

POLSKIE TOWARZYSTWO SEMIOTYCZNE

# STUDIA SEMIOTYCZNE

Tom XXXIX • nr 1

PÓŁROCZNIK

UNREAL BELIEFS:  
SPECIAL ISSUE ON THE BOOK  
BY KRZYSZTOF POSŁAJKO



WARSZAWA • 2025

Redaktorzy naczelni 1970–2023 (*Editors-in-Chief* 1970–2023):

Założyciel „Studiów Semiotycznych” (*Founding Editor*):

Jerzy Pelc (1970–2015)

Andrzej Biłat (2016–2023)

Zespół redakcyjny (*Editorial Board*):

Tadeusz Ciecierski (*Editor-in-Chief*)

Dominik Dziedzic (*Assistant Editor*)

Redaktorzy numeru (*Issue Editor*):

Agata Machcewicz-Grad, Maciej Tarnowski

Rada naukowa (*Advisory Board*):

Jerzy Bartmiński (Uniwersytet Marii Curie-Skłodowskiej), Paul Bouissac (University of Toronto), Andrzej Bronk (Katolicki Uniwersytet Lubelski Jana Pawła II), Roman Kossak (City University of New York), Stanisław Krajewski (Uniwersytet Warszawski), Idalia Kurcz (SWPS Uniwersytet Humanistycznospołeczny), Witold Marciszewski (Uniwersytet w Białymstoku, Fundacja na Rzecz Informatyki, Logiki i Matematyki), Genoveva Martí (ICREA & Universitat de Barcelona), Adam Nowaczyk (Uniwersytet Łódzki), Stefano Predelli (University of Nottingham), Mieczysław Omyła (Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie), Piotr Stalmaszczyk (Uniwersytet Łódzki), Anna Wierzbicka (Australian National University), André Włodarczyk (Université Paris-Sorbonne), Jan Woleński (Uniwersytet Jagielloński, Wyższa Szkoła Informatyki i Zarządzania)

Skład elektroniczny:

Dominik Dziedzic

Adres redakcji: Krakowskie Przedmieście 3, 00-047 Warszawa

e-mail: [studiasemiotyczne@pts.edu.pl](mailto:studiasemiotyczne@pts.edu.pl)

<http://studiasemiotyczne.pts.edu.pl/>

ISSN 0137-6608; e-ISSN 2544-073X

 Creative Commons Uznanie Autorstwa 4.0 (CC BY)

## CONTENT

|                                                                                                                |     |
|----------------------------------------------------------------------------------------------------------------|-----|
| Agata Machcewicz-Grad, Maciej Tarnowski, Unreal Beliefs: Editorial Introduction .....                          | 5   |
| Tadeusz Zawidzki, Locating Belief on the Spectrum of Existence .....                                           | 15  |
| Witold M. Hensel, Troubles With the Natural-Kind Approach to the Debate Over the Reality of Belief .....       | 29  |
| Amir Horowitz, Pośłajko on Belief Similarities and Belief Causation .....                                      | 49  |
| Piotr Kozak, Why Should We Be Realists About Beliefs? The Rational Indispensability Argument for Realism ..... | 67  |
| Annemarie Kalis, The Folk Are Doing Fine (But the Philosophers Are in Trouble) .....                           | 89  |
| Joanna Komorowska-Mach, Rethinking Folk Realism About Beliefs .....                                            | 101 |
| Devin Sanchez Curry, Beliefs Are Real(ly for Social Animals) .....                                             | 113 |
| Bruno Mölder, Belief: Minimal Non-Realism and the Constitutive Role of Interpretation .....                    | 131 |
| Brice Bantegnie, Dispositionalism and Realism .....                                                            | 147 |
| Eric Schwitzgebel, Superficialism About Belief, and How We Will Decide That Robots “Believe” .....             | 159 |
| Krzysztof Pośłajko, Defending Minimal Non-Realism .....                                                        | 173 |



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*<sup>\*</sup>, 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.



TADEUSZ ZAWIDZKI \*

## LOCATING BELIEF ON THE SPECTRUM OF EXISTENCE

**SUMMARY:** Krzysztof Pośłajko’s insightful and incisive book, *Unreal Beliefs*, successfully reignites the dormant debate about belief eliminativism. His key insight, inspired by Lewis’s (1983) distinction between natural and non-natural properties, is that a category can exist without being real: “completely arbitrary collections of modal objects” (Pośłajko, 2024, p. 68) exist, but since they are not natural kinds, naturalists need not accept that they are real. This opens up the possibility that beliefs exist in a minimal sense while not being real, since they may fail to be natural kinds, thus forestalling the counter-intuitive and paradoxical implications of belief eliminativism, in favor of Pośłajko’s favored “minimal non-realism”. However, as Pośłajko realizes, the distinction between natural and non-natural properties is “graded” (2024, p. 68), e.g., categories of the special sciences, like money, are not *as natural* as the categories of our best physics, but they are not completely arbitrary collections of modal objects either. This paper concerns where on this ontological spectrum to locate the category of belief, a question about which Pośłajko is not fully clear. Although he argues successfully that there is little evidence that belief constitutes a natural, *neurocomputational* kind, I argue that there are at least two other ways belief might qualify for natural kindhood: as a property of Dennettian intentional systems, inevitably produced by natural selection, or as a socially maintained deontic kind, i.e., Brandomian discursive commitment, inevitable in populations using natural language to coordinate. Pośłajko’s arguments for minimal non-realism about belief fail to rule out these versions of belief realism.

**KEYWORDS:** belief, eliminativism, minimal non-realism, special sciences, natural kinds, Boyd, homeostatic property cluster, projectibility, cognitive science, computational mechanism, Dennett, intentional systems, natural selection, Brandom, discursive commitment, deontic attitude.

---

\* George Washington University, Department of Philosophy. E-mail: zawidzki@gwu.edu. ORCID: 0000-0001-7307-659X.

## 0. Introduction

Krzysztof Posłajko's *Unreal Beliefs* is a very clear and innovative re-engagement with a crucial question in the metaphysics of mind, discussion of which had gone dormant since the 1990s: is the central category of folk psychology, belief, *real*? Despite vigorous debate in the 1980s and early 1990s, philosophers of mind largely lost interest in it afterwards due to the apparent incoherence of believing that beliefs do not exist, and more general incredulity regarding belief skepticism. However, Posłajko ably shows that this dismissal of the question rests on a conflation of two distinct issues: whether or not beliefs *exist* and whether or not beliefs are *real*. Drawing on Lewis's (1983) distinction between natural and non-natural properties, Posłajko argues that once the question of whether or not beliefs are natural is distinguished from the question of whether or not beliefs exist, belief realism is no longer as obvious as recent philosophy of mind has assumed. On Lewis's view, existence is cheap: "completely arbitrary collections of modal objects" (Posłajko, 2024, p. 68) are *existing* properties, but such properties should be distinguished from the "fully natural" (p. 68) properties postulated by our best science. On Posłajko's view, only the latter should count as *real* properties. Thus, the question of whether or not beliefs are real can be distinguished from whether or not beliefs exist: the former concerns the naturalness of belief, i.e., whether beliefs belong in the ontology of our best science of the mind. One can deflect incredulity about belief skepticism while taking this question seriously because beliefs exist whether or not they are natural, and hence, according to our best science, real. This way of framing the debate also has the virtue of refocusing it on what had always been its central concern: whether or not our everyday, "folk" understanding of the mind is consistent with what our best science shows.

As both Lewis and Posłajko recognize, the distinction between natural and non-natural properties is "graded" (p. 68). It is not the case that all existing properties are either fully natural denizens of the ontology of our best science, i.e., for physicalists like Posłajko, a completed physics, or completely arbitrary collections of modal objects. In particular, the categories of the special sciences, e.g., money, though undefinable in terms of fundamental physical properties, seem more natural than completely arbitrary collections of modal objects. According to Posłajko, this is mainly because they track objective similarities among individual objects, figure in surprising predictions, and can at least be *explained* in terms of more fundamental, physical properties (p. 64). Despite Posłajko's laudable metaphysical nuance concerning this general point, he is surprisingly inexplicit about *precisely where* on this ontological "spectrum" (p. 68) his "minimal non-realism" about belief locates the category. Clearly, the category of belief is more than a *completely arbitrary* collection of modal objects. If it were, its deep entrenchment in folk psychology would be an utter mystery. On the other hand, given the challenges relating belief to natural categories that Posłajko ably recounts (pp. 98–112), it is not clear how far towards the "natural" end of the spec-

trum belief belongs. Pośłajko argues not very far, on the grounds that there is little evidence that believers are objectively similar and that classifying them as believers supports surprising predictions, and belief is difficult to explain in terms of physical properties due to the challenges of naturalizing belief content (pp. 106–112) and classic puzzles about mental causation (pp. 98–106).

These arguments are not sufficient to settle the issue of belief's place on the spectrum of existence, and, in particular, its location relative to categories of the special sciences. For example, money seems at least as naturalistically problematic as belief because, as Pośłajko recognizes (p. 60), it can be explained in terms of more natural properties only via appeals to the preferences (and, presumably beliefs) of economic agents. The “objective” similarity of objects denoting the same economic value seems likewise dependent on psychological attitudes of relevant agents. It is true that money would not be a central category of economics if it did not support surprising predictions, but it is also hard to explain why the category of belief is so central to folk psychology unless it likewise supports surprising predictions. Indeed, our success in many domains seems otherwise inexplicable: think of detective work, for example.

Pośłajko also appeals to a more precise criterion of naturalness in support of minimal non-realism about belief: Boyd's analysis of natural kinds as “homeostatic property clusters” (Boyd, 1991; Pośłajko, 2024, p. 55). He argues that, unlike categories posited by special sciences, belief does not qualify as a natural kind according to this criterion (Pośłajko, 2024, pp. 91–97). His argument depends on the plausible conjecture that the neurally implemented causal mechanisms posited by our best cognitive science are unlikely to traffic in anything like beliefs, as the folk understand them. However, as I argue below, this is not necessary for belief to qualify as a natural kind in Boyd's sense. Instead, we might follow Dennett (1991) in grounding belief in natural selection, which, because it tends to respect principles of good design, reliably generates “intentional systems”, i.e., systems systematically interpretable as having beliefs. Or, inspired by Brandom's (1994) concept of discursive commitment, we might understand belief in terms of a certain kind of social role inevitably instituted by humans using language to coordinate.

In what follows, I first explain how Pośłajko deploys the view that natural kinds are homeostatic property clusters and the unlikelihood that cognitive science will vindicate the category of belief in support of his minimal non-realism. Next, I provide more detail about, (1) how belief might instead be grounded in a Dennettian appeal to natural selection, or (2) a Brandomian appeal to socially instituted discursive commitments. I conclude by addressing Pośłajko's concerns regarding revisionism about the folk concept of belief: although it is true that the Dennettian and Brandomian approaches are significantly revisionary, I argue that *any* systematic and coherent account of belief, including the reduction to neurally implemented cognitive mechanisms Pośłajko considers, must inevitably be comparably revisionary.

# 1. Natural Kinds as Homeostatic Property Clusters, Cognitive Science and Belief

Posłajko's arguments for minimal non-realism about belief hinge on the premise that it does not constitute a natural kind (2024, pp. 92–97). As he notes,

the traditional model of natural kinds, namely the one proposed by Putnam [...] and Kripke [...], is not well suited to the area of psychology [... because if] psychological kinds are defined functionally, then we cannot postulate a hidden structure common to all instances of such a kind (Posłajko, 2024, p. 92),

as the Putnam-Kripke model would require. However, he argues that, even on more liberal models of natural kinds, belief fails to qualify. In particular, even on Boyd's "homeostatic property cluster" approach (Boyd, 1991), or Khalidi's (2013) even more liberal approach, there is little evidence that belief is a natural kind. On Boyd's model, natural kinds support "successful and robust inductive generalizations" (Posłajko, 2024, p. 92), and this is explicable in terms of a mechanism that homeostatically maintains clusters of properties associated with the kind, in the way that, e.g., natural selection maintains biological species in a state of equilibrium against environmental perturbations. Khalidi dispenses with the requirement that there be a mechanism maintaining a homeostatic property cluster associated with a natural kind; all that matters is projectibility: "If there are some robust surprising generalizations about a certain category, then we might suspect that there is some joint in nature that we managed to discover using this category" (p. 93). But, according to Posłajko, there is no evidence that belief constitutes a natural kind even in this sense.

This claim seems implausible on its face. The reason is that, as noted above, it is very hard to explain belief's entrenchment as the central category of folk psychology if it supports no robust, surprising generalizations. Why would the folk constantly interpret each other in terms of a category that provides no predictive purchase on behavior? It is possible that the use of belief for quotidian behavioral prediction has been overstated. There is evidence that "reason explanations" are used primarily for justificatory rather than predictive purposes (Malle, 2006; Malle, Knobe, Nelson, 2007). Perhaps some sense can be made of the idea that belief would remain a stable category of folk psychology even if it were used primarily for ad hoc, retrospective rationalizations, in the way that folk appeals to supernatural categories often are. But this still seems insufficient to explain belief's multifarious quotidian uses. In particular, long-term strategic planning is often dependent on tracking others' beliefs in ways that support surprising, reliable predictions. As I remark above, detective work is a good example, but any time strategic interaction relies on taking into account differing perspectives on relevant facts, it seems that success depends on belief's projectibility.

On what grounds, then, does Posłajko claim that there is little evidence of belief's projectibility? He appeals to work in cognitive science, according to which:

There are [...] important reasons to be sceptical about the claim that beliefs are projectible as there seems to be insufficient evidence that this category is robust and unitary [...] there seems to be no convergence in the criteria for detecting beliefs, and the category of belief (as defined in cognitive psychology) seems to play many distinct theoretical roles. This last fact might reasonably lead to the conclusion that we are dealing with several distinct mechanisms and processes that we label “beliefs”, and the mechanisms and processes in question belong in fact to different cognitive kinds. The label “belief” can then reasonably be seen as used to describe a variety of phenomena which would be best considered as constituting more than one natural kind. (Posłajko, 2024, p. 96)

Even if this is right, it seems irrelevant to belief’s status as a natural kind in *Khalidi’s* sense: the fact that cognitive science fails to coherently apply the category of belief in ways that support surprising, robust predictions, does not show that the folk fail at this, and the latter seems sufficient for belief qualifying as a natural kind in *Khalidi’s* sense.

Posłajko also provides arguments that belief is unlikely to be causally relevant to behavior, or explicable in terms of natural properties (pp. 98–114). Both of these argumentative strategies turn on the status of the propositional contents in terms of which the folk individuate beliefs. Beliefs as individuated by content are unlikely to be causally relevant to behavior because, in any putative case of belief causation, the actual, occurrent, psychofunctional cause appears to belong to a narrower category, neutral between candidate beliefs differing in content (pp. 103–106). And every attempt to explain belief content in terms of more fundamental natural categories, like information or biological function, has proven deeply problematic (pp. 106–112). As I note in the introduction, such problems relating belief to more fundamental, causally potent natural kinds are not sufficient to disqualify it as a natural kind. The reason is that there are other natural kinds playing central roles in other special sciences which are comparably problematic, e.g., money. Money is clearly a natural kind because it supports surprising and robust economic predictions. But its status relative to more fundamental natural kinds is as problematic as that of belief, since what counts as a certain quantity of money depends entirely on the beliefs and preferences of relevant economic agents.

It seems then that Posłajko’s argument for minimal non-realism about belief, i.e., the claim that belief does not constitute a natural kind according to our best science, depends entirely on the claim that there is no evidence that it is a useful category for cognitive science. As he puts it:

[...] according to realists, attributions of beliefs allow us to attribute a property which makes [...] subjects genuinely similar. Also, the category “belief”, understood in a general sense, denotes a natural kind which is mind-independent and capable of functioning in empirical generalizations. These objective similarities between beliefs with particular content and between states belonging to the general category of “belief” *stem from the fact that belief attributions depict a deep internal cognitive architecture*. Anti-realists would deny all these claims. (Posłajko, 2024, p. 55, emphasis added)

And, in his view, there is little evidence that belief attributions depict a deep internal cognitive architecture. However, given the seemingly obvious fact that belief supports surprising, robust predictions in quotidian applications, this seems insufficient to disqualify belief as a natural kind, especially if we accept difficult to naturalize categories like money solely on the basis of their projectibility.

There appear to be only two responses available to Posłajko. According to the first, the projectibility of belief in quotidian applications might be a kind of illusion:

[...] the fact that people do very well at predicting and coordinating does not, by itself, prove that folk psychology is a successful theory. To justify the latter claim, one must show that we are successful at prediction and coordination because we use the theory of folk psychology, which makes an essential reference to states such as beliefs. And it is exactly this explanatory claim that can be problematized. (Posłajko, 2024, p. 148)

Although I agree with Posłajko that our reliance on explicit, theoretical applications of the concept of belief in successful, quotidian social cognition is typically overstated, the fact that we often get by without it is insufficient to show that it supports *no* surprising, robust predictions. Although the folk often coordinate successfully without theorizing, there are also many quotidian contexts in which explicit use of the folk concept of belief yields surprising, robust predictions. Any long-term, strategic planning which hinges on tracking different perspectives on relevant facts, e.g., fooling a suspected criminal into a self-incriminating decision, would seem to rely on belief’s projectibility.

The second response available to Posłajko is to adopt Boyd’s more stringent criteria for natural kindhood: not only must natural kinds support surprising, robust predictions; there must also be mechanisms which explain their stability as homeostatic property clusters. Posłajko could then appeal to his argument that belief plays no role in the internal, computational mechanisms posited by our best cognitive science to deny that there are such mechanisms in the case of belief. But this assumes that only internal, computational mechanisms postulated by cognitive science could play such a role. Indeed, Posłajko explicitly identifies the representationalist theory of belief as the paradigm of belief realism (pp. 85–89), suggesting that he views the question of whether belief constitutes a real, natural kind as closely related to its role in computational mechanisms postulated by our best cognitive science. However, as I argue in the next section, there are at

least two alternative kinds of mechanisms that could explain how belief is maintained as a homeostatic property cluster: Dennett's suggestion that natural selection inevitably produces intentional systems, systematically interpretable as believers, and the idea, inspired by Brandom's notion of "discursive commitment" (1994), that populations using language to coordinate must institute deontic statuses committing and entitling their members to sentences and behaviors implied by their utterances.

## 2. Alternative Mechanisms to Explain the Homeostatic Stability of Belief

### 2A. Intentional Systems and Natural Selection

As Posłajko notes (Posłajko, 2024, p. 81), Dennett defends a form of realism about belief, while agreeing with Posłajko that the category of belief cannot be neatly mapped onto any category likely to be posited by successful neurocomputational models. How, according to Dennett, can beliefs be real if they fail to map components of neurally implemented computations responsible for intelligent behavior? In Posłajko's terms, what "connection with fundamental properties" (p. 68) other than mapping components of successful neurocomputational models might ground realism about belief? Dennett's (1991) answer is clear: natural selection inevitably produces well-designed systems, i.e., *intentional systems*, systematically interpretable as believers:

[H]ow *could* the order be there, so visible amidst the noise, if it were not the direct outline of a concrete orderly process in the background? Well, it could be there thanks to the statistical effect of very many concrete minutiae producing, as if by a hidden hand, an approximation of the "ideal" order. Philosophers have tended to ignore a variety of regularity intermediate between the regularities of planets and other objects "obeying" the laws of physics and the regularities of rule-following (that is, rule-consulting) systems. These intermediate regularities are those which are preserved under selection pressure: the regularities dictated by principles of good design and hence homed in on by self-designing systems. That is, a "rule of thought" may be much more than a mere regularity; it may be a wise rule, a rule one would design a system by if one were a system designer, and hence a rule one would expect self-designing systems to "discover" in the course of settling into their patterns of activity. Such rules no more need be explicitly represented than do the principles of aerodynamics that are honored in the design of birds' wings. (Dennett, 1991, p. 43)

On Dennett's view, to treat a system as a believer, i.e., an "intentional system", is just to assume that it is optimally designed (Dennett, 1971): it can be usefully predicted on the assumption that it engages in the most rational behavior relative to a set of beliefs and desires.

This is what we typically do, for example, when playing chess against a computer program: rather than attempting to infer the computations responsible for its moves, we simply assume that it makes the most rational moves relative to

its beliefs about the rules of chess and chess board configurations, and its desire to checkmate us. This clearly supports surprising, robust predictions, and there is a mechanism which explains this: the work of its programmers. Similarly, in Dennett's view, we adopt this "intentional stance" (1987) toward other human beings and non-human animals. This practice supports surprising, robust predictions of naturally evolved systems, and there is a "fundamental" (Pośłajko, p. 68) mechanism which explains this: natural selection. For Dennett, it is not *just* that, as Pośłajko notes, "intentional interpretation reveals important behavioural patterns to us [...] attributions of beliefs seem to capture some important objective similarities" (p. 81). There is *also* a fundamental physical mechanism which explains this: natural selection. For this reason, Pośłajko underestimates Dennett's realism: he is more than "somewhat close to but not at the very end of the non-realist end of the spectrum of possible positions in the metaphysics of belief" (p. 81). Because he thinks natural selection inevitably discovers good designs, including intentional systems, Dennett thinks belief is as natural a category as any in the special sciences, and this seems consistent with Boyd's homeostatic property cluster view.

