-kind terms. Hensel formulates two problems for such construals: the problem of disappearing substance and the problem of disappearing reality. The first problem arises when the discussion is guided by the Putnamian-Kripkean essentialist, causal-historical account of natural kind terms. The account renders the discussion between realists and anti-realists about belief intractable: the realist can always withdraw a commitment to any putative law governing belief without giving up the thesis that the term "belief" refers to a natural kind. The second problem arises when the debate is underpinned by moderate accounts of natural kinds like Boyd's or Khalidi's, which have proved to be more accurate in describing how natural kinds are distinguished within actual science. Such accounts give up the demand for intrinsic essence and reducibility of natural kinds in lieu of the criterion of projectibility. As such, they improve the prospects for construing belief as a natural kind, yet it is doubtful whether they can establish its *reality*: after all, liberal accounts of natural kinds allow conventional, mind-dependent posits of science to be considered natural kinds. Natural kind realism about belief proves to be futile either

way: it is hardly disprovable if one adopts a causal-historical view of natural kind terms, or it becomes so weak to the point that it undermines the whole issue of the reality of beliefs if a moderate view of natural kinds is adopted.

The next, very interesting contribution is by Amir Horowitz, who criticizes some of the arguments proposed by Pośłajko to the effect that beliefs are not real and promotes, instead, his own view, which construes the posits of folk psychology to be real, though on a very revisionary understanding of them. He argues, pace Pośłajko, that beliefs understood as psychofunctional states ground objective similarities, are causally relevant with respect to behaviour, and their reality is corroborated by the predictive-explanatory success of folk psychology. A crucial caveat of Horowitz's argument, however, is that these psychofunctional states are not *intentional*—they are merely logico-syntactic structures devoid of representational properties. Indeed, Horowitz argues for strong intentional anti-realism, which says that “no concrete entity possesses or can possess any intentional property” (p. 62 of the current issue). In folk psychology, therefore, we discover some aspects of the internal architecture of the mind by tracking the causally efficacious syntactic structures, even though we falsely assume that these structures possess intentional properties. This conclusion, although revisionary with respect to folk psychology, inclines Horowitz to consider his position a realist rival of Pośłajko's mild irrealism.

Piotr Kozak's paper is exceptional in our collection in that it defends the view that beliefs, understood as inner, causally efficacious representations, not only exist but are real in the sense introduced by Pośłajko. The line of argument runs through the discussion of the relation between rules of rationality and beliefs. Kozak formulates a rational indispensability argument which is based on the premises that our ontology should be determined by our best theories, and that beliefs are indispensable posits of any plausible theory explaining why norms of rationality are applicable to our behaviour. From this, he draws the conclusion that we should be realists about beliefs. He contends, furthermore, that beliefs can guarantee the application of norms of rationality to behaviour only if they are causally efficacious and strictly individuable, and concludes that beliefs are not only existent but also real in the sense that they exhibit the natural properties Pośłajko denies them.

The next two papers in our collection approach Pośłajko's anti-realist project from a different angle—they take issue with his revisionist claim of the folk, commonsense psychology needing a thorough conceptual reengineering. This claim is substantiated by the common philosophical presupposition that the folk—or, more precisely, “the common sense of the contemporary, educated, western person” (Pośłajko, 2024, p. 118)—conceptualize belief in the same way as many philosophers and cognitive scientists do, as a causally efficacious, basic representational state. Both papers in turn dispute this assumption, interpreting empirical work in cognitive and social psychology and anthropology that suggests a less committal and more pluralist picture.

Annemarie Kalis argues that it is neither the case that Western folk are committed causalists, nor that a Western outlook on the mind should be presumed universal or privileged. Even though it is certainly true that much of the cognitive psychology and research on mindreading assumes a causalist framework and that the folk use a causal vocabulary, this need not mean that the folk have causalist commitments themselves. Kalis approves of Ryle's dictum that ordinary discourse should not be treated at face value as bringing in philosophical commitments, but rather admits a fair paraphrase. When one says "Mary quit smoking *because* she believes it is bad for her health", Kalis writes, they should not be understood as making a metaphysically heavy claim, for they are not philosophizing at all; it is a mistake to impose the causalist view on the folk based on such fragile data. But, even if such data were to be taken at face value, why do we fixate on a quite narrow set of them that comes from the Western discourse? In the latter part of the text, Kalis analyzes the linguistic and anthropological data that suggest that "folk psychological practices actually come in all shapes and sizes" (p. 94), when we decide to look a bit further: to how non-Western language doxastic verbs are used, or to different sociocultural customs of explaining action. While this heterodox nature of folk psychology does not, in Kalis' view, mean that we are necessarily wrong in interpreting the behaviour of agents in various cultures using a causalist "belief" terminology, it means that pointing criticisms of this picture against the folk, rather than philosophers and cognitive scientists, is misplaced. This article acts therefore both as a fair warning against over-interpreting the folk discourse and a too narrow focus on Western (or, more generally, WEIRD in the sense of Henrich, Heine, Norenzayan, 2010) folk psychological practices.