Now, as many have pointed out, Dennett's view is problematic for a host of reasons. Natural selection does not always discover optimal designs. The folk category of belief seems to differ significantly from the cleaned-up version Dennett formulates as part of intentional systems theory: the folk seem constrained neither by considerations of rationality nor by speculations concerning natural selection when attributing beliefs. As Pośłajko notes, the folk seem committed to the view that beliefs are concrete causes of behavior (p. 154), which Dennett explicitly denies (1991). Finally, Dennett acknowledges and accepts the fundamental indeterminacy of belief content implied by his view: "I see that there could be two different systems of belief attribution to an individual which differed substantially in what they attributed- even in yielding substantially different predictions of the individual's future behavior-and yet where no deeper fact of the matter could establish that one was a description of the individual's real beliefs and the other not" (1991, p. 49). In other words, to the extent that Dennett is a realist about belief, it is about a concept of belief that has undergone substantial revision relative to the folk concept. Pośłajko explicitly problematizes such revisionist proposals (Pośłajko, 2024, pp. 153–154); however, as I argue in the concluding section, it is unlikely that *any* theory of belief, including the representationalist theory Pośłajko claims is the paradigm of belief realism, can avoid radical revision of the folk concept. Hence, just the fact that the Dennettian notion of belief constitutes a revision of the folk notion is not enough to show that it is not a form of belief realism, grounded in a physical mechanism which is *not* neurocomputational, i.e., natural selection.

## 2B. Belief as Discursive Commitment

Posłajko considers another alternative to the representationalist theory of belief as grounding for belief realism: the idea that believer-that-*P* might constitute a social role, and hence a Boydian homeostatic property cluster maintained by social mechanisms. This idea derives from Mallon's (2016) proposal "that the social mechanism that creates a certain category might produce a 'range of effects that further differentiate putative members of the role'" (Mallon, 2016, p. 93), thus the category itself becomes a projectible and natural one. So even though category *X* is socially created in the sense that being an *X* is a socially ascribed social role, it might then turn out that all *X*s share important properties" (Posłajko, 2024, p. 161). Posłajko argues that such social mechanisms fail to ground the category of belief:

[W]hat would be needed is a story about what properties are shared by beliefs as a general category, and by people to whom we ascribe a belief with a specific content. Only by providing such a story could we claim to have positive reasons to think that beliefs are socially constructed natural kinds. However, this kind of evidence is precisely what we are lacking [...]. It is not that we know that there are no generalizations about belief; instead, we lack evidence to think that there are any, and this lack of evidence might be taken to support the view that beliefs should not, at the present moment, be conceptualized as being natural kinds. (Posłajko, 2024, p. 162)

However, I think Posłajko's skepticism here is premature. First, as noted above, the very fact that the folk reliably use the category of belief to make surprising, successful predictions constitutes evidence that there are generalizations about beliefs. Second, if we construe belief as a kind of "discursive commitment", in Brandom's (1994) sense, it is possible to identify plausible social mechanisms that ground this projectibility.

Brandom understands the attribution of states with propositional content in terms of playing "the game of giving and asking for reasons" (1994, p. 185). This is a fundamentally *deontic* practice: to attribute beliefs and other propositional attitudes is to situate subjects in a normative space of commitments and entitlements. For example, attributing the belief that Cleveland is north of Columbus involves attributing the whole constellation of other commitments and entitlements implied by this propositional content, e.g., commitment to Columbus being south of Cleveland and entitlement to Columbus being east (or west) of Cleveland. Now, although Brandom criticizes naturalistic accounts of propositional attitudes (1994), on the grounds that they fail to capture their normative import, he does ground discursive statuses in the deontic *attitudes* of those who attribute them: ultimately, interpretive targets have the discursive commitments and entitlements that attributors take them to have (Brandom, 1994). This interpretive practice is endlessly dynamic and recursive: discursive statuses are constantly being revised and negotiated, partly because their attributions also consti-

tute contestable moves in the game. However, such a game could never get off the ground if there were not some at least temporary régimes of interpretive consensus governing coordination within interpretive communities.

If we accept this general picture, then it is possible to identify a social mechanism grounding the reality of belief in the way Mallon proposes for social categories more broadly. We need only construe beliefs as commitments to sentences employed with enough interpretive consensus to facilitate coordination in specific linguistic communities. For example, we can *predict* that someone expressing commitment to Cleveland's being north of Columbus will also express commitment to, and otherwise act in accordance with, Columbus's being south of Cleveland, because the deontic attitudes of members of our community have shaped them appropriately. In effect, on this view, to believe that *p* is to adopt the very fine-grained social role of a believer-that-*p*, maintained in the same way as more canonical social roles, like motherhood: in virtue of deontic attitudes prevailing in communities. If we construe beliefs in this way, then they are as real as declarative sentences, linguistic communities, and deontic attitudes. Pośłajko himself acknowledges that "it seems perfectly reasonable to engage in inquiry into the patterns of language use" (Pośłajko, 2024, p. 143) in support of inferentialist theories of meaning like Brandom's (p. 143). And deontic attitudes might reasonably be made sense of in terms of Strawsonian "reactive attitudes" (Strawson, 1963), and "normative cognition" (Kelly, Westra, Setman, 2025) more broadly.

Thus, we have here another alternative to representationalism about belief: a social mechanism capable of grounding belief as a natural kind in Boyd's sense. As with the Dennettian appeal to natural selection inevitably producing intentional systems, this socially constructed yet fully natural belief category requires substantial revision of the folk concept. It is clear that the folk do not think of beliefs and other propositional attitudes as social statuses. Both Dennett's and the Brandom-inspired concepts of belief can be understood as products of conceptual engineering: ameliorative projects relative to the goal of explaining how beliefs can be real. In this sense, they are similar to Haslanger's (2012) conceptual engineering of concepts of race and gender, and Vargas's (2013) conceptual engineering of the concept of freedom of the will. Pośłajko explicitly notes the affinities between these projects and his own, as well as Haslanger and Vargas's rejection of anti-realism about race and free will, respectively (2024, pp. 128–133). However, he traces their refusal to give up realism to a conflation between anti-realism and eliminativism: they fail to appreciate his point that existing categories may fail to be real, in the sense of natural. This justifies his minimal non-realism about the *current folk* concept of belief: though beliefs exist, they are not real, even if revised versions, like Dennett's, or the Brandom-inspired one sketched above, or Gauker's (2021; Pośłajko, 2024, pp. 153–154) may be. However, as I argue next, in the concluding section, this attitude fails to appreciate that conceptual engineering is required for *any* naturalization project, and hence, in Pośłajko's view, any attempt to show beliefs are real.

### 3. Conclusion: If the Real Is the Natural then Revision Is Inevitable

Folk concepts are not known for their coherence and systematicity. This point has been appreciated since Carnap's (1945) discussions of "explication", a precursor to contemporary notions of conceptual engineering and ameliorative analysis. The folk use the same concepts for disparate purposes, with no concern for coherence or systematicity. For example, studies in "experimental philosophy" show that the folk concept of freedom of the will yields both compatibilist and incompatibilist intuitions depending on experimental design (Nahmias, 2014, p. 23, Footnote 10). Presumably, the same holds for the folk concept of belief. If we want a concept that is well behaved enough to attempt naturalization, revision of the unsystematic and incoherent folk concept is inevitable.

The question then becomes *which* aspects of the folk concept to maintain and which to jettison. Pośłajko often seems to assume that the representationalist theory of belief is a paradigm of realism about the *folk* concept. But this is obviously untrue. The folk are dualists. They are *not* computationalists. And they think concepts can be acquired through learning. These views are all denied by the most influential representationalists about belief. It is true that the folk seem committed to the view that beliefs are concrete, determinate causes of behavior, as representationalists assume. But why should *this* be the only aspect of the folk concept that matters for naturalization? If belief's status as a determinate, internal cause of behavior is all that matters, then representationalism seems to be the best candidate. But why is that aspect more important than, say, the folk assumption that beliefs cannot be identified with computational roles? This aspect of the folk concept is well grounded in appeals to intuition—not just reluctance to attribute beliefs to artificially intelligent systems,<sup>1</sup> but even basic intuitions that individuals in radically different computational states might share beliefs (Dennett, 1987). If we put more weight on the folk intuition that beliefs cannot be explained in terms of computational roles than on the folk intuition that beliefs are concrete causes of behavior, then Dennett's view that beliefs are inevitable features of optimally designed systems "discovered" by natural selection becomes a more viable route to naturalization.

Similarly, although the folk clearly do not treat beliefs as normative statuses, many aspects of the folk practice of attributing beliefs are more consistent with this view than the view that belief attributions are causal hypotheses. Consider the phenomenon of *akrasia*: one continues to attribute to oneself intentions that

---

<sup>1</sup> As one reviewer of an earlier draft notes, this seems in tension with my earlier endorsement of Dennett's observation that the folk adopt the intentional stance toward chess playing computers. The folk appear to be inconsistent in this regard: reluctant to explicitly attribute beliefs to artificially intelligent systems, while implicitly making this assumption in the context of game play. This is more grist for my mill: the folk concept of belief is incoherent and unsystematic. In any case, it is not clear that the folk adopt the intentional stance toward chess playing computers *in virtue of* the computational etiologies of their moves.

are constantly confuted by behavioral evidence, e.g., a compulsive smoker's intention not to smoke, presumably because such self-attributions are more like expressions of commitment than causal hypotheses. Another relevant phenomenon is the central justificatory role of belief attribution noted above (Malle, 2006; Malle, Knobe, Nelson, 2007). Belief attribution in general is deeply unlike the attribution of non-mentalistic causes in the following way: behavioral confutation of a belief attribution never *requires* its revision. One can always demand that the interpretive target revise their behavior to make it consistent with the attribution. For example, if someone tells you they believe a team will lose a game, but then proceeds to bet on the team, one need not revise the original attribution based on their avowal; one can always demand that they revise their betting behavior or at least explain the inconsistency. This is a significant disanalogy with the attribution of non-mentalistic causes: if one conjectures that one has COVID based on symptoms, but then the test shows up negative, the original conjecture must be withdrawn; there is no pressure to explain the inconsistency in terms of rational norms. The point here is not that the view that folk belief attributions are causal hypotheses is *false*; the point is that this is not *all* they are. If the road to naturalization requires ignoring these non-causal aspects of belief attribution, then it requires no less revision than proposals which ignore its causal aspects.

The upshot is that if belief realism requires naturalization, as Pośłajko assumes, then either the folk concept is too incoherent to be real, or belief realism will require making hard calls about which aspects of the folk concept to maintain and which to jettison. If beliefs must be concrete causes of behavior, then representationalism appears to be the best bet right now, and Pośłajko's arguments for minimal non-realism have bite. But if it is more important to agree with the folk that beliefs are *not* computational states, then Dennettian intentional systems theory, grounded in natural selection, seems more plausible, and Pośłajko's arguments are irrelevant to this kind of belief realism. And, if it is more important to agree with the folk that beliefs play normative roles, then a Brandom-inspired analysis in terms of discursive statuses maintained through social mechanisms seems more plausible, yielding another form of belief realism resistant to Pośłajko's arguments.

## REFERENCES

- Boyd, R. (1991). Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 61(1/2), 127–148.
- Brandom, R. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.
- Carnap, R. (1945). The Two Concepts of Probability: The Problem of Probability. *Philosophy and Phenomenological Research*, 5(4), 513–532.

- Dennett, D. C. (1971). Intentional Systems. *The Journal of Philosophy*, 68(4), 87–106.
- Dennett, D. C. (1987). *The Intentional Stance*. MIT press.
- Dennett, D. C. (1991). Real Patterns. *The Journal of Philosophy*, 88(1), 27–51.
- Gauker, C. (2021). Belief Attribution as Indirect Communication. *Groups, Norms and Practices: Essays on Inferentialism and Collective Intentionality*, 173–187.
- Haslanger, S. (2012). *Resisting Reality: Social Construction and Social Critique*. Oxford University Press.
- Kelly, D., Westra, E., Setman, S. (2025). The Psychology of Normative Cognition. *Stanford Online Encyclopedia of Philosophy*.
- Khalidi, M. A. (2013). *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*. Cambridge University Press.
- Malle, B. F. (2006). *How the Mind Explains Behavior: Folk Explanations, Meaning, and Social Interaction*. MIT press.
- Malle, B. F., Knobe, J., Nelson, S. E. (2007). Actor-Observer Asymmetries in Behavior Explanations: New Answers to an Old Question. *Journal of Personality and Social Psychology*, 93, 491–514.
- Mallon, R. (2016). *The Construction of Human Kinds*. Oxford University Press.
- Nahmias, E. (2014). Is Free Will an Illusion? Confronting Challenges From the Modern Mind Sciences. In W. Sinnott-Armstrong (Ed.), *Moral psychology* (Vol. 4, pp. 17–40). MIT Press.
- Posłajko, K. (2024). *Unreal Beliefs: An Anti-Realist Approach in the Metaphysics of Mind*. Bloomsbury Publishing.
- Strawson, P. (1963). Freedom and Resentment. *Proceedings of the British Academy*, 48, 187–211.
- Vargas, M. (2013). *Building Better Beings: A Theory of Moral Responsibility*. Oxford University Press.



WITOLD M. HENSEL \*

## **TROUBLES WITH THE NATURAL-KIND APPROACH TO THE DEBATE OVER THE REALITY OF BELIEF<sup>1</sup>**

**SUMMARY:** In the paper, I criticize the contemporary natural-kind approach to the reality of belief. My first argument is intended to show that the framework's focus on concepts, to the exclusion of theories and practices, conjoined with causal-historical views on the reference of natural-kind terms, has been pushing the debate into irrelevance by stripping it of content. My second argument aims to show that the natural-kind framework offers an illusory sense of tractability. On weak accounts of natural kinds endorsed by Pośłajko, the central question of whether belief is a natural kind might be tractable, but these accounts prove too weak to secure the conclusion that natural kinds are real. Strong accounts, by contrast, are descriptively inaccurate. At the end, I suggest that there is also a tradeoff between tractability and validity of the natural-kind framework.

**KEYWORDS:** folk psychology, belief, natural kinds, belief realism, belief anti-realism, objectivity, scientific realism, conventionalism.

### **1. Introduction**

According to Pośłajko (2024), contemporary debates over the reality of beliefs concern, at bottom, the question of whether belief is a reasonably natural kind. In fact, as he demonstrates, many discussions aim to determine the extent

---

\* University of Warsaw, Faculty of Philosophy. E-mail: w.hensel@uw.edu.pl. ORCID: 0000-0003-1786-3622.

<sup>1</sup> I would like to thank two anonymous referees of this journal for valuable feedback and Maja Kittel for helping me see how to make the manuscript clearer and more readable.

to which belief meets the criteria of a natural property proposed by Lewis (1983). The concept of belief, therefore, denotes a real entity just in case it is: 1) reasonably objective, rather than conventional, 2) causal-explanatory, rather than classificatory or normative, and 3) explainable in terms of fundamental properties. The first two criteria are supposed to be explicated by accounts of natural kinds such as (Khalidi, 2013) whereas the third corresponds with the project of naturalizing mental content (cf. Pośłajko, 2024, p. 55).

The proposed reconstruction boasts many advantages. Not only does it provide a framework for locating Pośłajko's position among its rivals, and show how the various arguments fit together, but it also throws into sharp relief key features of the debate. While I agree with Pośłajko that explicating the reality of belief in terms of natural-kindness reveals the structure of the debate, I have serious reservations about a further claim he makes about this approach: that it renders the central problem tractable, at least by philosophical standards.

Granted, the approach looks attractive. It breaks down a metaphysical puzzle into well-defined problems corresponding to the three criteria of a natural property accepted, as it happens, by nearly all parties. I also submit that the first two of those problems appear manageable. The third, which calls for assessing the viability of naturalizing propositional mental content, might be an exception.

It is hardly surprising that the debate has been conducted according to this schema. Yet, after over thirty years, I think the time is ripe to take stock and reevaluate.

My entry point to the discussion, in Section 2, is Churchland's eliminative materialism—the first concentrated attack on the reality of belief mounted from the positions of strong scientific realism. Churchland did not focus on concepts, but on theory (folk psychology). I explain the merits of this choice and contrast it later with the concept-centric approach favored by Pośłajko. I also reconstruct the philosophical community's reasons for resisting Churchland's theory-centered construal of the belief debate.

Section 3 introduces what I call *the problem of disappearing substance*: because realists focus on the concept of belief and its presumed natural-kindness, they tend to reply to criticisms by dropping “inessential” commitments and pretending that such concessions leave their position intact. I argue that this have-your-cake-and-eat-it dynamic, conjoined with neglecting embedding laws, strips the debate of content, rendering it irrelevant. In Section 4, I propose a solution, which is, essentially, to reunite the concept of belief with embedding laws.

Section 5 is devoted to what I call *the problem of disappearing reality*. In a nutshell: the strong essentialist accounts of natural kinds adopted to fend off the specter of eliminative materialism have become unacceptable, plagued by a plethora of counterexamples from scientific practice in many disciplines, including “hard” fields such chemistry and biology. However, realists cannot simply help themselves to moderate or weak accounts because those accounts fail to distinguish between objective and conventional groupings of objects. In effect,

establishing that belief is a natural kind, according to moderate or weak accounts, does not secure the conclusion that beliefs are real.

Section 6 offers a brief and somewhat speculative discussion of the likely reasons behind the idea that reality of a theoretical category could profitably be analyzed in terms of natural-kindness as it is explicated by the current literature on taxonomic practices. I suggest a fundamental tension between realist commitments and the tractability of assessing natural-kindness. The gist is that favoring tractability leads to natural-kindness's failure to be diagnostic of reality, whereas favoring realist commitments renders the problem of establishing natural-kindness intractable.

## 2. Setting the Stage

My point of departure in what follows will be a fairly strong realist attitude that seems to have dominated philosophy of mind since the early 1960's, when identity theory (Feigl, 1958; Smart, 1959) appeared on the scene (cf., e.g., Kim, 1989). I am also assuming, throughout the paper, that scientific realism is an overarching empirical hypothesis, supported, according to its proponents, by some version of the no-miracle argument (Putnam, 1978). The approach is often called convergent realism (Laudan, 1981). I am treating it as a default position not because I subscribe to it, but because it provides a clear rationale for a causal natural-kind semantics and, as far as I can tell, continues to be the majority view (Ghins, 2024; Leplin, 1984; Niiniluoto, 1999; Park, 2022; Psillos, 1999; Putnam, 1975; 1978).

Until the 1980's, virtually all discussions of the conditions and consequences of establishing whether selected psychological predicates refer to brain states took it for granted that those predicates actually refer (except, e.g., Feyerabend, 1963). Even now, the idea that a term like "belief" may lack reference appears ridiculous to many, perhaps most, writers. Pośłajko's (2024) defense of minimal non-realism poignantly illustrates this, reframing, as it does, the central problem in terms of the reality of belief, not its existence or the reference of the corresponding term.

Let me grant that, taken in isolation, asserting that "belief" lacks reference flies in the face of reason. But it acquires meaning in a broader context. So, if we do not wish to dismiss anti-realism about belief out of hand, we cannot define it as a position only about the concept of belief. Instead, we should treat that concept as part of a bigger whole. According to Churchland, that bigger whole is the theory of folk psychology. Although, after over forty years, it is possible to provide convincing extensions of this context—e.g., research tradition (Laudan, 1977), field (Darden, Maull, 1977), and scientific practice (Chang, 2011)—I will describe the gist of eliminative materialism in its original terms, grounded in traditional philosophy of science.

From the perspective of traditional philosophy of science, debates in philosophy of mind revolve around two imaginary theories that science can potentially

yield in the limit of inquiry: a “mature psychology” and a “completed neuroscience”, as Churchland (1981) called them. Both are true by idealizing assumption, at least according to scientific realists.

Now, a crucial question the scientific realists need to address is whether the actual accounts of mind and brain resemble the two imaginary theories *sufficiently closely to be approximately true*.

Unlike the question about the truth-value of “mature” psychology, the question about the approximate truth of the best available account of the mind is a matter of fact. The same, *mutatis mutandis*, goes for current vs “completed” neuroscience, but discussing them would take us too far afield. I suspect that Churchland used the term “completed”, rather than “mature” neuroscience, because he assumed current neuroscience to have reached approximate truth. For our purposes, though, it will suffice to interpret “mature” and “completed” as alternative words for approximately true.

As Churchland (1979; 1981) observed, interpreted realistically, any theory is either true or false, so it is logically possible that most tenets of folk psychology are false. It is therefore logically possible that folk psychology is not approximately true. But if folk psychology is not approximately true then, arguably, its ontological commitments lack license. In consequence, the corresponding terms, including “belief”, may lack reference. But that is not the main point.

The main point is that, if the approximate truth of folk psychology is in doubt, the central ontological posits of folk psychology lack warrant, and unwarranted posits have no place in our ontology, so they should be eliminated from our worldview. Eliminative materialism (eliminativism) asserts that, given available evidence, the central ontological posits of folk psychology, in fact, lack sufficient warrant. How that should affect ordinary language or social practices involving the concept of belief are side issues.

Because Churchland’s arguments for eliminativism are familiar, I will not rehearse them here. What is important for my purposes is that, although Churchland identified propositional attitudes as especially suspect, his criticisms targeted folk psychology as a whole, rather than its concepts. Churchland understood that, by treating folk psychology as a theory, one could evaluate it according to rigorous epistemic criteria used in science. That is also why he expressed the eliminativist claim in terms of inter-theoretic reducibility: common-sense psychology, he wrote, is “a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced, rather than smoothly reduced, by completed neuroscience” (Churchland, 1981, p. 67).

According to strong scientific realists, we are justified in accepting theoretical posits if they feature in a highly empirically successful and approximately true theory: that is, in a theory whose empirical success can be explained by its approximate truth, as per convergent realism (cf. Laudan, 1981).

Concepts can contribute to the overall success of a theory. Yet they act as cogs in a much larger machine, which is why evaluation of theoretical concepts involves embedding them in some theory and then gauging how that theory fares

against the competition. Even extracting a concept from the laws of a theory calls for questionable decisions, especially if the theory lacks a regimented formulation—a feature that has always plagued the debate over eliminativism. Of course, today, we would add that theories are themselves embedded in various scientific practices, including methods of collecting and analyzing data.

The reactions to Churchland's eliminativism have been overwhelmingly negative. Critics seem to have attacked every premise and every assumption of Churchland's reasoning, as well as a fair number of strawman positions (Baker, 1988; Boghossian, 1990; Dennett, 1991; Fletcher, 1995; Fodor, 1987; Gordon, 1986; Horgan, Woodward, 1993; Lycan, 1988; Ratcliffe, 2007; Smolensky, 1988; Stich, 1996). When the dust settled, realism appears to have won by "a sort of philosophical walkover: as the extreme version of anti-realism is treated as a non-starter, there is little pressure to engage in disproving it" (Pośłajko, 2024, p. 7).

Crucially, the rejection of eliminativism coincided with a return to the earlier concept-centered approach, as philosophers of mind accepted the idea that beliefs can be real regardless of the fortunes of common-sense psychology or any other embedding theory. The logic of this move is aptly captured by the Lycan-Stich argument (Lycan, 1988; Stich, 1996):

- 1) "Belief", whether it occurs in common-sense or scientific explanations of behavior, is a natural-kind term.
- 2) A theory featuring a natural-kind term may contain numerous false generalizations, and even misidentify key features of the natural kind, but the natural-kind term nonetheless picks out those samples of matter that really fall under that natural kind.
- 3) Natural kinds are salient mind-independent nodes in the causal structure of the world, which is to say they carve out objects along nature's joints.
- 4) Therefore, the term "belief" likely refers; beliefs exist and are real.

This appeared convincing for a number of reasons. Churchland's research program, calling for formal reconstructions of theories, had proved unworkable. We lack a satisfactory formulation of folk psychology, so it is impossible to measure the conceptual distance between it and any other theory of mind, even if we could agree on a model of reduction. But interest in scientific theory and models of reduction had already been waning in the 1970's, as philosophers of science were investigating other elements of scientific practice, on the one hand, and exploring the disunity of science, on the other (Cartwright, 1999; Dupré, 1993; Fodor, 1974).

Importantly, causal-historical accounts of reference and natural kinds had immediately become part of the realist package. According to Psillos, for example, the metaphysical component of scientific realism "asserts that the world has a definite and mind-independent natural-kind structure" (1999, p. vii). Indeed, the account of reference for natural-kind terms had been introduced together with

“flights to reference” (Bishop, Stich, 1998) similar to the Lycan-Stich argument in (Putnam, 1978); later instances include (Cummiskey, 1992; Hardin, Rosenberg, 1982; Kitcher, 1993).

### 3. Natural Kinds and the Problem of Disappearing Substance

Motivated by the view that laws of an embedding theory do not impose significant constraints on the reference of natural-kind terms, the move toward natural kinds has produced the following have-your-cake-and-eat-it dynamic:

- 1) A realist proposes an account of beliefs and argues that we have sufficient evidence for asserting that beliefs, as characterized by that account, are real.
- 2) An anti-realist criticizes the realist’s position by showing that we have strong reasons to doubt whether anything actually meets the conditions the realist account imposes on beliefs, so beliefs are not real.
- 3) A second realist comes along and observes that, although the anti-realist is right to contend that beliefs, as described by the first realist, are not real, it would be a mistake to adopt anti-realism. The rub is that the anti-realist has confused a *conception* of belief, proposed by the first realist, with the *concept* of belief, which is independent of any particular conception, because “belief” is a natural-kind term. Obviously, beliefs need not have feature *F*, which happens to be responsible for the untenability of the first realist’s account.

Let me illustrate this with two examples. The first comes from the “old” debate over beliefs whereas the second—from the more recent debate over mental representations in cognitive science.

#### 3.1. First Example

An important tenet of folk psychology, as initially reconstructed by both realists (e.g., Fodor, 1975) and eliminativists (Churchland, 1979, Chapter 5; Stich, 1983), was that thinking consists in constructing arguments couched in a language of thought. The link between language and thought, in turn, was supposed to buttress traditional sentential accounts of epistemic and instrumental rationality, endorsed by Fodor and rejected by Churchland. According to those accounts, an organism thinks rationally to the extent that the internal arguments it produces are valid and sound; and, by practical syllogism, an organism behaves rationally to the extent that its internal states, produced by rational reflection, determine the organism’s behavior.

To paraphrase Fodor (1987), common-sense belief/desire psychology is a deep account of the mind, because it explains the parallels between reasons and causes of actions. Belief, on this view, serves as a bridge between the norms of rationality and physical behavior (cf., e.g., Postajko, 2024, pp. 4–5). That,

more than anything else, makes the question of its reality an important philosophical problem.

Given the significance of the perceived link between language and thought to common-sense psychology, it was reasonable for eliminativists to attack belief realism by undermining Fodor's language-of-thought hypothesis. For example, in a well-known paper, Ramsey, Stich, and Garon (1991) argued that, if a certain type of connectionist architecture proves superior to classical symbol-based models of cognition, then propositional attitudes, understood as functionally discrete, semantically evaluable and causally active states, might disappear from our ontology (see also Garfield, 1997).

However, in reply to Ramsey, Stich, and Garon's (1991) argument, Heil defended belief realism by observing that

it is scarcely obvious that ordinary appeals to beliefs and other propositional attitudes *necessarily* require such things to be discrete entities. For some theorists, notably Fodor, they certainly seem to be [...]. But [...] Fodor's is not the only game in town. (Heil, 1991, p. 129)

He concluded that

pitting connectionism against folk theory provides scant support for eliminativism. (Heil, 1991, p. 132)

Heil was not alone. Armed with the idea that "belief" is a natural-kind term, the realist is free to cast doubt not only on the discrete character of beliefs, but also on other aspects of the putative link between language and propositional attitudes. For example, Loar (1981; 1982) appeals explicitly to natural-kindness to argue for a non-linguistic character of beliefs, although he does not cite eliminativist critiques of Fodor's Mentalese (for a rejoinder, see Fodor, 1982).

### 3.2. Second Example

Critics of belief realism, such as Schwitzgebel (2001; 2002; 2010; 2013) and Jenson (2016), attack two features attributed to beliefs by folk psychology and accepted by many contemporary belief realists, including Fodor (1987), Loar (1981) and Nichols and Stich (2003).

The first feature, call it *uniqueness*, is that a person who believes that *p* must stand in a particular relation to a single internal representation that affects both what they say and what they do. Barring deception, then, people who profess to believe that *p* but act in ways inconsistent with believing that *p* do not really believe that *p* (hence, determining whether someone believes that *p* involves an analysis of what they say and what they do).

The second feature, call it *approximate coherence*, is that most people's beliefs and desires form a largely consistent set because, otherwise, the practical syllogism would rarely yield successful explanations of human behavior. Alt-

though a person may, on occasion, hold contradictory beliefs, belief/desire psychology is silent on what kind of behaviors can be expected to occur in such cases. As a result, confronted with a person holding apparently contradictory beliefs on a subject relevant to understanding that person's actions, folk explanations of those actions usually fall back on non-propositional factors such as personality traits and emotions.

Jenson (2016) marshals a broad range of theoretical and empirical arguments to show that a reasonable notion of belief in cognitive science requires both uniqueness and approximate coherence, and that, in light of available findings, it is implausible to hold that any internal, causally active mental states possess both properties. The conclusion he draws is that belief is a likely candidate for elimination from cognitive science.

However, a realist can handle Jenson's critique by simply dropping uniqueness or approximate coherence as properties of belief. After all, if belief is a natural kind then it need not possess any property attributed to it by any available theory, whether folk or scientific. I am going to illustrate this response by appeal to Quilty-Dunn and Mandelbaum's (2018) argumentation for representationalism. Although Quilty-Dunn and Mandelbaum (2018) answer Schwitzgebel's charges against belief realism, not Jenson's, the arguments involved are sufficiently similar.

According to Quilty-Dunn and Mandelbaum (2018), beliefs may be *fragmented*, or stored at many separate locations. This implies that a person can have multiple mutually incompatible beliefs about a state of affairs and behave in a way that conflicts with a belief they profess to hold. In other words, I may have the belief that Marty loves cheesecake stored in one part of my brain, and the belief that Marty hates cheesecake stored in another part of my brain. Depending on various contextual factors, some of my actions and utterances are caused by the representation *Marty-loves-cheesecake* and some other of my actions and utterances are caused by the representation *Marty-hates-cheesecake*.

As Quilty-Dunn and Mandelbaum (2018) demonstrate, the idea of storing semantically evaluable representations of the same state of affairs at separate information silos not only coheres with existing empirical evidence but can also contribute to explaining it. Interestingly enough, some of this evidence speaks against the standard accounts of rationality. Available findings indicate that people acquire beliefs automatically, by simply accepting whatever they hear or read; only belief rejection involves effortful critical thought, which can be disrupted by cognitive load (Porot, Mandelbaum, 2021). I would add that, worse still, achieving coherence in a large set of representations seems computationally intractable, so our belief systems likely abound in inconsistencies (Sommer, Musolino, Hemmer, 2023).

Given that Quilty-Dunn and Mandelbaum (2018) distance themselves from folk psychology while Jenson (2016) seems to criticize it, the two positions might appear to be mutually compatible. Yet Jenson (2016) explicitly argues for the elimination of belief (not common-sense psychology) from cognitive science

(not lay discourse or philosophy)—a claim that conflicts directly with Quilty-Dunn and Mandelbaum’s (2018) realist thesis (see also Porot, Mandelbaum, 2021). While Jenson (2016) argues for the centrality of uniqueness and approximate coherence to explanations based on propositional attitudes, Quilty-Dunn and Mandelbaum (2018) insist that uniqueness and approximate coherence are merely posits of a false theory of belief.

\* \* \*

Having given themselves permission to abandon theoretically significant claims about belief without weakening their position, many realists have been tempted to do so. In their view, they merely discarded empirical hypotheses of secondary importance. On the assumption that “belief” is a natural-kind term, it is impossible, however, to determine which features attributed to beliefs, if any, should enjoy the status of essential commitments of belief realism and which serve as auxiliary empirical hypotheses. In effect, the question of the reality of beliefs becomes practically intractable.

The problem, though, is not that increasingly impoverished characterizations of belief make substantive argumentation ever more difficult for the anti-realist. The problem is that the have-your-cake-and-eat-it dynamic dilutes the metaphysical claims involved, pushing the discussion into irrelevance. It siphons off philosophically interesting content.

Therefore, while I applaud Pośłajko for characterizing belief in the thinnest possible terms acceptable to any belief realist, this tactic backfires by removing the motivation to pursue the central question of the debate. If belief is understood as any internal, causally active state endowed with propositional content (Pośłajko, 2024, p. 31) then the putative reality of beliefs coheres with a vast array of highly diverse accounts of mind and rationality, many of which conflict, at a fundamental level, with common-sense psychology and received philosophical wisdom.

But, if that is the case, why should we even care?

#### **4. How to Deal With the Problem of Disappearing Substance**

Luckily, the problem of disappearing substance can be easily fixed. Essentially, all we have to do is reunite the concept of belief with its natural companions—the putative laws of psychology in which it occurs (and, ideally, also research traditions, fields, and practices). It is those laws that connect the notion of belief with the phenomena it is intended to explain, giving substance to belief realism, on the one hand, and enabling empirical assessment, on the other.

This is a conservative remedy, equivalent to resetting the debate to one of its original parameters. In a well-known paper, Fodor defined strong realism about propositional attitudes as asserting the existence of semantically evaluable “mental states whose occurrences and interactions cause behavior *and do so*,

moreover, in ways that respect (at least to an approximation) the generalizations of common-sense belief/desire psychology” (Fodor, 1985, p. 78, emphasis added).

Since an agreed-upon reconstruction of common-sense belief/desire psychology is missing, I propose that we treat philosophers’ and psychologists’ considered opinions on the subject as reference points, acknowledging the diversity of offered accounts. This is what we do anyway. For the time being, and perhaps forever, the plurality of plausible accounts of common-sense belief/desire psychology is a fact of life that we have to accept, frustrating as it might be.

Although the laws in question may deviate from common-sense belief/desire psychology, a version of realism about belief can be located in the debate only if its proponents articulate the relationship they take to hold between their account of belief and folk psychology. In fact, that is probably the main difference between my theory-oriented approach and Pośłajko’s concept-centered framework. Roughly speaking, on a theory-oriented approach, realism about belief may come in two distinct varieties depending on its attitude toward common sense. On the one hand, there are the traditional mental-content theories that attempt to naturalize mental content with the aim of vindicating folk psychology, and on the other, there are the revisionists that expressly reject common-sense belief/desire psychology (cf. Rupert, 2018, p. 206).

As an example of revisionist belief realism, consider Quilty-Dunn and Mandelbaum’s (2018) representationalism. On a concept-oriented approach favored by Pośłajko, Quilty-Dunn and Mandelbaum are strong belief realists, just like Fodor and other friends of folk psychology, because they claim that belief serves an important explanatory function in current cognitive science and social psychology (see also Porot, Mandelbaum, 2021). Yet, on Fodor’s (1985) classical definition of strong realism, they are not, because the generalizations accepted as approximately true by Quilty-Dunn and Mandelbaum conflict with those of folk psychology. I would argue that my theory-oriented approach captures something relevant to the debate that Pośłajko’s framework ignores. That something is the fact that, unlike traditional belief realism, revisionist realisms, such as Quilty-Dunn and Mandelbaum’s representationalism, do not conflict with eliminativism.

This is not merely a point of logic. Churchland—the most famous eliminativist—defends representationalism, and has done so unwaveringly, as far as I can tell, since the beginning of his career (the title of his latest book is *Plato’s Camera: How the Physical Brain Captures the Landscape of Abstract Universals*; Churchland, 2013). In other words, on a purely concept-oriented framework, Churchland espouses both a version of strong realism about belief (a representationalism similar to that advocated by Quilty-Dunn and Mandelbaum) and a version of strong anti-realism about belief (eliminative materialism)!

The contradiction in Churchland’s position is illusory, of course. The eliminativist thesis concerns the ontological posits of common-sense belief/desire psychology, not those of just any old account that happens to employ the term “belief”. Strictly speaking, then, Quilty-Dunn and Mandelbaum’s realism about belief and Churchland’s eliminativism are mutually independent.

In fact, Quilty-Dunn and Mandelbaum's representationalism may lend support to belief realism or belief anti-realism depending on theoretical context. On the one hand, realists, as we saw, may claim that Quilty-Dunn and Mandelbaum's account is approximately true, in which case beliefs come out as real. On the other, eliminativists can argue that the existence of Quilty-Dunn and Mandelbaum's account confirms Churchland's prediction that scientific psychology will, over time, move beyond folk psychology and, eventually, develop a superior theoretical framework that will lack the category of belief. In a similar vein, one can argue that we are already witnessing the conceptual fragmentation of lay "belief" (Taylor, Vickers, 2017).

### **5. Theories of Natural Kinds and the Problem of Disappearing Reality**

My second argument against reframing the belief debate in terms of reasonable natural-kindness is more serious and more complicated than the first. Here is the gist:

Acceptable criteria of natural-kindness are constrained by commitments to scientific realism, on the one hand, and descriptive accuracy, on the other. When these constraints are considered separately, the situation is unremarkable. Strong scientific realists have good philosophical reasons to adopt only strong accounts of natural kinds, while moderate and weak scientific realists attempt to retain only those features of strong accounts that are necessary to support their version of realism. When it comes to descriptive accuracy, the literature on scientific taxonomies has been steadily gravitating toward relaxing the notion of natural-kindness in order to accommodate a growing number of real-life counterexamples to stronger accounts (Chakravartty, 2023; Craver, 2009; Dupré, 1981; 1993; Havstad, 2018; Hendry, 2006; Khalidi, 2013; Ludwig, 2018; Tobin, 2010).

At present, the literature indicates that only the weakest accounts of natural kinds achieve acceptable levels of descriptive accuracy. However, from the perspective of the realism debate, these weak accounts cannot distinguish between objective and conventional groupings of objects. The criteria of naturalness those accounts propose coincide with factors that anti-realists cite to explain the success of science (Chakravartty, 2023). The upshot is that rating belief as a reasonably natural kind, based on weak accounts of natural-kindness endorsed by Pośłajko, fails to justify the claim that beliefs are real. On the flip side, rating belief as non-natural serves to undermine the realist thesis.

In order to make my case, I will provide a brief overview of recent accounts of natural kinds, showing the key tensions between realist desiderata and scientific practice that have led theorists to gravitate toward ever weaker accounts (for a historical overview of natural-kind theories, see Hacking, 1991). In the process, I will highlight the most relevant connections between accounts of natural kinds, scientific realism, and realism about belief.

Following the publication of (Kripke, 1972) and (Putnam, 1975), scientific realists envisaged natural kinds as mind-independent, scientifically discoverable divisions of objects in nature that formed a unique hierarchy such that if an object fell under two natural kinds then one of those kinds had to be subsumed under the other (cf. Khalidi, 2013, p. 69; Tobin, 2010, p. 179). The objectivity and the uniqueness of natural-kind structure dovetailed nicely with causal-historical accounts of reference because, once a natural-kind term had been properly introduced, successive scientific theories could be expected to converge, in the long run, on the nature of the term's referents.

This picture of natural kinds was traditionally conjoined with *non-gradability* (natural kinds have sharp boundaries) and *essentialism* (all members of a natural kind satisfy a specific set of necessary and sufficient conditions, discoverable, in the long run, by science). Lastly, scientific realists often leaned toward *micro-essentialism*, according to which the properties of higher-level natural kinds are determined and explained by the micro-structural properties shared by their members—hence the widespread use of examples such as tiger and H<sub>2</sub>O.

On strong accounts, belief is a natural kind if it jointly satisfies the following conditions:

- 1) It has sharp boundaries—it is purely a matter of objective fact whether a particular mental state tokens a belief, rather than a supposition, and whether its content is *Marty loves cheesecake* or *Marty likes cheesecake*;
- 2) Its properties can be explained by the properties and interactions of lower-level kinds.
- 3) It features in a unique hierarchy.

However, as to 1, many respectable scientific kinds have blurred boundaries. These kinds include an indeterminate number of chemical compounds (e.g., water; Havstad, 2018) and biological species (e.g., tiger; Zachos, 2016). As to 2, most scientific kinds have not been reduced, in an appropriate way, to lower-level kinds. Quantum chemistry, for example, cannot explain the chemical properties of atoms and molecules in purely physical terms (Chang, 2015). As to 3, there is no single concept of acid (Hendry, 2006) and over twenty concepts of biological species exist side by side, each giving rise to a different but legitimate taxonomy (Zachos, 2016). Given the strength of conditions 1–3, it is doubtful that science has discovered any natural kinds in the required sense.

The scientific realist can, of course, insist that science will eventually settle on a unique hierarchy of kinds, taking solace from such developments as widespread acceptance of De Queiroz's (2007) proposal to focus on species delimitation, not species concepts. That, however, would amount to issuing a promissory note and abandoning the idea that realism explains empirical success. The more usual option is to offer a less demanding view of what constitutes a natural kind.

According to an influential account proposed by Boyd (1989; 1991; 2010), natural kinds are homeostatic property clusters (HPC), or sets of inductively

important properties held together by causal mechanisms. As such, members of a natural kind need not share essential properties, and natural kinds themselves need not have sharp boundaries or feature in a unique hierarchy.

Although the HPC view abandons most features proposed by strong accounts of natural kinds—dropping uniqueness and non-gradability, and replacing essentialism and micro-essentialism by an identifiable causal mechanism—it has been depicted as an attractive mid-way position between essentialism and conventionalism (Craver, 2009). However, applying it to psychological constructs would require identifying specific mechanisms responsible for holding the relevant properties together, which, at present, is a tall order. At any rate, Pośłajko (2024, p. 93) recommends that we focus on even more liberal proposals.

That is how we arrive at weak accounts of natural-kindness. According to Khalidi's simple causal theory, "a natural kind is associated with a set of properties whose co-instantiation causes the instantiation of other properties" (2013, p. 80). Causal efficacy of natural kinds admits of degrees in several respects, such as *inductive richness* (number of connections with other natural kinds), *robustness* (range of background conditions under which those connections obtain), and *strength of causal connection* (deterministic vs statistical).

On such a construal of natural-kindness, realists about psychological kinds are back in the game. To justify the claim that belief is a natural kind, they have to show that the notion of belief can be expected to feature in a "considered scientific theory [...] at least in the limit of inquiry" (Khalidi, 2013, p. 42). A standard way of doing so involves arguing that the notion of belief occurs in well-confirmed empirical generalizations that are unlikely to get rejected.

But adopting a moderate or weak account of natural kinds comes at a cost. If no unique natural-kind structure exists then the picture of scientific investigation associated with realism and the causal-historical account of reference collapses. Figuratively speaking, we cannot first sink our referential hooks into a natural kind that awaits discovery and then slowly reel it in to understand its essence. Not unless we already know the domain of investigation in sufficient detail to be able to solve the *qua* problem. Rather, the picture that emerges, especially in the special sciences, is that a typical natural-kind term denotes a long disjunction of crosscutting natural kinds, each of which may potentially turn into a focus of research and thereby become recognized as a natural kind. Which of those kinds is investigated, when, and for how long depends on the vagaries of the scientific process, not on the world.