In a second paper concerned with the interpretation of the folk discourse, Joanna Komorowska-Mach also argues against its causalist reading—but she does so by offering an alternative account of the data cited in support of it. In her view, commonsense psychology does not posit discrete, causally efficacious belief states—it is the language that pressures us into doing so. Building upon empirical findings from developmental psychology, Komorowska-Mach argues that our basic folk understanding of the mind, embedded in mindreading practices, is actually a more holistic, context-dependent process that need not distinguish particular representational and motivational states the way our ordinary language does. It is our need to communicate and conceptualize the findings through linguistic means, that enforces the rigid, discrete categories theorized about by philosophers. It is, in her view, crucial to distinguish between folk psychology as a *craft* (after Dennett, 1998), which we use in our mindreading practice, and as a *theory*, which involves theoretical posits such as belief; and we should not think of the folk as committed to such theoretical entities, given they do not play an important role in our actual cognitive life. Therefore, Komorowska-Mach's paper forces us to rethink the role that the concept of belief plays in philosophical discussions of the success and failures of folk psychology and whether it is actually as central as many claim it to be.

While the preceding five papers take issue with particular diagnoses of the anti-realist and revisionary picture advanced by Pośłajko, the latter four may be thought of as essentially agreeing with the diagnosis, but suggesting alternative treatment. The skepticism toward substantial realist commitments of “belief”-talk is shared by two important traditions in philosophy of mind: interpretivist and dispositionalist approaches, which suggest that beliefs should not be seen as concrete contentful mental states, but rather as theoretical devices used to interpret actions of others, or as dispositions to act, reason, or feel. We are happy to see the contributions from both schools of thought represented here, which forcefully argue for alternative accounts that nevertheless agree with Pośłajko’s minimal non-realism in opposing the dominant representationalist picture.

On the interpretivist side, we start with David Sanchez Curry’s *Beliefs Are Real(ly for Social Animals)*. In his paper, Curry revisits, from an interpretivist perspective, Pośłajko’s thesis that beliefs do not constitute a natural kind. While he remains sympathetic to Pośłajko’s claim that beliefs should not be seen as a natural kind understood along the Lewisian, physicalist lines, he doubts that we cannot assume that beliefs constitute a natural kind in a wider, socially determined sense. After all, we seem to use many terms in everyday discourse and science alike that have a natural-kind semantics, while being, at the same time, dependent on the environmental and social conditions we inhabit. Interpretivists can, in Curry’s view, provide an account of belief as a natural kind essentially constituted by interpretation practices, and thus claim that beliefs are *real* in Pośłajko’s sense without embracing the psychofunctionalist picture he criticizes. Curry also notices crucial disanalogies between socially constituted kinds, such as the concept of race and the concept of belief, an analogy which Pośłajko uses to dismiss the social-constructivist realism about belief that Curry favors. This paper thereby provides an original, interesting investigation into metaphysical presuppositions and commitments of the interpretivist view, much needed in the literature and challenging for non-realists with similar intuitions.

Bruno Mölder, the second interpretivist among our authors, signals his agreement with Pośłajko’s outlook on the prospects of belief naturalization (unlike Curry), but takes a more detailed look at how interpretivism may develop in line with it. In that, he responds both to Pośłajko’s anti-interpretivist argument from the book and its more elaborate version from an earlier (2020) article; in short, the argument stresses the judgment-dependence of beliefs in the interpretivist framework, and challenges it to provide an account of interpretation that does not rely on actual practices and judgments of interpreters (which make it problematic to ascribe beliefs *falsely*) without falling into deflationism (making “interpretation” ontologically redundant). Replying to this objection, Mölder clarifies his view—*ascription theory*—and shows how it escapes Pośłajko’s challenges by thoroughly relying on the canonical ascription as constitutive. In line with this position, he also counters Pośłajko’s claim that folk psychology entails substantial metaphysical commitments and disputes the usefulness of a revised concept of belief that Pośłajko advocates for. For readers interested in

the metaphysical commitments of interpretivism and ordinary “belief”-discourse, this paper will prove interesting, serving as a counterpoint to both Pośłajko’s criticism and Curry’s position defended in the former paper.

The next paper in our collection, by Brice Bantegnie, introduces the final view—dispositionalism—and provides its defense against Pośłajko’s charges. Bantegnie argues that, contrary to Pośłajko’s classification, dispositionalism need not be anti-realist—identifying beliefs with multitrack dispositions makes them in no way less real, as they may be effectively studied by relevant sciences and constitute a natural kind. More importantly, however, Bantegnie defends dispositionalism against the conceptual inadequacy challenge, arguing that our concept of belief need not be a causalist one, as the explanations in which the concept features are not necessarily causal explanations. By doing this, Bantegnie addresses an important question of causalist commitment of folk psychology that recurs throughout many of the papers featured here, and interestingly argues for an alternative role of scientific explanations and naturalistic metaphysics of mind. Therefore, the paper is of interest not only to those interested in conceptual analysis of belief, but also to those focused on the methodology and ontology of cognitive science.