Moderate and weak views on natural kinds do not provide a framework for distinguishing between realism and conventionalism because the criteria of natural-kindness proposed on those accounts appeal only to what is accepted as a given by both approaches. The main indicator of naturalness, on both realism and conventionalism, is inductive success. However, the issue that divides realists and conventionalists is not whether inductive success exists, but rather, how best to explain it (for a detailed argument, see Chakravartty, 2023).

That is why, even despite professing realism, proponents of moderate and weak accounts of natural kinds flirt with, or lapse into, conventionalism. Chakravartty (2023) provides several quotations to prove the point. I will restrict myself to one: “natural kinds are social constructions suited to the inductive/explanatory projects of particular disciplines. In a certain sense they are mind, interest, and project dependent” (Boyd, 2010, p. 220).

Because moderate and weak accounts of natural kinds cannot distinguish between objective and subjective groupings of objects or properties, the very notion of naturalness that they allegedly explicate remains unclear. As Chakravartty writes,

once we appreciate the depth of connection between positing kinds and inductive success, scientific and other classificatory practices inevitably drive kind realism in a more pluralistic direction than was once the norm. This realization, however, immediately raises a serious question regarding whether there are any constraints at all, beyond inductive success, on what is properly recognized as a kind as per realism. (Chakravartty, 2023, p. 77)

Apparently, then, explicating the reality of belief in terms of natural-kindness is not such a great idea, after all. The criteria of objective similarity and causal efficacy furnished by moderate and weak accounts of natural kinds may render the reality question tractable, but cannot deliver the goods. In a way, that only strengthens Pośłajko’s case, revealing for all to see the infirmities of realist arguments. Complacency breeds failure, as they say.

## 6. Concluding Remarks

The problem of disappearing substance and the problem of disappearing reality have something in common: they both arise from the same kind of reaction to empirical counterevidence. That reaction, essentially, amounts to diluting one’s metaphysical commitments and pretending that nothing is amiss. In the former case, the counterevidence comes from studies on belief; in the latter, from studies on taxonomic practices in science. As the philosophical accounts become ever weaker, the problems involved appear to become ever more tractable, which creates the illusion of progress. In both cases, also, the reaction can seem reasonable only if one chooses to ignore the significance of theory and approximate truth for realism. Let me close this paper with a sketchy, and somewhat speculative, reconstruction of how that occurred in the case of scientific realism and natural kinds.

On a very general way of looking at things, the realist might expect science to produce a true theory that would explain phenomena at a fundamental level. Such a theory would comprise fundamental laws, formulated in terms of fundamental concepts. Fundamental concepts, in such a case, would partition matter into objects with boundaries determined by the world. Fundamental kinds would be perfectly objective and explanatory.

Yet fundamental objects form a stunning variety of ensembles whose behavior defies explanation by fundamental theory, so we need non-fundamental theories to handle systems of non-fundamental objects. How do these theories divide the various groups of fundamental objects into non-arbitrary wholes? And what does it mean for a type of a complex object to be non-arbitrary, in the first place?

These are vexing quandaries. An answer that seems to underlie the appeal of contemporary accounts of natural kinds is that scientists distinguish between natural and conventional groupings by attending to the number of true generalizations the groupings afford—and, arguably, the number of true generalizations afforded by a kind is a purely objective criterion of naturalness.

To a rough approximation, if you imagine the class of all possible rival taxonomies, natural taxonomies will be those that support the largest number of true generalizations. That does not mean, however, that only natural taxonomies are useful or scientific. As we regard some generalizations to be more important than others, we may favor a taxonomy that maximizes the number of interesting generalizations, rather than true generalizations simpliciter. Although an interest-driven taxonomy may not coincide with a natural taxonomy, we may perhaps insist that interest-driven taxonomies are reasonably natural if they bear a close enough resemblance to a natural taxonomy.

Obviously, one can fill in this sketch in a number of ways, which I do not want to get into. As we saw, Khalidi (2013) refines the picture by recognizing that, rather than just counting generalizations, we need to assess them with respect to inductive richness, robustness, and strength of causal connection. Posłajko (2024) adds that reducibility to lower-level natural kinds might also be necessary, but he does not say why or what kind of explanatory link he has in mind.

Embellishments aside, I want to draw attention to a fundamental difficulty of this general approach: namely, that, on this picture, the problem of assessing rival taxonomies is intractable unless we already know all true generalizations. Which, of course, we do not.

We can only speculate about the proportion of true generalizations that have been entertained in the course of human history to true generalizations *tout court*. Likewise, we have no way of assessing the rate of well-confirmed generalizations in the set of all true generalizations. And, on a more practical side, if we mean business, we should partition well-confirmed generalizations into those that are likely approximately true and those that are not. While tractable in principle, even this task is daunting.

In effect, judgments of natural-kindness are massively underdetermined by the evidence and subject to bias. Nobody counts generalizations associated with rival taxonomies. Instead, authors seem to assume that if one can hang several significant generalizations on a kind then that kind must be reasonably natural: “If [...] a given category turns out to not form a natural kind, then we should not expect any significant generalizations to be discoverable about it” (Posłajko, 2024, p. 55).

That assumption, I submit, bears much of the responsibility for the illusion of tractability surrounding natural kinds. Yet the literature on actual taxonomic practices reveals that it is probably false. Generally, unscientific and non-natural taxonomies that have enjoyed acceptance can all boast inductive success (Ludwig, 2018). Only contrived philosophical examples do not. Indeed, apparent causal efficacy, analyzed in terms of generalizations, does not coincide with the approximate truth of underlying theory or the scientific character of the kinds in question.

Let me end with an illustration. Posłajko (2024, p. 93) cites choleric personality as an arbitrary (non-natural) kind. Yet that category was part of a highly successful theory that had unified physiology, medicine, and psychology and dominated medical science for well over a millennium (roughly, from the second to the seventeenth century). The theory was inductively successful, by the standards of the day, and the concept of a choleric personality offered a wealth of useful predictions (for an overview, see Bos, 2009). Indeed, for all we know, the category of a choleric personality may still be more projectible than, say, the kind argon (being a noble gas, argon is mostly chemically inert)—and that might be so even if contemporary chemistry is approximately true.

As far as I can tell, the only reason why choleric personality is no longer considered to be a natural kind is its association with the now-defunct humoral theory. But, alas, if natural-kind concepts can be real even if they feature in mistaken theories, then choleric personality might be as reasonable a natural kind as any.

## REFERENCES

- Baker, L. R. (1988). *Saving Belief: A Critique of Physicalism*. Princeton: PUP.
- Bishop, M. A., Stich, S. P. (1998). The Flight to Reference, or How Not to Make Progress in the Philosophy of Science. *Philosophy of Science*, 65(1), 33–49. <https://doi.org/10.1086/392625>
- Boghossian, P. A. (1990). The Status of Content. *The Philosophical Review*, 99(2), 157–184. <https://doi.org/10.2307/2185488>
- Bos, J. (2009). The Rise and Decline of Character: Humoral Psychology in Ancient and Early Modern Medical Theory. *History of the Human Sciences*, 22(3), 29–50. <https://doi.org/10.1177/0952695109104422>
- Boyd, R. (1989). What Realism Implies and What it Does Not. *Dialectica*, 43(1–2), 5–29. <https://doi.org/10.1111/j.1746-8361.1989.tb00928.x>
- Boyd, R. (1991). Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 61(1/2), 127–148.
- Boyd, R. (2010). Realism, Natural Kinds, and Philosophical Methods. In H. Beebe, N. Sabbarton-Leary (Eds.), *The Semantics and Metaphysics of Natural Kinds* (pp. 212–234). Routledge.

- Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science*. Cambridge University Press.
- Chakravartty, A. (2023). Last Chance Saloons for Natural Kind Realism. *American Philosophical Quarterly*, 60(1), 63–81. <https://doi.org/10.5406/21521123.60.1.06>
- Chang, H. (2011). The Philosophical Grammar of Scientific Practice. *International Studies in the Philosophy of Science*, 25(3), 205–221. <https://doi.org/10.1080/02698595.2011.605244>
- Chang, H. (2015). Reductionism and the Relation Between Chemistry and Physics. In A. Simões, J. Renn, T. Arabatzis, *Relocating the History of Science* (Vol. 312, pp. 193–209). Springer.
- Churchland, P. M. (1979). *Scientific Realism and the Plasticity of Mind*. CUP.
- Churchland, P. M. (1981). Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy*, 78(2), 67–90. <https://doi.org/10.2307/2025900>
- Churchland, P. M. (2013). *Plato's Camera: How the Physical Brain Captures the Landscape of Abstract Universals*. MIT Press.
- Craver, C. F. (2009). Mechanisms and Natural Kinds. *Philosophical Psychology*, 22(5), 575–594. <https://doi.org/10.1080/09515080903238930>
- Cummiskey, D. (1992). Reference Failure and Scientific Realism: A Response to the Meta-Induction. *The British Journal for the Philosophy of Science*, 43(1), 21–40. <https://doi.org/10.1093/bjps/43.1.21>
- Darden, L., Maull, N. (1977). Interfield Theories. *Philosophy of Science*, 44(1), 43–64. <https://doi.org/10.1086/288723>
- Dennett, D. C. (1991). Two Contrasts: Folk Craft vs Folk Science and Belief vs Opinion. In J. D. Greenwood (Ed.), *The Future of Folk Psychology: Intentionality and Cognitive Science* (pp. 135–148). Cambridge University Press.
- De Queiroz, K. (2007). Species Concepts and Species Delimitation. *Systematic Biology*, 56(6), 879–886. <https://doi.org/10.1080/10635150701701083>
- Dupré, J. (1981). Natural Kinds and Biological Taxa. *The Philosophical Review*. <https://doi.org/10.2307/2184373>
- Dupré, J. (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Harvard University Press.
- Feigl, H. (1958). The “Mental” and the “Physical”. *Minnesota Studies in the Philosophy of Science*, 2, 370–497.
- Feyerabend, P. (1963). Materialism and the Mind-Body Problem. *The Review of Metaphysics*, 17(1), 49–66.
- Fletcher, G. J. O. (1995). Two Uses of Folk Psychology: Implications for Psychological Science. *Philosophical Psychology*, 8(3), 221–238. <https://doi.org/10.1080/09515089508573155>
- Fodor, J. A. (1974). Special Sciences (Or: The Disunity of Science as a Working Hypothesis). *Synthese*, 28(2), 97–115.
- Fodor, J. A. (1975). *The Language of Thought*. Harvard University Press.

- Fodor, J. A. (1982). A Reply to Brian Loar's "Must Beliefs be Sentences?". *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1982(2), 644–653. <https://doi.org/10.1086/psaprocbienmeetp.1982.2.192449>
- Fodor, J. A. (1985). Fodor's Guide to Mental Representations. *Mind*, 94(373), 76–100.
- Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.
- Ghins, M. (2024). *Scientific Realism and Laws of Nature: A Metaphysics of Causal Powers* (Vol. 483). Springer International Publishing. <https://doi.org/10.1007/978-3-031-54227-5>
- Gordon, R. M. (1986). Folk Psychology as Simulation. *Mind & Language*, 1(2), 158–171. <https://doi.org/10.1111/j.1468-0017.1986.tb00324.x>
- Hacking, I. (1991). A Tradition of Natural Kinds. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 61(1/2), 109–126.
- Hardin, C. L., Rosenberg, A. (1982). In Defense of Convergent Realism. *Philosophy of Science*, 49(4), 604–615. <https://doi.org/10.1086/289080>
- Havstad, J. C. (2018). Messy Chemical Kinds. *The British Journal for the Philosophy of Science*, 69(3), 719–743. <https://doi.org/10.1093/bjps/axw040>
- Heil, J. (1991). Being Indiscrete. In J. D. Greenwood (Ed.), *The Future of Folk Psychology: Intentionality and Cognitive Science* (pp. 120–134). CUP.
- Hendry, R. F. (2006). Elements, Compounds, and Other Chemical Kinds. *Philosophy of Science*, 73(5), 864–875. <https://doi.org/10.1086/518745>
- Horgan, T., Woodward, J. (1993). Folk Psychology is Here to Stay. In J. D. Greenwood (Ed.), *Folk Psychology and the Philosophy of Mind* (pp. 149–175). Psychology Press.
- Jenson, J. C. (2016). The Belief Illusion. *The British Journal for the Philosophy of Science*, 67(4), 965–995. <https://doi.org/10.1093/bjps/axv006>
- Khalidi, M. A. (2013). *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*. Cambridge University Press.
- Kim, J. (1989). The Myth of Nonreductive Materialism. *Proceedings and Addresses of the American Philosophical Association*, 63(3), 31–47. <https://doi.org/10.2307/3130081>
- Kitcher, P. (1993). *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. Oxford University Press.
- Kripke, S. A. (1972). Naming and Necessity. In D. Davidson, G. Harman (Eds.), *Semantics of Natural Language* (pp. 253–355). Springer Netherlands. [https://doi.org/10.1007/978-94-010-2557-7\\_9](https://doi.org/10.1007/978-94-010-2557-7_9)
- Laudan, L. (1977). *Progress and Its Problems: Towards a Theory of Scientific Growth*. University of California Press.
- Leplin, J. (1997). *A Novel Defense of Scientific Realism*. Oxford University Press.
- Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61(4), 343–377.
- Loar, B. (1981). *Mind and Meaning*. Cambridge University Press.

- Loar, B. (1982). Must Beliefs be Sentences? *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1982(2), 626–643. <https://doi.org/10.1086/psaprocbienmeetp.1982.2.192448>
- Ludwig, D. (2018). Letting Go of “Natural Kind”: Toward a Multidimensional Framework of Nonarbitrary Classification. *Philosophy of Science*, 85(1), 31–52. <https://doi.org/10.1086/694835>
- Lycan, W. G. (1988). *Judgement and Justification*. Cambridge University Press.
- Nichols, S., Stich, S. P. (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. OUP. <https://doi.org/10.1093/0198236107.001.0001>
- Niiniluoto, I. (1999). *Critical Scientific Realism*. Oxford University Press.
- Park, S. (2022). *Embracing Scientific Realism* (Vol. 445). Springer International Publishing. <https://doi.org/10.1007/978-3-030-87813-9>
- Porot, N., Mandelbaum, E. (2021). The Science of Belief: A Progress Report. *WIREs Cognitive Science*, 12(2). <https://doi.org/10.1002/wcs.1539>
- Posłajko, K. (2024). *Unreal Beliefs: An Anti-Realist Approach in the Metaphysics of Mind*. Bloomsbury Publishing.
- Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. Routledge.
- Putnam, H. (1975). The Meaning of “Meaning”. In K. Gunderson (Ed.), *Language, Mind, and Knowledge: VII* (pp. 131–193). University of Minnesota Press.
- Putnam, H. (1978). *Meaning and the Moral Sciences*. Routledge & Kegan Paul.
- Quilty-Dunn, J., Mandelbaum, E. (2018). Against Dispositionalism: Belief in Cognitive Science. *Philosophical Studies*, 175(9), 2353–2372. <https://doi.org/10.1007/s11098-017-0962-x>
- Ramsey, W., Stich, S., Garon, J. (1991). Connectionism, Eliminativism, and the Future of Folk Psychology. In J. D. Greenwood (Ed.), *The Future of Folk Psychology: Intentionality and Cognitive Science* (pp. 93–119). CUP.
- Ratcliffe, M. (2007). *Rethinking Commonsense Psychology: A Critique of Folk Psychology, Theory of Mind and Simulation*. Palgrave Macmillan UK.
- Smart, J. J. C. (1959). Sensations and Brain Processes. *The Philosophical Review*, 68(2), 141–156.
- Smolensky, P. (1988). On the Proper Treatment of Connectionism. *Behavioral and Brain Sciences*, 11(1), 1–23. <https://doi.org/10.1017/S0140525X00052432>
- Sommer, J., Musolino, J., Hemmer, P. (2023). A Hobgoblin of Large Minds: Troubles With Consistency in Belief. *WIREs Cognitive Science*, 14(4). <https://doi.org/10.1002/wcs.1639>
- Stich, S. P. (1996). *Deconstructing the Mind*. Oxford University Press.
- Taylor, H., Vickers, P. (2017) Conceptual Fragmentation and the Rise of Eliminativism. *European Journal for Philosophy of Science*, 7, 17–40. <https://doi.org/10.1007/s13194-016-0136-2>
- Tobin, E. (2010). Crosscutting Natural Kinds and the Hierarchy Thesis. In *The Semantics and Metaphysics of Natural Kinds* (pp. 179–191). Routledge.
- Zachos, F. E. (2016). *Species Concepts in Biology*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-44966-1>



AMIR HOROWITZ \*

## POŚLAJKO ON BELIEF SIMILARITIES AND BELIEF CAUSATION<sup>1</sup>

**SUMMARY:** Poślajko (2024) suggests viewing beliefs (and possibly other mental attitudes) as holding a unique ontological status: they exist but are not real. This is Poślajko's belief non-realism. Poślajko also argues that the folk concept of belief is realistic in the sense in question (it is a concept of natural entities), and the conjunction of this claim and belief non-realism yields his belief anti-realism, namely the view that beliefs do not satisfy this folk concept. The purpose of the present paper is to challenge Poślajko's case for belief non-realism. In Section 2, I argue that since Poślajko's argument for the claim that mental kinds are not natural kinds because they fail to ground similarities presupposes that functional similarities cannot fulfil the similarity requirement for naturalness, this argument establishes the claim that beliefs are non-natural and hence non-real in a rather weak sense. In Section 3, I argue that Poślajko's claim that beliefs are not causally relevant states, as well as its entailment of the claim that beliefs are non-real, depends upon a terminological choice. In Section 4, I take issue with Poślajko's claim that the soundness of the success argument for the existence of beliefs requires, first, that we be successful at prediction and coordination because we use the theory of folk psychology, and second, that folk psychology provide an accurate description of the internal structure of the mind (requirements that are not met). However, I point out my own reasons for the view that the success argument fails to establish the existence or reality of beliefs qua psychofunctional content-bearing states, and further, that there are no such states, though, in my opinion, predictively and explanatorily successful ascriptions of propositional attitudes do expose aspects of the mind's architecture, namely its logico-syntactic structures.

---

\* The Open University of Israel, Department of History, Philosophy, and Judaic Studies. E-mail: amirho@openu.ac.il. ORCID: 0000-0003-1913-0553.

<sup>1</sup> I thank two anonymous referees for their helpful comments on an earlier version of this paper.

KEYWORDS: Krzysztof Posłajko, belief anti-realism, natural kinds, mental causation, folk psychology.

## 1. Introduction

In his book, *Unreal Beliefs: An Anti-Realist Approach in the Metaphysics of Mind*, Krzysztof Posłajko suggests viewing beliefs (and possibly other mental attitudes) as holding a unique ontological status: they exist but are not real. This might sound contradictory, but it is not, since in Posłajko's sense "real" refers to a specific nature of some existents, namely to naturalness in David Lewis's (1983) sense.<sup>2</sup> So, the ontological category of existence includes both real and unreal entities, and beliefs are among the latter. This is Posłajko's *belief non-realism*. Posłajko also argues that the folk concept of belief is realistic in the sense in question (it is a concept of natural entities), and the conjunction of this claim and belief non-realism yields his *belief anti-realism*, namely the view that beliefs—as they really are—do not satisfy this folk concept.<sup>3</sup> Thus, (at least) part of mental reality is different from what it appears to the folk to be. An additional thesis of the book, one which is not part of its anti-realism, is that the folk concept of belief should be reformed.

The purpose of the present paper is to challenge Posłajko's case for the view that beliefs are non-real. To be natural in the Lewisian sense, properties must meet the following three criteria: function in causal explanations, ground objective similarities—which Posłajko takes to amount (in the context of the issue of mental realism) to constituting natural kinds—and be connected to more natural properties. Posłajko argues that beliefs meet none. In part I, I will argue that since Posłajko's argument for the claim that mental kinds are not natural kinds—inasmuch as they fail to ground similarities—presupposes that functional similarities cannot fulfil the similarity requirement for naturalness, then this argument establishes the claim that beliefs are non-natural and hence non-real in a rather weak sense. In part II, I will argue that Posłajko's claim that beliefs are not causally relevant states, as well as its entailment of the claim that beliefs are non-real, depends upon a terminological choice. In part III, I will take issue with Posłajko's claim that the soundness of the success argument for the existence of beliefs requires, first, that we be successful at prediction and coordination *because* we use the theory of folk psychology, and second, that folk psychology

---

<sup>2</sup> The feel of tension between the claim that beliefs exist and the claim that they are not real may be further mitigated upon bearing in mind that the sense of "exist" under consideration is the minimal deflationist one. I will not discuss this issue.

<sup>3</sup> I refer to Posłajko's views as "belief anti-realism" and "belief non-realism", respectively, despite the fact that he does not employ these expressions (though he does refer to his views as "anti-realism" and "non-realism").

provide an accurate description of the internal structure of the mind (requirements that are not met). Still, for reasons different from Posłajko's, I share his views that the success argument fails to establish the existence or reality of beliefs *qua psychofunctional content-bearing states*, and further, that there are no such states. But I take folk psychology to be more faithful to mental reality than Posłajko does, for in my opinion predictively and explanatorily successful ascriptions of propositional attitudes do expose aspects of the mind's architecture, namely its logico-syntactic structures. I will briefly point out my reason for this view.

## 2. Natural Kinds and Similarities

In this part, I address Posłajko's arguments for the claim that mental kinds fail to ground objective similarities, and thus are not natural kinds. One of the arguments is based on the claim that, as functional, beliefs do not possess hidden essences. I argue that since this argument in fact presupposes that functional similarities cannot fulfil the similarity requirement for naturalness, then it establishes the claim that beliefs are non-natural and hence non-real in a rather weak sense, one which does not make room for multiple realizability of mental categories. In his other argument, Posłajko correctly points out that our standard folk content ascriptions cannot track non-trivial similarities, but, I argue, fine-grained, fully detailed, belief ascriptions do track non-trivial similarities.

Posłajko mentions the traditional model of natural kinds, proposed by Putnam (1975b) and Kripke (1980), according to which natural kinds are characterized by the fact that their members share hidden essences, e.g., atomic structure or DNA. Posłajko agrees with the view (of, *inter alia*, Beebe, Sabbarton-Leary, 2010; Cooper, 2013) that

this model cannot be extended to psychological kinds as there is little sense in claiming that they might be seen as possessing hidden micro-essences. If psychological kinds are defined functionally, then we cannot postulate a hidden structure common to all instances of such a kind. (Posłajko, 2024, p. 92)<sup>4</sup>

So, it is in virtue of the conceptual identification of the psychological with the functional that makes psychological kinds such as beliefs non-natural kinds on this model. Posłajko appears to endorse the analytic version of functionalism (mainly due to Lewis, 1972), according to which mental terms are defined by the folk-psychological generalizations about the relations among mental states, their environmental causes, and their behavioral effects.<sup>5</sup> Note that psychological

---

<sup>4</sup> All further references to Posłajko's work are to his (2024).

<sup>5</sup> Specifically, Posłajko's argument presupposes (the standard) role functionalism, rather than realizer functionalism (proposed by Lewis, 1980), which identifies mental states not with functional roles but with their particular physical realizers, and takes functional specifications merely to refer to whichever lower-level properties satisfy them.

kinds' being non-natural in the sense in question means neither that mental states do not have micro-essences—if they are physical, then they do—nor that psychological concepts such as “belief” fail to capture similarities among their referents—*ex hypothesi*, they capture functional similarities (both between type-different beliefs and between different tokens of the same belief-types). But according to Pośłajko, those micro-essences do not make kinds psychological, and those functional similarities do not make kinds natural.

Pośłajko does not dwell much on the hidden-essence conception of natural kinds, which he takes to be unsuitable for psychological kinds, and turns to discuss the more liberal conception of Khalidi (2013) and Cooper (2013), according to which what distinguishes between psychological categories which are arbitrary and those which are objective is the latter's projectability. He considers this conception to be more suitable for psychology. Still, the claim that the category of beliefs is not natural due to its functional nature is significant to the issue of belief realism. Since the issue of the naturalness (or lack thereof) of psychological categories such as beliefs is, for Pośłajko, a route to the issue of their reality (or lack thereof), it appears that he takes their functionality to be at least part of what prevents them from being real. Furthermore, note the following three claims: first, their functionality prevents beliefs from being natural and hence real, according to Pośłajko, because, as functional, they do not possess hidden essences; second, functional categories can be realized physically; third, Pośłajko shuns away from dualism. The conjunction of these claims implies that the reason for the non-naturalness of beliefs that is pointed out in this context must be the fact that functions are not related to any specific physical realization. In other words, the reason is the (in principle) multiple realizability of functions; in still other words, it is the fact that functional categories are not physical categories.

However, to the extent that the non-naturalness and non-reality of beliefs depend on this characteristic, the sense in which beliefs are non-natural and non-real is rather weak. For in this sense, properties that figure in the special sciences (e.g., biological properties) and software properties, for example, would come out non-real. That is, on this conception, beliefs would come out as real (or non-real) as those properties. Pośłajko would certainly not be satisfied with such a mild belief non-realism, and indeed he asserts his denial of the reality of beliefs to be stronger than what mere multiple realizability implies. Acknowledging the plausibility of the idea that higher-level states such as beliefs are multiply realizable, he notes that this characteristic “might be (and probably would be) accepted by the strong realist about beliefs” (p. 67). And, as mentioned, in spite of his claim that beliefs are not natural kinds on the hidden-essence conception of natural kinds, he does not attach significance to this conception. But it is worth raising the issue of multiple-realizability and functionality not only because Pośłajko does claim that the category of beliefs is not natural due to its functional nature and so should confront the challenge that it leads to a rather weak belief non-realism, but also because of the relevance of this issue to another essential aspect of beliefs as natural kinds, namely that of whether the concept “belief” captures similarities.

Let's move to Poślajko's argument for a negative answer to this question. It is one tenet of Poślajko's belief non-realism that "There are no objective similarities between states which are described by attributions of the form '*x* believes that *p*'" (p. 64). Poślajko asks us to consider two people who believe that JFK went to Harvard. The attribution of this belief to both is justified on the grounds that both meet some common-sense criteria for attributing it, e.g., assert "John F. Kennedy went to Harvard", or exhibit some behavioral dispositions that are related to it. He admits that there is something that makes each attribution of such a belief true, when it is: "this might be some combination of behavioral and verbal dispositions, sub-personal cognitive states, and phenomenal feels" (p. 65). However, he denies that "there is any deep similarity between the factors that make such attributions true in different cases" (p. 66). To show this, he asks us to further imagine that the two people to whom the belief that JFK went to Harvard is attributed differ in ways that are relevant to their possession of this belief:

The one is an academic who studies American presidents [...] the other is a distracted student who is utterly disinterested in the life of past politicians but has read some stuff about Kennedy and somehow remembered the fact that he went to Harvard. The dispositions of these two persons—both behavioral and verbal—would most probably be diverged in this case, and we might suppose that the more basic sub-personal cognitive states would also be very different [...] We have also no reason—at least according to a minimal non-realist—to suppose that the phenomenal aspect of belief would be the same (or reasonably similar) in these two cases. (Poślajko, 2024, p. 66)

I agree with much of what is claimed here, though not with all. Indeed, Poślajko illustrates that the individuation of beliefs—as we standardly ascribe them—is wild and undisciplined. Stich (1983) presented examples that illustrate this phenomenon, and so did Crane (2017) and Curry (2021; whom Poślajko mentions in this context). We may also add to the list the example of people who are attributed the same belief despite speaking and thinking in different languages—they certainly differ with respect to some relevant sub-personal states.<sup>6</sup> Note that the point of all these examples concerns the ascription of content. The examples do not concern the psychological mode of believing—it is content whose ascription is supposed to be wild and undisciplined. And I agree that it is. To be precise, I agree that our standard folk content ascriptions—or at any rate part of them—are wild and undisciplined to the extent that they cannot be said to track non-trivial similarities between the supposed bearers of content (e.g., belief-states).

But a defender of the reality of beliefs might protest that this falls short of undermining her view (or at any rate a version of it that is worth considering).

---

<sup>6</sup> They also differ with respect to some verbal dispositions, but this is true only under a non-intentional individuation of these dispositions and arguably only the intentional individuation matters.

One possible way for her to base this protest is to argue that there is a gap between beliefs as they are and belief as they are ascribed by us. Brian Loar (1988) distinguishes between “social contents”—the contents we ascribe to people by means of *de dicto* or oblique ascriptions, and “psychological contents”—real contents of people’s mental states, which supervenes on conceptual role, that is, on the state’s relations with other mental states, perceptual input, and behavioral output.<sup>7</sup> As applied to the JFK belief example, the belief realist may argue that the identical content ascriptions in question—those which ascribe to both subjects the belief that JFK went to Harvard—do not accurately describe the real contents of the subjects, which differ, as the functional roles associated with their respective beliefs differ. If so, then this example does not put to test the thesis that type-identical beliefs enjoy similarities, since the beliefs under consideration are not type-identical.

Trivially, the distinction between contents that are ascribed to people and contents they really have presupposes intentional realism, which Pośłajko appears to reject, but such a presupposition is legitimate as a challenge to belief non-realism. That is, it is upon Pośłajko’s shoulders to undermine this distinction. One possible way for him to undermine Loar’s distinction is to argue against the idea that mental content can be reduced to psychological conceptual role (for my arguments against conceptual role theories of content, see Horowitz, 1992; 2024), but the very idea of this distinction can incorporate any other realist approach to content, and assuming that all realist approaches to content are doomed to fail does not appear to be a legitimate move on the non-realist part in this debate.

However, there is a deeper problem with approaches such as Loar’s. The distinction in question is committed to the claim that the whole practice of content ascription may be misguided, and this claim makes little sense in an intentional realist picture. At least, it is strongly unmotivated. For if the whole practice of content ascription is, in principle, wrong, what reasons could we have for identifying any mental property or relation that is not captured by content ascriptions with content? How at all could we identify content in complete separation from our content ascriptions? To assume that there is such a property or relation appears utterly arbitrary. So, Pośłajko’s JFK argument is not challenged by any approach that is based on a distinction between real and psychological content.

Still, there is a less radical alternative that challenges Pośłajko’s argument to the effect that beliefs do not ground similarities. On this alternative, on many occasions content ascriptions yield only partial descriptions of our mental contents. In other words, they are often elliptical, and in still other words, they standardly individuate mental contents in coarse-grained ways while it is possible (though for most practical purposes not necessary) to individuate many of them in more fine-grained ways. Thus, the academic and the distracted student

---

<sup>7</sup> Andrew Woodfield argues for a similar view (1982; for a conceptual role account of mental content, see Block, 1986).

have different beliefs with respect to JFK, even though the same content ascription is standardly used to describe both beliefs. For example, we may say that Sam (the academic) believes that JFK, who was the president of the USA from 1961 until his assassination in 1963, and whose main achievements as a president were [...] excelled in his studies at Harvard, and that Tom (the distracted student) believes that the American president named JFK went to Harvard. Now, you might object that some of the details that figure in this ascription of Sam's belief are not part of the content of his belief, and that, generally, all content ascriptions resist analysis of this kind. I think there are reasons to doubt this claim, but suppose it is true, and thus that we cannot truly describe the content of Sam's belief in a way that is more fine-grained than the just-mentioned description of Tom's belief. If so, then there is no reason to say that their respective beliefs differ in behavioral dispositions, for example. (Of course, they may say different things about JFK on various occasions because they differ with respect to some other JFK beliefs, but what matters is the contribution of the beliefs under consideration to the behavioral dispositions under the same circumstances, that is, how these beliefs would have been interacted with identical beliefs or desires). So, if a fine-grained individuation of these two people's relevant JFK beliefs is feasible, then these beliefs differ in type and so this example does not put to test the idea that beliefs ground similarities. If, on the other hand, we justly ascribe to these people the same JFK belief-type and no fine-grained individuation is feasible, despite their different relevant backgrounds and the like, then they share behavioral dispositions.

Now, behavioral dispositions, like dispositions of any kind, do not arise *ex nihilo*, and, most plausibly, identical behavioral dispositions are underlain by factors of the same type. Indeed, these factors need not be of the same physical type, and assuming that beliefs are in fact multiply realized, then they *are not* of the same physical type. But factors that underlie the same behavioral dispositions are, most probably, structurally similar and functionally similar (see Putnam, 1975a argument for the claim that functionalism is the most plausible theory of mental state types). So, it appears that fine-grained, fully detailed, belief ascriptions do track similarities and beliefs do ground similarities at some level of reality. The insistence that the fact that beliefs ground merely high-level similarities supports the case for belief non-realism (no other characteristic of structural and functional similarities may be relevant to their exclusion), presupposes that only type-physicalism is a (physicalist) realist account of mental kinds such as beliefs.

### 3. Are Beliefs Causes?

The reasoning that I have just presented for the obtaining of functional similarities among beliefs that are treated as identical is concerned with the causal basis of the manifestations of the dispositions associated with beliefs (though functional properties are not, strictly speaking, causal properties). So, it is related to the issue of mental causation and, specifically, belief causation. Poślajko at-

tacks this issue directly on several fronts, because he takes the question of belief realism to hinge upon the question of whether we can accept the idea of beliefs as (higher-level) causes. His reply to the latter question—and hence to the former—is negative. I will now discuss his argument to this effect, and argue that the truth of his answer turns out to be a matter of mere terminological choice.

Central to Pośłajko's argument is the idea that "the [supervenience] base of beliefs is wider than the occurrent psychofunctional state of a believer", and therefore "the difference-[making] account of causation does not vindicate the idea that beliefs are causes" (p. 101). According to the difference-making account, to say that property *A* causes property *B* means that making a difference in the value of property *A* would result in a change of value in property *B* (e.g., List, Menzies 2017; Menzies, 2008; Woodward, 2008). Pośłajko assumes, for the sake of argument, that this account resolves the problem of higher-order causation; his claim is that, even then, the problem of the causal relevance of content persists. The claim that the supervenience base of contentful mental states such as beliefs is wider than the occurrent internal mental state is usually understood as the semantic externalist view (mainly associated with Burge, 1979; Putnam, 1975b), according to which factors external to the possessor of a mental state—be these environmental-physical ones or social ones—participate in determining its content. According to Pośłajko, however, semantic internalism is also committed to that claim, for it is committed to the view that whatever constitutes the content of a belief is *temporally* external to the occurrent mental state of the believer. Whether the internal factor that constitutes mental content is physical or phenomenal, it must be located in the past of the believer. So,

the supervenience base of my current belief contains at least two states: the first is the past event that is partly constitutive of my concept, which is part of my belief; the second is my current psychofunctional state related to this belief. (Pośłajko, 2024, p. 103).

Let's refer to this view as "diachronic internalism".<sup>8</sup>

Pośłajko argues (pp. 103–104) that this view entails, in an interventionist framework,<sup>9</sup> that beliefs are not causally active. The argument is this: Suppose that, due to being presented by fake scientific evidence, I drop belief *B*<sub>1</sub> and adopt, instead, belief *B*<sub>2</sub>, and as a result perform action *A*. This appears to be a case of an interventionist mental causation. However, suppose that a doppelgänger of mine in a nearby possible world has the same psychofunctional state as the one that is associated with my *B*<sub>1</sub>. After being exposed to the same interven-

---

<sup>8</sup> As noted, according to the standard understanding of externalism, such a view should be classified as externalist, but my classification of it as a unique internalist view aligns with Pośłajko's terminology, which reflects his claim that not only under (standard) externalism do beliefs lack causal efficacy.

<sup>9</sup> Interventionism is Woodward's (2008) version of the difference-making account. The uniqueness of this version does not matter to the present discussion.

tion he comes to have the same psychofunctional state as the one that is associated with my *B2*, and as a result he comes to perform action *A*. Due to the different ways in which we acquired our concepts, his beliefs are *B1\** and *B2\**, respectively, that is, his beliefs differ in content from my corresponding beliefs. Now, the intervention in question (the presentation of the fake information) cannot change those past events that constitute the contents of our respective beliefs. It can only change the psychofunctional state in question. Since this state is part of the belief, a change in it is a change in the belief. And so, since a change in it brings about action,<sup>10</sup> this case involves counterfactual dependence between belief and action. However, this dependence is not causal, because it is the change in the psychofunctional state that brings about the action. Thus, in an interventionist framework, beliefs are not causes—they are not causally responsible for changes in behavior.

As Poślajko notes, this argument is a variant of the arguments to the effect that wide content—content that depends on factors external to the subject—is causally inert. It is worth noting that these arguments have been challenged, notably by Tyler Burge (1986; 1989),<sup>11</sup> and that Poślajko's argument for the claim that beliefs are not causes pertains to a temporal element and does not apply to externalism. Perhaps Poślajko is assuming that those arguments can survive the challenge. Otherwise, for his argument to be effective, it must be that content is partly determined by a past factor regardless of whether it is partly determined by an external factor. While it appears that Poślajko only argues that diachronic internalism is entailed by internalism, it might be that he intends to argue for the categorical claim that content is always partly determined by a past internal factor. At any rate, I will now discuss his claim that beliefs are not causes on the assumption that this categorical claim is true.

First, we should distinguish between the causal status that can be attributed to properties and the causal status that can be attributed to events or states. Events are the paradigmatic causes. It is only in a loose sense that we can say that properties are causes. *Instantiations* of properties (which, according to many philosophers, are events) are causes, whereas properties can, at most, be causally efficacious, in the sense that events have their causal power *by virtue of* possessing certain properties. A notion that is weaker than that of the causal efficacy of properties is that of properties' causal relevance. According to Jackson and Petit, for example, what makes a non-efficacious property causally relevant to the causal production of an event is the fact that its realization ensures that a crucial

---

<sup>10</sup> The claim expressed by this clause is not stated by Poślajko. It is my understanding that his argument leans on it.

<sup>11</sup> Burge's interlocutor on this issue, Jerry Fodor, was later convinced that Burge's view is closer to truth than his own (Fodor, 1994).

productive property is realized and, in the circumstances, that the event occurs (e.g., Jackson, Petit, 1990).<sup>12</sup>

Now, how should we understand Pośłajko's claim that beliefs are not causes? He does not deny that the internal psychofunctional states that are associated with beliefs are causes of behavior. For him, this does not suffice to endow *beliefs* with the status of causes. So, he must be rejecting the identification of those psychofunctional states with beliefs. Indeed, they are not the supervenience base of beliefs, but only parts thereof. Further, "What turns out to be genuinely causally relevant in explanations of actions, even in the interventionist framework, are not beliefs *qua* content-bearing states but psychofunctional occurrent states" (p. 105, emphasis added).<sup>13</sup> This claim is also concerned with the issue of the causal relevance of properties or categories. According to it, the category of beliefs is causally irrelevant: those properties that make entities beliefs and the beliefs that they are (according to their individuation by our practice of belief ascription) are not causally relevant.<sup>14</sup> Indeed, assuming either externalism or diachronic internalism, this appears to be true. However, Pośłajko is also making here the stronger claim that the psychofunctional states in question—which are the causes of human action—are not content-bearing states. So, they are not beliefs, and beliefs are not causes of human action. We should note that the claim that content-bearing states are not causes is different and does not follow from the claim that content, which makes beliefs beliefs and the specific beliefs they are, is not causally relevant.

Here is another way to look at the matter. We can take beliefs to be those internal psychofunctional states as individuated in terms of content, and hence as individuated in terms of factors external to themselves, while these external factors can be environmental-physical, social, or (in the framework of Pośłajko's diachronic internalism) past ones. On this suggestion, what beliefs are is one thing—they are *internal* states; how they are individuated is another—they are *externally* individuated. If this suggestion is true, then neither semantic externalism nor diachronic internalism undermines the idea of beliefs as internal psychofunctional states and thus the idea of beliefs as causes of human actions.

Thus, on both ways of viewing the matter, first, human actions are caused by internal psychofunctional states; second, beliefs are individuated externally, i.e.,

---

<sup>12</sup> The difference-making account can also be considered an account of causal relevance, although it is often presented—and Pośłajko presents it, as we saw—in terms of being a cause rather than in terms of being causally relevant.

<sup>13</sup> Pośłajko's conditioning this claim upon interventionism should not be taken lightly. He claims that beliefs *qua* beliefs are causally irrelevant *even* in the interventionist framework, but interventionism is not innocuous. For example, it clashes with Burge's abovementioned view of mental causation, a view which allows beliefs to play a causal role in the framework of semantic externalism.

<sup>14</sup> The claim that beliefs *qua* content-bearing states are not causes is reminiscent of Stich's (1978) claim that there are no entities that are both semantically evaluable and causes of behavior.

by factors external to the current psychofunctional states in question; third, the properties that individuate beliefs are not causally relevant. On one of these ways, these properties (externally) individuate *those internal states* and make them beliefs and the specific beliefs they are, and so it is beliefs that are causes of action; on the other way (Pośłajko's), since those properties that determine the identity of certain mental entities as beliefs and as the specific beliefs they are (be their ontological status what it may) are external to those internal states, *those states are not beliefs*—they are not content-bearing states.

How shall we decide whether those internal psychofunctional states are *beliefs qua content-bearing states* or not? According to Pośłajko, "... the problem is that beliefs lose their status as genuinely causally relevant factors for sub-personal psychological states which are individuated purely psychofunctionally" (p. 105). But why say that these states are individuated purely psychofunctionally, rather than that they are also externally individuated *qua* content-bearing states? No criterion can decide on the matter except for stipulation. The difference between these views is a difference that makes no difference: no possible metaphysics—no possible concealed facts—can settle the issue. It is a matter of nothing but terminological choice. This point is significant since if this issue is a matter of sheer stipulation, then, to the extent that it is supposed to contribute to the case for the non-naturalness of beliefs, and in turn to the case for their non-reality, then the issue of the reality of beliefs itself becomes a matter of sheer stipulation. In fact, taking Lewisian naturalness to be the mark of the real is itself a stipulation. Pośłajko argues that

The issue of realism about beliefs and the question of whether belief properties are natural seem to strongly coincide: in both cases, the positive answer is partly justified by accepting the claim that beliefs play a genuine causal-explanatory role [...]. The indicators are similar in both cases. (2024, pp. 58–59)

However, if, as this passage appears to suggest, there is an independent test for the reality of beliefs—namely the playing of a genuine causal-explanatory role—why is the appeal to naturalness needed?<sup>15</sup>

Of course, I do not claim that the existence of beliefs is a matter of stipulation. My claim regarding stipulation in this context is only concerned with the issue of the "reality" of beliefs as addressed by Pośłajko. Additionally, this claim does not diminish the significance of Pośłajko's observation about the uniqueness of beliefs and their role—if any—in mental causation. However, I think that this uniqueness does not challenge their reality and provides no reason to assign them

---

<sup>15</sup> Pośłajko may argue that his terminological choice is preferable on the grounds that the folk psychological concept of belief is a concept of a mental cause of behavior *qua* content-bearing state. However, first, I doubt that this folk psychological concept is so specific and fine-grained. (I am indebted to an anonymous referee for this point). Second, the current issue concerns the nature of mental states themselves, rather than their match with folk psychology.

an inferior ontological category within the realm of existence, unless it is thus stipulated.<sup>16, 17</sup>

That said, I believe that the idea of the separation between the internal psychofunctional states and intentional properties, which figures prominently in Poślajko's discussion, is very significant. Furthermore, I believe that no concrete entity possesses intentional properties. I will refer to this issue in the next section.