Our collection of commentaries closes with a paper by Eric Schwitzgebel involving a defence of a broad family of superficialist views on belief and considerations about the prospects for applying superficialism to the debate about whether LLMs and robots can hold beliefs. Superficialist views on belief construe it as constituted by such properties as phenomenal, behavioural, and/or inferential dispositions and reject its identification, along representationalist lines, with any “deep” facts about the cognitive architecture of the subjects of belief. Schwitzgebel argues that the views of belief of this sort are both pragmatically justified (we care about the dispositional manifestations of belief and not about its neurocomputational realization) and scientifically projectible (the superficialist notion of belief can be extended to subjects that are very different from humans in terms of their internal architecture). Schwitzgebel observes that superficialism about belief suggests that, at the current stage of development of AI systems, they seem to lack the phenomenal experience characteristic of belief and as such they should rather be attributed with *belief*^{*}, involving the behavioural and inferential dispositions characteristic of belief but not its phenomenal aspect. Schwitzgebel also comments on Pośłajko’s critique of dispositionalism, which hinges on the objection that the folk notion of belief is more robust than Schwitzgebel’s preferred view suggests. Schwitzgebel’s response recalls the point made by Zawidzki in the opening paper: the folk concept of belief has several dimensions, some of which are in line with dispositionalism and some with representationalism. Our pragmatic purposes in the use of the notion of belief should decide between these tendencies on each occasion, or so Schwitzgebel says.

In an often-cited passage from *The Analysis of Mind*, Bertrand Russell says: “Psychology, theory of knowledge and metaphysics revolve about belief, and on

the view we take of belief our philosophical outlook largely depends” (1921, p. 231). Perhaps this quote illustrates why the debate on the nature of belief is so complex, with many moving parts affected by adopting a particular stance. But more emphatically, it shows why the debate is not just another philosophical puzzle with low theoretical and practical stakes. What we take belief to be is important—and we are very proud of this collection, which allows both deep and wide-ranging reflection on this question. We are very grateful to all our authors and Krzysztof Pośłajko for their outstanding contributions; we would also like to thank the referees, who provided invaluable insight and care for each of the submitted papers. Last but not least, we also thank the team of *Semiotic Studies* for making this issue happen: Dominik Dziedzic for his impeccable work and management, and Tadeusz Ciecierski for his encouragement and professional advice.

REFERENCES

- Boyd, R. (1991). Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds. *Philosophical Studies*, 61(1/2), 127–148.
- Churchland, P. M. (1981). Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy*, 78(2), 67–90.
- Curry, D. S. (2021). How Beliefs Are Like Colors. *Synthese*, 199(3–4), 7889–7918.
- Demeter, T. (2022). A Mental Fictionalism Worthy of Its Name. In T. Demeter, T. Parent, A. Toon (Eds.), *Mental Fictionalism: Philosophical Explorations* (pp. 186–103). New York; London: Routledge.
- Dennett, D. C. (1998). Two Contrasts: Folk Craft versus Folk Science and Belief versus Opinion. In D. Dennett (Ed.), *Brainchildren* (pp. 81–95). Cambridge, MA: MIT Press.
- Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.
- Henrich, J., Heine, S. J., Norenzayan, A. (2010). The Weirdest People in the World? *Behavioral and Brain Sciences*, 33(2–3), 61–83.
- Khalidi, M. A. (2013). *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*. New York: Cambridge University Press.
- Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61(4), 343–77.
- Mölder, B. (2010). *Mind Ascribed. An Elaboration and Defence of Interpretivism*. John Benjamins.
- Parent, T. (2013). In the Mental Fiction, Mental Fictionalism is Fictitious. *The Monist*, 96(4), 605–621.
- Pośłajko, K. (2020). Can Deflationism Save Interpretivism? *Philosophia*, 48(2), 709–725.
- Pośłajko, K. (2024). *Unreal Beliefs: An Anti-Realist Approach in the Metaphysics of Mind*. London: Bloomsbury Academic.
- Russell, B. (1921). *The Analysis of Mind*. London: George Allen & Unwin Ltd.

- Schwitzgebel, S. (2026). “Dispositionalism, Yay! Representationalism, Boo!”. In J. Jong, E. Schwitzgebel (Eds.), *The Nature of Belief*. Oxford: OUP.
- Stich, S. P. (1983). *From Folk Psychology to Cognitive Science: The Case Against Belief*. MIT Press.
- Toon, A. (2023). *Mind as Metaphor: A Defence of Mental Fictionalism*. Oxford, GB: Oxford University Press.