#### 4. The Success Argument

Beliefs, alongside other propositional attitudes, notably desires, are mentioned in successful explanations and predictions of behavior. To the extent that these explanations and predictions are causal, this fact might be taken to challenge Poślajko's view that beliefs are non-real. The success argument for realism about propositional attitudes *qua* content-bearing states (Fodor, 1987; Lahav, 1992; Pylyshyn, 1980) in fact challenges views such as Poślajko's along such lines. I will now present Poślajko's treatment of this argument and criticize it. Finally, I will briefly show why I share his view that the success argument fails to establish the existence or reality of beliefs *qua* psychofunctional content-bearing states, and, relatedly, why I take folk psychology to be more faithful to mental reality than Poślajko does.

Since the success argument purports to establish that content-bearing states exist and are causes of behavior, it should be taken to pertain not merely to their existence but also to their reality in Poślajko's sense. Indeed, this is how Poślajko (p. 148) presents the argument (restricting it to beliefs):

---

<sup>16</sup> There are philosophers who endorse "Alexander's dictum", according to which to be real is to have causal powers. But even if we understand this principle as a requirement for having some superior ontological status within the set of existents rather than for membership in this set, it makes no sense to take it to distinguish in such terms between those two abovementioned ways of viewing the matter, and the claim regarding stipulation remains intact.

<sup>17</sup> Recall that Poślajko's belief anti-realism is not the claim that beliefs are non-real, but the claim that they do not match the folk concept of belief. In particular, Poślajko argues that the folk concept of belief is a concept of a causally active entity, whereas beliefs are causally irrelevant. Let me just briefly remark that I am not sure that the folk concept of belief is indeed a concept of a cause. I tend to think that although people would agree that beliefs play a causal role vis-à-vis behavior, the concept of belief itself is silent on the issue of belief causation. If, for example, we are reported that science has discovered the causal mechanism responsible for behavior and that beliefs are not part of it, it seems that those of us who will be persuaded that this report is correct, will not take it to undermine the existence of beliefs. On the other hand, it seems to be a crucial part of the folk notion of belief that beliefs have contents—that they are beliefs *that...*—and a psychological mode (that of believing).

1. Folk psychology is extremely predictively successful.
2. Beliefs are posits of folk psychology.
3. Posits of extremely predictively successful theories are real.

Conclusion:

Beliefs are real.

If the argument is sound, it undermines Poślajko's contention (p. 112) that there is no empirical support for belief realism, for the first premise is supposed to be supported empirically. It is this premise that Poślajko attacks. He accepts that people do remarkably well at predicting the behavior of their fellow-subjects, but argues that for establishing the claim that folk psychology is a successful theory "one must show that we are successful at prediction and coordination *because* we use the theory of folk psychology [...]" (p. 148, emphasis added). According to Poślajko, this condition does not obtain. Leaning on Dennett's (1998) distinction between folk psychology understood as a craft and understood as ideology, he argues: "Fodor's idea that we predict and coordinate behaviour because we rely on the theoretical apparatus of folk psychology is just one of many possible interpretations of the success of our folk-psychological craft" (p. 149).

I think the focus on the theory of folk psychology on the part of Fodor and, consequently, of Poślajko leads the discussion astray. The issue concerns the reality of beliefs, while the appeal to folk psychology is supposed to be merely a means to vindicate their reality. As Poślajko agrees, we successfully predict behavior by ascribing beliefs. The issue is whether we can conclude from this that beliefs are real, and this question comes down to whether the appeal to that which makes those states beliefs is essential to this success. This issue need not refer to folk psychology as a theory. The only relevance of the theory of folk psychology to the success argument is that it is the enterprise that posits beliefs. This is the only theoretical apparatus to which we should appeal. In fact, we can dispense with reference to folk psychology and present the argument (as it is indeed often presented) as one that simply leads from the predictive and explanatory success of ascribing beliefs to the reality of beliefs. If to ascribe beliefs is, by definition, to use the theory of folk psychology, then trivially the predictions succeed because we use the theory, but in such a case nothing needs to be shown beyond the appeal to beliefs. And if the use of this theory is something beyond and above the ascription of beliefs, it is not required for the predictions to work. In either case, it is the appeal to beliefs that is responsible for the success. Indeed, the simple and natural explanation of the success in question is the assumption that we have beliefs and desires, but whether this is true—whether this responsibility means that beliefs *qua* content-bearing states underlie this success—is another matter, to which I shall turn shortly.

Poślajko adds another requirement to the soundness of the success argument: that folk psychology provide an accurate description of the internal structure of

the mind. Following Hochstein (2012), he doubts that this is the case. The validity of this requirement depends on what “the internal structure of the mind” is taken to mean. Certainly, to ascribe beliefs in terms of content is not to ascribe the physical causal mechanisms of the mind (at least in the *de dicto* sense of this claim). It is of the very essence of folk psychology to bypass such mechanisms, without, of course, denying their existence and reality. (If the presumed fact that beliefs cannot be identified with such mechanisms is taken to show that beliefs are not real, then, it seems to me, we are again in the realm of stipulation). But to bypass them is not to deny their existence or reality. Further, predictively successful ascriptions of beliefs may expose some aspects of those mechanisms; it is the gist of the success argument that those mechanisms possess causally relevant aspects that are ascribed by such ascriptions. That is, postulating some aspects of them may account for their predictive success. I think it does, though I share Pośłajko’s suspicion about the real import of the success argument. Specifically, I do not take it to establish the existence or reality of beliefs *qua content-bearing states*, which is what it is usually taken to establish. The argument fails to show that the causes of behavior possess intentional aspects.

I discuss this issue extensively in my *Intentionality Deconstructed: An Anti-Realist Theory* (2024), the main thesis of which is intentional anti-realism, namely the claim that no concrete entity possesses or can possess any intentional property. I will now present my reason for believing that the success argument does not vindicate the existence of intentional properties, and then my reason for believing that it does vindicate the existence of other aspects of the structure of the mind. I contest the argument’s third premise. I share the view of selective scientific realists such as Philip Kitcher (1993) and Stathis Psillos (1999), according to which the explanatory and predictive success of a theory only indicates the truth of those parts of it that are explanatory or active in the theory’s explanations and predictions. The question for folk psychology, therefore, is whether it makes room for a distinction between such parts or aspects and others. And it does, for in ascribing a propositional attitude to one, we thereby ascribe to one not only a psychological mode and content but also a logico-syntactic structure (the logico-syntactic structure of the ascribed propositional content). Postulating such structures suffices to account for the success under consideration, since they can stand in logical relations of all sorts and thus can guarantee both theoretical reasoning and (what is directly relevant to the success argument) practical reasoning. If this is correct, then the success argument fails to vindicate the existence of mental states with intentional properties, for the minimal postulation of the existence of mental states with logico-syntactic properties suffices to explain the success.

In fact, the success argument is a compelling argument to the effect that the explanatory and predictive success of folk psychology justifies the idea that we have internal psychofunctional states. Pośłajko accepts the existence and reality of such states, which are causes of behavior and which he regards to be parts of the supervenience base of beliefs (and probably of other propositional attitudes).

I assume he would agree that this is also established by the success argument. In my view, it is not only that the possibility of these states' having logico-syntactic properties blocks the success argument as an argument for intentional realism, as their postulation suffices to account for the explanatory and predictive success; I take this argument a step further. I take its rationale—that the explanatory and predictive success of a theory indicates the truth of those parts of it that are explanatory or active in the theory's explanations and predictions—to also support the view that these states in fact have logico-syntactic properties.

Indeed, low-level physical properties of these states are causally efficacious vis-à-vis behavior and they can account for the explanatory and predictive success of ascriptions of propositional attitudes in particular cases, but we cannot be satisfied with postulating such properties since content ascriptions often succeed to explain or predict instances of behavior as belonging to behavioral patterns that are not physical patterns (even though their instances are physical); recall the well-known examples such as those of stopping a cab or buying a share in the stock market. Postulating either intentional or logico-syntactic properties—both of which are conceptually implied by folk psychological explanations and predictions—would account for this phenomenon,<sup>18</sup> but as the latter are minimal, we are warranted to postulate them but not the former.

In this respect, Poślajko's approach and mine differ in the ways they conceive of the truth of folk psychology, for I take those internal psychofunctional states that underlie human behavior to have a closer affinity to the folk explanations and predictions of behavior: the former, though lacking intentional properties, often possess those structural characteristics that the latter take them to possess. Poślajko's belief non-realism takes folk psychology to be less faithful to mental reality than I do, and he denies, whereas, as noted, I accept, that predictively and explanatorily successful ascriptions of propositional attitudes reveal aspects of the mind's architecture. Folk psychology consists of ascribing *contents*, but contents necessarily involve logico-syntactic structures and thus content ascriptions are ascriptions of logico-syntactic structures (they conceptually imply logical structures, which must take some syntactic form or another). So, folk psychology's commitment to the existence of such structures (though arguably indirect) is as strong as its commitment to contents. This commitment is a commitment to language-like structures, but it is not a commitment to how such structures are realized.<sup>19</sup> The issue of the realization of the structures in question warrants wider discussion, but here I am only concerned with comparing Poślajko's view

---

<sup>18</sup> The claim that postulating either intentional or logico-syntactic properties would account for the fact that content ascriptions often succeed to explain or predict instances of behavior as belonging to behavioral patterns that are not physical patterns does not depend on the properties in question being causally efficacious. It suffices that they be causally relevant (in Jackson and Petit's sense mentioned above), and we can see that they are.

<sup>19</sup> For example, it clashes with radical connectionism (e.g., Smolensky, 1988), but it is compatible with implementational connectionism (e.g., Marcus, 2001).

and mine in terms of their fit with the core commitments of folk psychology, and as said, my view credits folk psychology with more faithfulness to mental reality than does Poślajko's.

But as we saw, both Poślajko and I reject the notion that the success argument—or any other argument—provides a good reason to postulate internal psychofunctional content-bearing states. I also share Poślajko's view that there are no such states.<sup>20</sup>

## 5. Summary

Poślajko suggests viewing beliefs as holding a unique ontological status: they exist but are not real. I tried to challenge three of his arguments to the effect that beliefs are non-real: the argument that beliefs do not ground similarities; the argument that leans on his diachronic internalism to show that beliefs are causally irrelevant, and the argument against the first premise of the success argument for belief realism, namely that folk psychology is predictively successful. I argued, respectively, that since Poślajko has not ruled out the option that beliefs ground functional similarities, then his argument makes beliefs non-real in a rather weak sense; that his claim against the causal relevance of beliefs—and thus his claim against belief non-realism—turns out to be stipulative; and that the relevant understanding of that premise of the success argument is that it is the appeal to beliefs that is responsible for the predictive and explanatory success of ascribing beliefs, and that thus understood, it is true. But I also argued that this still fails to establish that beliefs *qua content-bearing states* are real (a point on which I agree with Poślajko), since we cannot rule out the possibility that logico-syntactic properties of our psychofunctional states underlie the predictive and explanatory success of folk psychology. Further, it seems that they do, and so I take folk psychology to be more faithful to mental reality than does Poślajko, even though I share his view that this reality does not include psychofunctional content-bearing states.

---

<sup>20</sup> I could only address here several aspects of Poślajko's approach to intentionality, which he discusses widely. Likewise, a discussion and defense of my intentional anti-realism (presented in *Intentionality Deconstructed*; Horowitz, 2024) lies beyond the scope of the present paper, for it does not play a role in my critique of Poślajko's arguments and in my approach to the success argument. Here I only explained my reasons for believing that this argument fails to vindicate the existence of mental intentional properties while it does vindicate the existence of mental logico-syntactic ones.

## REFERENCES

- Beebe, H., Sabbarton-Leary, N. (2010). Are Psychiatric Kinds Real? *European Journal of Analytic Philosophy*, 6(1), 11–27.
- Block, N. (1986). Advertisements for a Semantics for Psychology. *Midwest Studies in Philosophy*, 10(1), 615–678.
- Burge, T. (1979). Individualism and the Mental. *Midwest Studies in Philosophy*, 4(1), 73–121.
- Burge, T. (1986). Individualism and Psychology. *Philosophical Review*, 95(1), 3–45.
- Burge, T. (1989). Individuation and Causation in Psychology. *Pacific Philosophical Quarterly*, 70(4), 303–322.
- Cooper, R. (2013). Natural Kinds. In K. W. M. Fulford, M. Davies, R. Gipps, G. Graham, J. Sadler, G. Stanghellini, T. Thornton (Eds.), *Oxford Handbook of Philosophy and Psychiatry* (pp. 950–965). Oxford: OUP.
- Crane, T. (2017). The Unity of Unconsciousness. *Proceedings of the Aristotelian Society*, 117(1), 1–21.
- Curry, D. S. (2021). How Beliefs Are Like Colors. *Synthese*, 199, 7889–918. <https://doi.org/10.1007/s11229-021-03144-1>
- Dennett, D. C. (1998). Two Contrasts: Folk Craft Versus Folk Science and Belief Versus Opinion. In D. Dennett (Ed.), *Brainchildren* (pp. 135–148). Cambridge, MA: MIT Press.
- Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge (MA): MIT press.
- Fodor, J. A. (1994). *The Elm and the Expert*. Cambridge, MA: MIT Press.
- Hochstein, E. (2012). Minds, Models and Mechanisms: A New Perspective on Intentional Psychology. *Journal of Experimental & Theoretical Artificial Intelligence*, 24(4), 547–557.
- Horowitz, A. (1992). Functional Role and Intentionality. *Theoria*, 58, 197–218.
- Horowitz, A. (2024). *Intentionality Deconstructed: An Anti-Realist Theory*. Oxford: Oxford University Press.
- Jackson, F., Pettit, P. (1990). In Defense of Folk Psychology. *Philosophical Studies*, 59, 31–54.
- Khalidi, M. A. (2013). *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*. Cambridge: Cambridge University Press.
- Kitcher, P. (1993). *The Advancement of Science*. Oxford: OUP.
- Kripke, S. (1980). *Naming and Necessity*. Cambridge (MA): HUP.
- Lahav, R. (1992). The Amazing Predictive Power of Folk Psychology. *Australasian Journal of Philosophy*, 70(1), 99–105.
- Lewis, D. (1972). Psychological and Theoretical Identifications. *Australasian Journal of Philosophy*, 50(3), 249–258.
- Lewis, D. (1980). Mad Pain and Martian Pain. In N. Block (Ed.), *Readings in Philosophy of Psychology* (Vol. 1, pp. 216–222). Cambridge (MA): HUP.
- Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61(4), 343–377.

- List, C., Menzies, P. (2017). My Brain Made Me Do It: The Exclusion Argument Against Free Will, and What's Wrong With It. In H. Beebe, C. Hitchcock, H. Price (Eds.), *Making a Difference: Essays on the Philosophy of Causation* (pp. 269–285). Oxford: Oxford University Press.
- Loar, B. (1988). Social Content and Psychological Content. In R. H. Grimm, D. D. Merrill (Eds.), *Contents of Thoughts* (pp. 99–110). Tucson: University of Arizona Press.
- Marcus, G. A. (2001). *The Algebraic Mind: Integrating Connectionism and Cognitive Science*. Cambridge (MA): MIT Press.
- Menzies, P. (2008). The Exclusion Problem, the Determination Relation, and Contrastive Causation. In J. Hohwy, J. Kallestrup (Eds.), *Being Reduced: New Essays on Reduction, Explanation, and Causation* (pp. 196–217). Oxford: Oxford University Press.
- Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. London: Routledge.
- Putnam, H. (1975a). The Nature of Mental States. In H. Putnam, *Mind Language and Reality: Philosophical Papers Volume 3* (pp. 429–440). Cambridge: CUP.
- Putnam, H. (1975b). The Meaning of 'Meaning'. *Minnesota Studies in the Philosophy of Science*, 7, 131–193.
- Pylyshyn, Z. (1980). Cognitive Representation and the Process-Architecture Distinction. *Behavioral and Brain Science*, 3, 154–169.
- Smolensky, P. (1988). On the Proper Treatment of Connectionism. *Behavioral and Brain Sciences*, 11, 1–23.
- Stich, S. P. (1978). Autonomous Psychology and the Belief-Desire Thesis. *The Monist*, 61(4), 573–591.
- Stich, S. P. (1983). *From Folk Psychology to Cognitive Science: The Case against Belief*. Cambridge (MA): MIT Press.
- Woodfield, A. (1982). On Specifying the Contents of Thoughts. In A. Woodfield (Ed.), *Thought and Object* (pp. 259–297). Oxford: Clarendon Press.
- Woodward, J. (2008). Mental Causation and Neural Mechanisms. In J. Hohwy, J. Kallestrup (Eds.), *Being Reduced: New Essays on Reduction, Explanation, and Causation* (pp. 218–262). Oxford: Oxford University Press.

PIOTR KOZAK \*

## **WHY SHOULD WE BE REALISTS ABOUT BELIEFS? THE RATIONAL INDISPENSABILITY ARGUMENT FOR REALISM**

**SUMMARY:** This paper defends realism about beliefs by presenting the rational indispensability argument, which claims that beliefs are indispensable for explaining how norms of rationality shape human behaviour. I begin by examining the normative aspect of belief ascription and the essential connection between beliefs and rationality highlighted in logic and cognitive psychology. Mild anti-realism, as articulated by Pośłajko, allows beliefs to exist for explanatory purposes but denies them natural properties, claiming they are not causally efficacious or strictly individuable. Against this, I argue—drawing on Wundt’s argument about the relationship between normative disciplines and descriptive sciences—that if rationality norms influence human reasoning, then beliefs must be real in a robust, naturalistic sense: causally efficacious and individuable. I adapt the Quine-Putnam indispensability argument from the philosophy of mathematics to argue that our best theories of rationality require ontological commitment to beliefs as real entities. I conclude, contra Pośłajko, that recognising the explanatory role of rationality norms commits us to a realist metaphysics of belief.

**KEYWORDS:** realism, mild anti-realism, the rational indispensability argument for realism, Wundt’s argument, naturalism.

---

\* University of Białystok, Institute of Philosophy. E-mail: piotr.kozak1@gmail.com.  
ORCID: 0000-0001-9734-4640.

## 1. Introduction

Some people believe that the Earth is flat. Others believe in trickle-down economics. Sometimes, we listen and do not judge, as to err is human. However, there are times when we are, and should be, more judgmental.

Our beliefs might not always be accurate. But, more importantly, they can also be irrational. Imagine some irrational students who simultaneously believe that  $3 > 2$  and  $2 > 1$  but also hold that  $3 < 1$ . We would say this is not only incorrect but also irrational. If they believe that  $3 > 2$  and  $2 > 1$ , then they are compelled to believe that  $3 > 1$ . If they do not believe this, then they should.

Furthermore, the principles of rationality influence the *existence* of beliefs and the way we attribute them. Suppose the same students believe that Frege's Puzzle consists of a thousand pieces and that Popper's Gambit is a cunning chess move. In these cases, we should clarify these concepts for them and at least reconsider our teaching methods. That is what they pay us for.

However, once we clarify the concepts of Frege's Puzzle and Popper's Gambit, students cannot claim that these concepts refer to their standard objects of reference while still asserting that Frege's Puzzle consists of a thousand pieces and Popper's Gambit is a cunning chess move. If they do so, they do not truly understand these concepts and do not hold these beliefs.

The examples above demonstrate the normative aspect of the concept of belief and the practice of attributing beliefs. There are circumstances where we are justified in expecting someone to change their mind, whether due to new facts or conflicting beliefs. Some beliefs may be irrational, and others might make it rational to form additional beliefs. In summary, rationality puts constraints on our beliefs.

This intricate relationship between rationality norms and beliefs is one of the primary reasons for introducing the concept of beliefs into our scientific practice. For instance, it is expected that logic establishes normative standards for reasoning. Logic has been taught with the expectation that it guides our thinking and, consequently, our beliefs. We aim to avoid inconsistencies in our beliefs. We criticise others for failing to recognise the logical consequences of their beliefs. We endeavour to adhere to the principles of logic. Admittedly, this goal has not always been fully realised. However, there have been occasions when success has been achieved.

Secondly, analysing beliefs in relation to rationality is a key aspect of the research paradigm on reasoning in modern cognitive psychology. To cite a few examples, Wason's selection tasks (1968) have highlighted the challenges we encounter in conditional reasoning, which has significantly influenced subsequent theories of reasoning within psychology over the past fifty years. The Heuristics and Biases research programme, started by Tversky and Kahneman (1974), has shown how our decisions often differ from the traditional rationality model, prompting debates on the nature of decision-making in the social sciences during the 1970s and 1980s. Lastly, the so-called Great Rationality Debate

(Gigerenzer, 2015; Stanovich, 1993) has reignited discussions about the rationality of human beliefs.

Granted, most psychological research shows that we make mistakes in our reasoning and do not blindly follow the rules of logic. Most probably, human rationality is bounded (e.g., Gigerenzer, 1993; 2000; Simon, 1990), meaning that reasoning norms are context-dependent and only meaningful in relation to specific contents and situations. Determining how well formal norms align with content and context requires empirical investigation.

Furthermore, the link between normative (how we ought to reason) and descriptive theories of rationality (how we actually reason) remains an open question (e.g., Oaksford, Chater, 2007). On one side, some psychologists, without dismissing the validity of the normative perspective, contend that cognitive psychology should concentrate on the descriptive aspect (e.g., Elqayam, Evans, 2011). On the other side, some psychologists, inspired by the Quinean naturalised epistemology, aim to integrate the descriptive and normative dimensions (e.g., Schurz, Hertwig, 2019). Finally, some believe that purely descriptive research cannot proceed without reference to normative considerations, and empirical results can only be understood in relation to norms of rationality (e.g., Hahn, 2025; for an overview, see, e.g., Chater et al., 2018).

However, no one denies that we are relatively rational, and sometimes we do adhere to norms of rationality (e.g., Stanovich, West, 2000). After all, we can learn the rules of logic and/or probability theory, and there is no argument about our ability to apply them. Moreover, psychologists study how to improve our reasoning to be more rational (e.g., Knauff, Spohn, 2021; Sturm, 2012). In other words, examining whether and how rationality norms influence human beliefs is essential to cognitive psychology. For some thinkers (e.g., Davidson, 1970), these questions distinguish psychology from other special sciences.

In cognitive psychology, the concept of beliefs is fundamentally linked to the idea of rationality. The term “belief” is not just a folk psychological term we use daily; it is a technical term used across various fields of logic and psychology, particularly in relation to how standards of rationality influence our cognitive processes. In other words, the main reason for introducing the concept of belief into philosophy and psychology has been to explain the rationality behind our behaviour.

Exploring the connection between rationality and beliefs is outside the scope of this paper. Some (e.g., Engel, 2018; Shah, Velleman, 2005) link the rationality of beliefs to their truth conditions, suggesting that what sets beliefs apart from other mental states like feelings is that truth determines their correctness conditions, and that attributing beliefs follows norms of truth. Others (e.g., Helton, 2020; Zangwill, 2005) argue that rationality or irrationality is a necessary characteristic of a belief. For instance, believing that  $3 < 1$  requires the ability to revise that belief in accordance with rationality norms.

For the purposes of this paper, it is sufficient to hold that beliefs can be rational or irrational. However, this does not imply that the concept of belief consists solely of normative properties or that beliefs are constitutively normative (e.g., Bykvist, Hattiangadi, 2007; Ciecierski, 2017). Recognising that norms of rationality constrain beliefs does not fully determine the metaphysics of beliefs, either. The idea that beliefs can be rational or irrational aligns with most views on the nature of beliefs, such as representationalism (e.g., Burge, 2010; Fodor, 1975) or interpretationism (e.g., Dennett, 1987).

The only assumption I make is that it is reasonable to believe (pun intended!) that any full-fledged psychological or philosophical theory of beliefs must address the relationship between rationality and beliefs. It is enough to say that explaining how rationality can be integrated into the mind was the main motivation behind Fodor's (1975) language of thought hypothesis and Frege's (1956) views on thinking and language.

I will not explore the nature of rationality. It surpasses both the author's philosophical expertise and the readers' patience. For our purposes, let us accept that the core of the complex concept of rationality involves using beliefs, justifications, and logic to guide behaviour—judgments, decisions, actions—and to achieve goals (cf., e.g., Audi, 2001; Nozick, 1993).

It is debatable whether rationality is universal, what role context plays, or how to define different types of rationality (e.g., structural or practical). Nevertheless, it is widely accepted that rationality, in any form, requires norms. Rationality norms are rules that determine how we think, act, and make decisions to be considered rational. For instance, it is irrational to hold incoherent beliefs—such as believing that you are a great mathematician, that great mathematicians can count, and yet believing that you cannot count—since this contradicts the principle of non-contradiction. Similarly, it is irrational to follow a horoscope, as it violates the norm that one should not believe in something against the evidence.

Norms of logic are a subset of rationality norms. The latter also includes rules of probability and norms of decision theory, practical rules, such as guidelines for efficiency or basing judgments on evidence, etc. However, for the sake of brevity, throughout the paper, I refer to the concept of rationality norms, mainly focusing on the rules of logic.

For the purpose of analysis, I assume that the norms of rationality can influence our behaviour. For instance, based on the principle of non-contradiction, you may abandon one of the inconsistent beliefs and, based on the evidence, cease following a horoscope. If the norms of rationality could not influence our behaviour, a large part of research in cognitive psychology, logic, the philosophy of mind, not to mention common knowledge, would be at risk.

Now, a careful reader of Pośłajko's (2024) ingenious book might shrug. The author does not discuss the topic of rationality. Why on earth should anyone care about the lengthy introduction on the relationship between rationality and beliefs if we are discussing the topic of the existence of beliefs? Moreover, a benevolent

reader of the book might argue that the connection between rationality and beliefs is secured in Posłajko's mild anti-realism framework. One might say that the topic of rationality is not relevant for being a mild anti-realist.

The core of Posłajko's sophisticated argument for mild anti-realism is based on a distinction between being real in terms of existence and being real in terms of possessing natural properties. According to the first understanding, being real involves being individuated by and used within explanatory contexts in scientific practice. Using Posłajko's examples, the concepts of race and gender refer to existing phenomena and are employed in scientific explanations. Therefore, race and gender exist.

According to the second understanding, being real means representing natural kinds and possessing natural properties. A kind is *natural* iff it reflects the structure of the natural world rather than our interests and actions. Sharing natural properties, as proposed by Lewis (1983), refers to the idea that objects sharing these properties are objectively similar; that natural properties are relevant in causal explanations; and that they can be defined in terms of fundamental natural properties. Atoms are considered real because they meet these criteria. The concepts of race and gender probably do not meet them. Therefore, race and gender are not considered real, at least in the same sense as atoms are.

The main idea of the book is that we can apply this distinction to beliefs. Posłajko navigates between the pitfalls of eliminativism and industrial-strength realism, arguing that, just as race and gender, beliefs exist. However, they do not possess natural properties. Therefore, they are not real.

The same distinction swiftly applies to the idea of rationality. Do norms of rationality exist? I see no good reason to deny it. Do norms of rationality have natural properties? Not really. Therefore, just as one can be a mild anti-realist about beliefs, one can also be a mild anti-realist about rationality. Although it says little about the nature of the connection between rationality and beliefs, it makes it conceptually plausible that these two ideas can be described within the same anti-realist framework. The case is settled.

I argue that this conclusion is flawed. I contend that recognising the influence of norms of rationality on reasoning and actions requires us to adopt a realist perspective on beliefs. In essence, beliefs are indispensable in explaining how the norms of rationality can influence our behaviour and, therefore, must be considered real in terms of possessing natural properties. Since my argument partly resembles the Quine-Putnam indispensability argument for realism in the philosophy of mathematics, I call it the rational indispensability argument for realism about beliefs.

*Main thesis:* If norms of rationality were to be applicable and shape human behaviour, then beliefs must be real in terms of possessing natural properties.

In short, I argue that our best theories of reasoning in cognitive psychology assume that rationality norms can influence our behaviour. This assumption involves specific ontological commitments. The ontological commitments of a theory are the kinds of entities that must exist for the theory to be true. The ontological commitments of our theories of reasoning are that beliefs mediate between norms and behaviour, and, thus, are causally efficacious and strictly individuatable; therefore, they are real.

In light of the rational indispensability argument, mild anti-realists can formulate three kinds of responses. They can either argue that there are good reasons not to accept the conclusion; they can deny that norms of rationality apply to human behaviour; or, lastly, they can defend their position by claiming that the distinction between mild anti-realism and realism is merely a matter of definition. I argue that avoiding anti-realist positions is the most convincing choice.

I proceed as follows. I start by presenting Wundt's argument for psychologism in the debate on the nature of laws of logic. Although this argument may not be sufficient to establish that laws of logic are psychological entities, it provides us with valuable insights into the relationship between normative and descriptive properties. Building on Wundt's argument, in the next section, I introduce the rational indispensability argument for realism about beliefs. I contend that if norms were to influence our behaviour, then we must recognise that beliefs exist, are causally efficacious and are strictly individuatable. Therefore, beliefs are real in terms of possessing natural properties. In the final section, I outline possible responses to the rational indispensability argument and demonstrate that they fall short.

## 2. Wundt's Argument

These days, philosophers love naturalism. However, there was a time when, and a place where naturalism was a philosophical non-starter. Just before the Great War, German-speaking philosophers were preoccupied with the so-called *Psychologismusstreit*. The heated discussion revolved around whether logic is part of psychology. Psychologists<sup>1</sup> answered yes. Nonetheless, the central figures of this controversy, Frege and Husserl, strongly opposed any attempts to naturalise logic. Long story short, after fierce critiques from both philosophers, the term "psychologism" became an insult, denoting the fallacy of confusing psychological entities with non-psychological ones.

This paper does not aim to evaluate the advantages and disadvantages of psychologism, especially since it has never been a precisely defined stance (Beiser, 2014; Gray, 2008). Instead, it functions more as a collection of arguments that

---

<sup>1</sup> Notably, one of the most prominent psychologists was also a founding father of modern logic, George Boole. The aim of his groundbreaking work on propositional calculus was to study the laws of reasoning: "The design of the following treatise is to investigate the fundamental laws of those operations of the mind by which reasoning is performed" (Boole, 1951, p. 1).

could be labelled psychologistic. Likewise, I do not intend to assess the validity of Frege and Husserl's arguments. Instead, I will present one of the strongest arguments in favour of psychologism, with the sole aim of using it to support the realism of beliefs. Since it was originally introduced by Wundt (1880/1883), let us refer to it as Wundt's argument.

The structure of Wundt's argument is quite clear. The main premise states that any normative discipline setting prescriptive rules for behaviour must be based on a descriptive discipline. The idea is that for a rule to be valid, there must be a descriptive foundation outlining the conditions under which the rule applies. Let us compare a legal rule requiring moving faster than the speed of light and the one mandating turning left. Only the second one is binding. The first one is invalid, since moving faster than the speed of light, according to all that we know about the world, is physically impossible.

The second premise is also quite uncontroversial. It asserts that logic is a normative discipline that prescribes rules for thinking. For example, if from  $p \rightarrow q$  and  $\neg p$  I infer that  $\neg q$ , I commit the inverse error and should reconsider my reasoning. I am bound to do so because the rules of logic instruct me to act accordingly.

The final premise asserts that only one science can serve as a descriptive foundation for logic. It is psychology.

From these three premises, Wundt concludes that logic must be based on psychology.

#### *Wundt's Argument*

(P1) Normative disciplines that establish prescriptive rules for behaviour are founded on descriptive disciplines.

(P2) Logic is a normative discipline concerning rules of reasoning (laws of thought).

(P3) Psychology is the only descriptive science that can form a basis for logic.

---

(C) Logic is based on psychology.

The strength of Wundt's argument was recognised very early by both Frege and Husserl. Despite claims sometimes made (e.g., Føllesdal, 1958, p. 49), Frege and Husserl never argued that the distinction between psychology and logic rests on the is-ought distinction. Frege (2007) argues that logic includes both descriptive rules, such as if it is true that  $a$ , then  $a$ , and prescriptive rules, like if you accept the truth of  $a$ , then you must accept  $a$ . However, the descriptive rules form the foundation for the prescriptive ones. According to Frege, the fallacy is concealed in the third premise. It is the confusion between being true, which is an objective property and belongs to the realm of logic, and what is held to be true, which is subjective and falls under psychology, that underpins the psychological fallacy.

Husserl is even more straightforward. He recognises that normative disciplines have descriptive foundations (2001, §§14–16) and openly criticises the idea that psychology and logic can be distinguished by the is-ought distinction (2001, §17). On this matter, he agrees with psychologists and asserts that thought as it ought to occur is a special case of thought as it actually occurs. He dismisses psychologism due to its consequences, believing it inevitably results in the pitfalls of relativism and scepticism.

Clearly, the aim of the paper is not to settle the psychologism debate nor to assess the validity of Frege and Husserl's arguments. It is plausible, following Frege, that psychologists have been misled by the confusion surrounding the ambiguous concept of laws of thought. It may be that psychologists confuse two different types of enterprises: developing a theory of deductive logic concerning the relation between truth-preserving entities, on one side, and a theory of the psychological principles of belief revision, on the other. Citing Harman (1984, p. 107): "Logical principles are not directly rules of belief revision. They are not particularly about belief at all".

Therefore, Wundt does not need to be correct about the conclusion. Nonetheless, the first two premises of Wundt's argument remain valuable to retain.

The main idea of Wundt's argument is that if logic were to establish rules for our behaviour, it must be based on a descriptive foundation that explains why these rules are applicable. In other words, if logic is to prescribe reasoning rules, it must be consistent with the principles of the psychology of reasoning. We can refer to this as the "truth" of Wundt's argument. This thesis comes with two significant caveats.

Firstly, the "truth" of Wundt's argument does not require us to believe that logic has a privileged role in reasoning or that its principles form the basis of thinking. Suppose one believes that  $p$  and  $p \rightarrow q$ . It does not necessarily force me to believe that  $q$ . For example, believing that  $q$  can be at odds with empirical evidence, in which case it could be irrational to form a belief that  $q$ .

Similarly, suppose some rule of logic requires considering all alternatives of our scientific hypothesis before reaching a conclusion. Although the content of this logical rule does not seem very different from the rule of Modus tollens, due to the psychological limits of our cognition, attempting to follow this rule could render any hypothesis undecidable, thereby invalidating the rule.

Thus, logical rules cannot be identified with the norms of rationality. They cannot be identified with psychological reasoning rules either. As Frege and Husserl indicated, this would be a clear case of the mistaken identification of non-psychological entities with psychological ones. Logical rules are, at best, an idealisation of both rationality norms and psychological rules.

The "truth" of Wundt's argument simply involves recognising an interesting normative connection between logic and human reasoning. In other words, the principles of logic are *applicable* and can shape the structure of human reasoning. It does *not* require us to believe that the principles of logic are identical to the psychology of reasoning.

The rules of logic can influence our minds without being identified with the psychological mechanisms involved. For example, consider that human reasoning aims to accurately represent the world, and that this goal cannot be achieved if beliefs contain inconsistencies. Logic is a theory about the relationships between propositional contents. Therefore, understanding the rules of logic can be important for eliminating inconsistencies in our beliefs.

Furthermore, recognising the theoretical applicability of the rules of logic is crucial if we want to keep logic useful for understanding the world. If someone believed that logic has no real-world application at all, it would reduce logic to a childish game, similar to puzzles. We all enjoy puzzles, just as we enjoy chess. However, unlike departments of logic, universities do not need departments dedicated to puzzles and chess.

Secondly, the “truth” of Wundt’s argument can be understood in two ways. On the one hand, it reduces to the formula ascribed to Kant that “ought implies can”, expressed in deontic logic with the axiom  $OA \rightarrow \Diamond A$ . Thus, if we are obliged to perform a certain action, then it must be possible to perform it, and, by the rules of Modus tollens, if it is not possible, then it is omissible. For example, we do not expect newborns to follow the rules of logic, since their cognitive skills are not sufficiently developed to follow these rules.

On the other hand, the “truth” of Wundt’s argument asserts that if there is a normative link between logic and psychological facts of human reasoning, then a way of bridging the gap between them must exist. Psychology explains how the mind functions causally. Since psychology underpins the applicability of logical norms, they must be integrated into the causal mechanisms of the human mind. Without this integration, it would be difficult to understand how these rules can be valid for human reasoning or how they influence our behaviour. In what follows, I mainly focus on this second interpretation of the argument.

And obviously, Wundt’s argument says very little, if anything, about how to bridge this gap between rules of logic and reasoning. It does not identify any crucial causal factors that facilitate the transfer from the content of a logical rule to applying this rule in actual reasoning. It only states that this causal transfer, in principle, should be possible if the rules of logic were applicable to human reasoning.

In the next section, I will argue that recognising the “truth” of Wundt’s argument helps us to develop an argument for the realism of beliefs. I will demonstrate that the applicability of the norms of rationality requires beliefs to be causally effective and strictly individuatable, and, therefore, real in Pośłajko’s sense.

### **3. The Rational Indispensability Argument for the Realism of Beliefs**

Let us move on to the second stage of the argument. In the philosophy of mathematics, there is a long-standing debate about the existence of mathematical entities. One of the most influential arguments supporting mathematical realism is the so-called Quine-Putnam indispensability argument (Putnam, 1979; Quine,

1980a; 1980b). In this section, I will formulate an analogous argument for the reality of beliefs. Because it pertains to the indispensability of beliefs for the applicability of the norms of rationality, I will refer to it as the rational indispensability argument for the realism of beliefs.

Let us begin with the Quine-Putnam indispensability argument. According to the first premise of the argument, we should have ontological commitments to all and only the entities that are indispensable to our best scientific theories. The second premise states that mathematical entities are indispensable to our best scientific theories. Therefore, we should have ontological commitments to mathematical entities and, consequently, be realists about them.

*The Quine-Putnam Indispensability Argument*

(P1) We ought to have ontological commitments to the entities that are indispensable to our best scientific theories.

(P2) Mathematical entities are indispensable to our best scientific theories.

---

(C1) We ought to have ontological commitments to mathematical entities.

(C2) We ought to be realists about the existence of mathematical entities.

Two clarifications are necessary. First, the concept of indispensability should be understood as the idea that mathematical entities cannot be eliminated from our best theories because doing so would make these theories less attractive. They would lack empirical success, explanatory power, and unifying ability, among other qualities. Clearly, this is a matter of mathematics.

Secondly, although the second premise seems non-controversial, the first requires support. The validity of the first premise derives from accepting confirmational holism, which holds that scientific theories are confirmed and disconfirmed as wholes (Quine, 1980b). If a theory is confirmed by empirical findings, the entire theory is confirmed. Therefore, if mathematical entities are employed in a theory, their existence is also affirmed.

Clearly, there is no need to delve into the metaphysics of mathematics. Nor is it necessary to analyse the responses to the Quine-Putnam indispensability argument. For this paper, it suffices to observe that a similarly structured argument for the reality of beliefs can be formulated. Let me clarify this.

The core of the rational indispensability argument holds that the concept of belief is a technical term used in theories of reasoning within cognitive psychology to explain how rationality requirements can influence our behaviour. If these rationality requirements are to be applied—and we accept that they are—then we must also recognise that beliefs, such as my acceptance of the truth of the *Modus tollens* rule, can affect our behaviour. This, in turn, suggests that beliefs must have causal powers and generally remain consistent in content, which amounts to claiming they are real in Posłajko's sense.

The structure of the argument, in simplified form, is as follows. The first premise remains unchanged and maintains that we ought to make ontological

commitments regarding the entities our best theories require. The second premise argues that beliefs are indispensable to theories explaining the applicability of norms of rationality to human behaviour. Therefore, we ought to have ontological commitments to beliefs and be realists about them.

*The Rational Indispensability Argument (Simplified)*

(P1) We ought to have ontological commitments regarding the entities required by our best theories.

(P2) Beliefs are indispensable to theories explaining the applicability of norms of rationality to human behaviour.

---

(C1) We ought to have ontological commitments to beliefs.

(C2) We ought to be realists about the existence of beliefs.

Explaining the second premise is essential. Note that if the argument only claimed that we use the concept of beliefs in our scientific theories, it would face two counterarguments. The first is that belief is a folk psychological concept that can and should be replaced by a more precise one. In other words, beliefs are dispensable. This forms the core of the eliminative strategy concerning the metaphysics of beliefs. The second counterargument is that if the argument simply asserts that beliefs must exist, it does not imply that they are real in terms of possessing natural properties. This is the main idea of Pośłajko's mild anti-realist approach. However, the purpose of the argument is to show that beliefs are indispensable and real in Pośłajko's sense. Let me start with the claim of indispensability.

According to eliminativists, the concept of beliefs is a folk psychological term mainly used in theories of other minds. In brief, we assume beliefs exist to explain others' behaviour. However, according to eliminativists, the concept of belief is a prescientific term used as a shortcut in reasoning and should, in principle, be replaced by a more sophisticated concept.

However, this objection does not apply to the rational indispensability argument. In logic and in theories of reasoning in cognitive psychology, the concept of belief is not a folk-psychological one. Instead, it is a technical term referring to attitudes towards propositional content. I do not know many folks, but I doubt they would be happy to substitute the expression "I believe that  $p$ " with "I have a cognitive propositional attitude that  $p$ ". Granted, there are many things we do not know about propositional attitudes. And yet, we do know something about them. We can describe the relationship between different attitudes. We know they are related to truth and so on. In this technical sense, the concept of belief is indispensable, as any attempt to eliminate it would make our best theories of reasoning in logic and cognitive psychology less attractive.

Eliminativists might argue that this response can be easily challenged by distinguishing between two different understandings of the concept of belief. Psychologists use one, while philosophers use the other. Philosophers use the technical term, which pertains to idealised entities, such as propositional attitudes;

psychologists refer to causally grounded mental states. Their objections pertain only to the psychological understanding of this term, whatever such understanding might be. However, although distinctions are generally useful, they do not resolve the core issue.

According to the rational indispensability argument, beliefs, defined as attitudes towards propositional content, can causally influence certain psychological states. If these psychological states were categorically different from propositional attitudes, then any attempt to explain the causal influence of propositional attitudes on these states would risk making a categorical error. However, we assume that propositional attitudes do causally affect some psychological states; therefore, philosophical and psychological understandings of the belief concept cannot belong to different logical categories. If we presuppose that there is a way to bridge the gap between propositional attitudes and psychological states, philosophical and psychological understandings of the belief concept cannot be categorically different. Let me revisit this point shortly.

According to mild anti-realists, the rational indispensability argument might not succeed. They could accept the premises and the conclusion but remain sceptical about the reality of the beliefs. Following Pośłajko, they may argue that the argument only shows that beliefs exist, which they willingly endorse, but that they are not real in the sense of lacking natural properties. According to mild anti-realists, beliefs are neither strictly individuatable, causally efficacious, nor definable in terms of fundamental natural properties. Can mild anti-realists rest easy? Well, easy does it!

Let us reconsider Wundt's argument. According to its "truth", if logic, in particular, and rationality norms, in general, were to influence human reasoning, then there must be a way to bridge the gap between the norms of rationality and the psychological mechanisms of reasoning. If there were no way to bridge the gap, then the norms of rationality would be useless in altering our behaviour. Yet, we reluctantly assume they do have a use.

What are our ontological commitments regarding the applicability of rationality norms? Norms of rationality can influence our behaviour only if we have cognitive access to them. For example, we need to know the Modus tollens rule to apply it to our reasoning. Even though we do not have to be aware of the exact formulation of the rule, we are required to know that if we believe that  $p \rightarrow q$  and  $\neg q$  then it would be incorrect to believe that  $p$ .

Moreover, we must adopt a special attitude towards propositional contents. We must hold that the Modus tollens rule is valid and take for granted that  $p \rightarrow q$ , and  $\neg q$ . We cannot doubt it or deny it. In other words, to follow the rule of Modus tollens, we are required to possess a specific attitude towards the proposition that it is not true that  $p$  if  $p \rightarrow q$  and  $\neg q$ . It is another way of saying that I have a *belief* that  $\neg p$ . Thus, the applicability of rationality norms implies the existence of beliefs.

Furthermore, we must recognise that norms of rationality influence our reasoning. Only through this influence can we establish the validity of these

norms. If they could not affect our reasoning, they would be deemed inapplicable and pointless.

The only way that norms of rationality can influence our reasoning is by being part of a causal chain. For instance, being aware of  $p \rightarrow q$  and  $\neg q$ , as well as the validity of the Modus tollens rule, makes one aware that  $\neg p$  and, in turn, *can* (though it does not have to) change one's behaviour. Therefore, the applicability of the norms of rationality implies not only that beliefs exist, but also that they are causally efficacious. In a word, if one argues that the applicability of the norms of rationality requires us to accept the existence of beliefs and that these norms can influence our reasoning, then it follows that beliefs must possess causal powers. Let us call this the no-gaps argument. Allow me to explain it.

Suppose you believe that  $p \rightarrow q$  and  $\neg q$ , as well as the validity of the Modus tollens rule. Next, suppose it does indeed make you believe that  $\neg p$ . In this case, you start with a set of propositional attitudes towards the content  $p \rightarrow q$ ,  $\neg q$ , and the validity of the Modus tollens rule. Let us call it the set of beliefs  $A$ . You end with a set of propositional attitudes towards the content of a set  $A$ , plus the attitude towards the content that  $\neg p$ . Let us call it  $A^*$ . The causal relation between  $A$  and  $A^*$  may not be clear to us. Yet, assuming that we are naturalists, we are bound to hold that there are no causal gaps. There must be some sort of causal relationship between  $A$  and  $A^*$ . If not, you must accept some sort of supernatural interaction between our attitudes towards propositions. Therefore, assuming you are a naturalist, beliefs are causally efficacious.

Holding that beliefs are causally efficacious does not oblige us to accept the language of thought hypothesis. It has never been, contra Fodor, the only game in town. Let me provide two examples of possible theories. First, the idea that logic must be causally embedded within human reasoning was the main reason why Frege was interested in language. In short, Frege believes that language is necessary for thinking, as only through language can we grasp the content of our thoughts. If words can influence our behaviour, and they in principle can, then beliefs, through language, can do the same. Second, according to dispositionalism, having beliefs is equivalent to manifesting a certain pattern of behaviour. Belief is a relational property that manifests a specific structure of physical events. Generally, relational properties have causal powers. For example, a poorly placed balance point can cause the bridge to collapse. Similarly, patterns of behaviour can give rise to other patterns. For dispositionalists, the causal powers of beliefs are a feature, not a bug.

Thus, not only must we assume that beliefs are causally efficacious, but we also have theories that attempt to explain how this is possible.

Finally, the applicability of the norms of rationality requires that beliefs be strictly individuatable. This means that the content of our beliefs should remain stable across time and between individuals. For example, to follow the rule of Modus tollens and expect others to do the same, I must assume that it is the same propositional attitude and the same rule content that applies universally across time. If rules and attitudes towards them were not stable, it would be im-

possible to determine whether one is following rules, and therefore, they would be inapplicable.

Suppose I follow the Modus tollens rule and expect others to do likewise. Clearly, to adhere to this rule, I must assume that its content is stable and distinguishable from other contents. It cannot be the case that at a time  $t_1$ , I accept the truth of the statement that if a conditional statement  $p \rightarrow q$  is true, and the consequent is false, then the antecedent must also be false, and at a time  $t_2$ , I hold that in the same case, the antecedent must be true.

Furthermore, to adhere to the Modus tollens rule, I must hold the same and distinguishable attitude towards its content. I must accept the truth of its content. Doubting or denying the truth of this rule counts as a different propositional attitude. Therefore, assuming that we can follow the Modus tollens rule, beliefs must be strictly individuable.

To sum up, the applicability of the norms of rationality implies that beliefs are causally efficacious and strictly individuable. The fully developed form of the rational indispensability argument can be presented as follows.

*The Rational Indispensability Argument (Fully Developed)*

- (P1) We ought to have ontological commitments regarding the entities required by our best theories.
  - (P2) Beliefs are indispensable to theories explaining the applicability of norms of rationality to human behaviour.
  - (C1) We ought to have ontological commitments to beliefs (from P1–2).
  - (C2) We ought to be realists about the existence of beliefs (from P1–2).
  - (P3) Norms of rationality influence human reasoning (general assumption).
  - (P4) Applicability of the norms of rationality commits us to hold that (from Wundt's argument):
    - (P4.1) Beliefs are causally efficacious
    - (P4.2) Beliefs are strictly individuable.
- 
- (C3) Beliefs exist and are causally efficacious and strictly individuable.

Let us examine the final feature mentioned by Pośłajko: being definable in terms of fundamental natural properties. To my knowledge, it is the most debated property introduced by Pośłajko. Essentially, there is no consensus on what constitutes the most fundamental natural properties (e.g., Dorr, Hawthorne, 2013). Possessing spin can be regarded as such. Nonetheless, a significant grey area remains beyond that.

There are two problems with the condition of being definable in terms of fundamental natural properties. According to functionalists and dispositionalists, beliefs can be defined by their relational properties: their causal roles and behavioural patterns. Do relational properties count as being fundamental and natural?

I see no good reason why not. For example, according to the theory of relativity, the geometrical properties of space are considered primitive and among the most fundamental natural properties of space. According to Newtonian physics, energy is a relational *and* fundamental property. The topological structure of DNA is natural and influences the expression of genetic information. If we have no trouble agreeing on that, then we would be applying double standards if we deny this status to the relational properties of beliefs.

Furthermore, and most importantly, introducing the concept of fundamental natural properties appears to predetermine the outcome of the debate. We can agree with Lewis (1984) that the concept of naturalness aids in developing the idea of realism. We can also agree with him that natural properties are causally efficacious and build up the world. However, excluding relational properties from the set of natural properties is question-begging. If we exclude relational properties from the set of fundamental natural properties, then the debate over whether beliefs are real is effectively settled, since the only remaining positions are identity theory and the Fodorian industrial-strength realism. Non-Fodorian functionalism and dispositionalism are dismissed prematurely. Consequently, we could achieve an easy victory in the debate, but only through definitional tricks. Such victories bring no glory.

Therefore, it is not a good move to start with the idea of fundamental natural properties and then argue that mental entities proposed by some theories lack causal powers and cannot be individuated. Instead, it is more prudent to claim that if some mental events are causally efficacious and individuable, then they possess natural properties. According to the rational indispensability argument, beliefs are causally efficacious and individuable. Therefore, they possess natural properties and are real in Pośłajko's sense.

#### 4. Possible Responses

There may be at least three responses to the rational indispensability argument. First, one can accept the premises but reject the conclusion. If there is strong evidence against the conclusion, then it would be irrational to accept it. Second, one can argue that one of the premises is flawed, rendering the entire argument unsound. Third, one can accept the conclusion but claim that the argument overlooks an important aspect of the mild anti-realist position. Let us begin with the first.

According to the first response, as carefully described by Pośłajko, there is strong evidence that the folk-psychological concept of belief does not refer to any mental entity that can possess causal powers and be strictly individuable. Let us start with being individuable. Pośłajko (2024, p. 95), following Connors and Halligan (2015, p. 3), argues that in cognitive psychology, beliefs, although they share several common properties, are best seen as varying across different dimensions within these properties. For example, they can differ in their effects

on cognition and behaviour. If they vary across dimensions, then they must lack natural properties and, therefore, are not real.

However, varying across dimensions does not imply that something lacks natural properties or is unreal. It simply indicates that it is a multidimensional concept. Velocity (but not speed) is a multidimensional concept that varies across different dimensions, yet this does not mean we cannot measure velocity, nor does it suggest that it is unnatural or unreal. Biodiversity is also a multidimensional concept that impacts the environment in various ways. Nevertheless, we can measure biodiversity, and no one should deny its reality. Being multidimensional and varying across dimensions is, at most, a sign of a coarse-grained concept. However, this is not problematic, as most scientific concepts are coarse-grained.

Posłajko's argument against the causal powers of beliefs is more sophisticated. He argues (2024, p. 98, Footnote) that proponents of beliefs are unable to "provide an account of how contentful mental states might be seen as being genuinely causally active" (2024, p. 106). For the sake of brevity, I will not delve into the details of his argument. Instead, I will raise two doubts regarding his conclusion.

First, based on the no-gaps argument, if you are a naturalist and accept that the norms of rationality apply to reasoning, then you are compelled to recognise the causal relevance of beliefs. Admittedly, there is another story to tell about how this causal relationship operates. However, it is different to say that we do not understand how something works and that it cannot work.

Second, proponents of dispositionalism and functionalism seem to be immune to Posłajko's causal power argument. Dispositionalists and functionalists define beliefs in terms of relational properties of dispositions and causal roles. We can discuss in what sense relational properties can be causally efficacious. However, it is a reasonable and plausible option that they *are* causally effective (e.g., McKittrick, 2005). Therefore, the reasons to doubt the conclusion of the rational indispensability argument do not appear to be conclusive.

Let us consider the second type of response. One may oppose the rational indispensability argument by rejecting one of its premises. The most plausible option is rejecting the premise that norms of rationality apply to human reasoning. Since this response was suggested to me in a private conversation by Maja Kittel, I refer to it as Kittel's counterargument from a-rationality.

According to Kittel's counterargument, it is a philosophical wishful thinking to believe that norms of rationality apply to human behaviour. We wish for them to apply, but they only serve to justify the existence of courses in logic. Similarly, we argue that Santa exists to justify the tradition of gift-giving.

Kittel's counterargument can be understood in two ways. The first proposes that norms of rationality in general, and rules of logic in particular, are not the most effective tools for describing human reasoning psychology. Even if they were, it does not specify which types of rationality norms (such as practical or theoretical), and rules of logic (like classic, non-monotonic, or modal), rules of

probability or decision theory, are most appropriate for that purpose. The second suggests that these norms and rules do not influence human reasoning. Let me begin with the first interpretation.

According to the indispensability argument, norms of rationality, including rules of logic, may, though they need not, influence human reasoning. However, this does not require us to believe that these norms and rules depict actual mechanisms of human reasoning and belief revision processes. Admittedly, it would be easier to understand how rationality norms are applied if they were part of human psychology. This is precisely what Wundt concludes in his argument. Nevertheless, the indispensability argument remains valid even if we reject Wundt's conclusion.

Norms of rationality and rules of logic do not necessarily concern beliefs at all. They might describe relations between propositional contents and/or actions. At most, they could be seen as idealisations of the real process of belief revision.

What matters for the indispensability argument is that norms of rationality in general, and rules of logic in particular, apply to human reasoning not in terms of describing it, but in terms of affecting it. We can be guided by rationality norms and follow rules of logic. Metaphorically speaking, even if we are not rational animals, we sometimes act as if we were (and this is what makes us rational). Acting as if we were rational, however, commits us to believe that beliefs are real.

Therefore, regarding the indispensability argument, it is irrelevant which rationality norms or rules of logic best describe human reasoning and belief revision processes. It could be classical logic (cf., e.g., Stenning, Lambalgen, 2008), but it does not have to be (cf., e.g., Bishop, Trout, 2005). It might also be a combination of problem-solving techniques. The only premise we need to accept is that, on occasion—though rarely—people do use some, but not all, norms of rationality in reasoning.

The second interpretation of Kittel's argument challenges this very premise. Kittel's counterargument is supported by extensive empirical evidence on the irrationality of human reasoning. Even this paper begins by referencing the renowned Wason's selection task paradigm and the Heuristics and Biases programme in psychology. Have not they demonstrated that rationality is merely a myth?

Granted, most subjects fail in Wason's selection tasks. They rely on dependable heuristics rather than deductive reasoning. Nevertheless, we should not be overly pessimistic about this.

Let us examine Wason's selection tasks. You are presented with four cards on a table, each having a number on one side and a colour on the other. The visible sides show the digits 3 and 8, along with the colours blue and red. Which card(s) should you turn over to verify the rule that if a card displays an even number, then its opposite side must be blue?

After brief consideration, you can easily solve this task using *Modus ponens* and *Modus tollens*. The correct answer is to flip the 8 card and the red card.

However, less than 10% of the subjects in the original study answered correctly. Is not this definitive proof that we are irrational and that logic has no influence on our reasoning? No, it is not. Even though correct answers were uncommon, they did occur. Wason's selection test shows that we struggle with nonintuitive tasks. Still, no one claims that logic must be intuitive. And yet, some people do use logic to reason. This makes the applicability of the rules of logic even more remarkable.

Nonetheless, there is one final way to challenge the premises of the argument. One can argue that, although norms of rationality apply to human reasoning, their applicability does not necessarily mean that beliefs are real (P4). Consider a possibly existing legal norm banning witchcraft. Does its existence imply that witchcraft is real? Not necessarily. Similarly, even if rationality norms are binding, they do not force us to believe that beliefs are real.

Well, is that so? No! The difference between the legal rule banning witchcraft and the Modus tollens rule is that only the latter is valid. Banning witchcraft, although such a rule could exist, is not applicable because, to the best of our knowledge, witchcraft does not exist. The rule may exist, but it is not valid. Nevertheless, it would be binding if we discovered that witchcraft is real.

In contrast, the Modus tollens rule exists and, to the best of our knowledge, is valid and applicable. If it is valid and can be applied, then (according to Wundt's argument) it commits us to the view that it can be integrated with the causal mechanism of reasoning. It can only be integrated if we adopt certain propositional attitudes towards its content, which boils down to holding that we possess beliefs. Therefore, beliefs must be real.

Finally, let us consider the last response. The proponents of mild anti-realism may accept the conclusion of the rational indispensability argument, even if it means biting the bullet. However, they might still wish to preserve the spirit of mild anti-realism. According to this view, recognising that beliefs are real in both senses mentioned by Pośłajko overlooks an important distinction between beliefs and other mental states. As Pośłajko argues, beliefs are not simply mental objects that exist in some Cartesian mental space. Asking about the nature of beliefs is not the same as asking about natural properties of the mind. Differentiating between existing and being real helps us understand that beliefs exist, but they do so differently from atoms and quarks. Similar to the concept of gender, they are constructed for scientific purposes. Therefore, beliefs are largely constructed concepts (e.g., Mathews, 2013) and do not constitute natural kinds.

Now, according to most realists, excluding perhaps the proponents of the identity theory, beliefs exist but they also exist differently from atoms and quarks. Therefore, the difference between realists and mild anti-realists can be subtle and might simply be a matter of using different terms.

Ecumenism is a virtue, but sometimes it can lead us astray from the path of salvation. The success of the last response depends on how we understand being constructed. On the one hand, being constructed may mean not referring to any object of a natural kind. This is where mild anti-realists and some realists con-

verge. On the other hand, being constructed could mean that a category is a product of how we describe the world. That is what interpretivists would argue.

However, there is another way to develop the idea of being constructed. It could mean being a product of our measurement practices. For example, being 10 on the Apgar scale is a product of our measurements. This property did not exist before the introduction of the Apgar scale. Nonetheless, being 10 on the Apgar scale, although multidimensional, remains an individuatable property and does possess causal powers. That is what the realists like Dennett (1991) support.

According to the rational indispensability argument, if norms of rationality are to influence behaviour, beliefs must exist and possess certain natural properties. Admittedly, they do not need to exist like atoms and quarks; they can be more akin to a score of 10 on the Apgar scale. Nonetheless, in all these cases, they must be causally efficacious and individuatable. If they were not, then, according to the rules of *Modus tollens*, norms of rationality would be ineffective.

Therefore, we do not quarrel over words. We discuss the metaphysics of beliefs and the essence of rationality. That is why we should be realists about beliefs.

## 5. Conclusion

In this paper, I have examined the intricate relationship between rationality and beliefs, arguing that beliefs are essential for explaining how norms of rationality shape human behaviour. I start by discussing the normative role of beliefs and how rationality constrains our beliefs. Then, I introduce the concept of mild anti-realism, which suggests that beliefs exist but lack natural properties. I criticize this view, asserting that recognising the influence of rationality norms on reasoning requires us to be realists about beliefs. I present the rational indispensability argument, claiming that beliefs are indispensable for the application of rationality norms and therefore must be real. I argue that the applicability of rationality norms commits us to the existence of beliefs that are causally efficacious and strictly individuatable. Consequently, beliefs are real in the sense defined by Pośłajko.

And although I do not agree with Pośłajko's main thesis, I do find his book both challenging and thought-provoking. There is no better recommendation for the book than discussing it. Furthermore, I do share Pośłajko's view that we are misled by futile attempts to localise beliefs as some objects existing in the Cartesian theatre. The only point where I differ is that beliefs are not real. I have argued that we have good reasons to treat them as real. Whether they are more like the concept of gender or like being 10 on the Apgar scale, and what that even means, is a different matter for discussion.

## REFERENCES

- Audi, R. (2001). *The Architecture of Reason: The Structure and Substance of Rationality*. Oxford: Oxford University Press.
- Beiser, F. C. (2014). *The Genesis of Neo-Kantianism, 1796–1880*. Oxford: Oxford University Press.
- Bishop, M. A., Trout, J. D. (2005). *Epistemology and the Psychology of Human Judgment*. Oxford: Oxford University Press.
- Boole, G. (1951). *An Investigation of the Laws of Thought, on Which Are Founded the Mathematical Theories of Logic and Probabilities*. New York: Dover.
- Burge, T. (2010). *Origins of Objectivity*. Oxford: Oxford University Press.
- Bykvist, K., Hattiangadi, A. (2007). Does thought imply ought? *Analysis*, 67(296), 277–285.
- Chater, N., et al. (2018). Mind, Rationality, and Cognition: An Interdisciplinary Debate. *Psychonomic Bulletin & Review*, 25, 793–826.
- Ciecierski, T. (2017). Attitudes and Normativity. *Axiomathes*, 27(3), 265–283.
- Davidson, D. (1970). Mental Events. In L. Foster, & J. W. Swanson (Eds.), *Experience and Theory* (pp. 79–101), Amherst: University of Massachusetts Press.
- Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.
- Dennett, D. C. (1991). Real Patterns. *Journal of Philosophy*, 88(1), 27–51.
- Dorr, C. Hawthorne, J. (2013). Naturalness. In K. Bennett, D. W. Zimmerman (Eds.), *Oxford Studies in Metaphysics* (pp. 3–77). Oxford: OUP.
- Elqayam, S., Evans, J. St. B. T. (2011). Subtracting “Ought” From “Is”: Descriptivism Versus Normativism in the Study of Human Thinking. *Behavioral and Brain Sciences*, 34(5), 233–248.
- Engel, P. (2018). The Doxastic Zoo. In A. Coliva, P. Leonardi, S. Moruzzi (Eds.), *Eva Picardi on Language, Analysis and History* (pp. 297–316). London: Palgrave Macmillan.
- Fodor, J. A. (1975). *The Language of Thought*. New York: Cromwell.
- Føllesdal, D. (1958). *Husserl und Frege: Ein Beitrag zur Beleuchtung der Entstehung der phänomenologischen Philosophie* [Husserl and Frege: A Contribution to Elucidating the Origins of Phenomenological Philosophy]. Oslo: Aschehoug.
- Frege, G. (1956). The Thought: A Logical Inquiry. *Mind*, 65(259), 289–311.
- Frege, G. (2007). *Foundations of Arithmetic*. New York: Routledge.
- Gigerenzer, G. (1993). The Bounded Rationality of Probabilistic Mental Models. In K. I. Manktelow, D. E. Over (Eds.), *Rationality: Psychological and Philosophical Perspectives* (pp. 284–313). London: Routledge.
- Gigerenzer, G. (2000). *Adaptive Thinking*. New York: Oxford University Press.
- Gigerenzer, G. (2015). *Simply Rational: Decision Making in the Real World*. Oxford: Oxford University Press.
- Gray, J. (2008). *Plato’s Ghost: The Modernist Transformation of Mathematics*. Princeton: Princeton University Press.
- Hahn, U. (2025). Human Rationality. *Annual Review of Psychology*, 77, 193–221.

- Harman, G. (1984). Logic and Reasoning. *Synthese*, 60(1), 107–127.
- Helton, G. (2020). If You Can't Change What You Believe, You Don't Believe It. *Noûs*, 54, 501–526.
- Husserl, E. (2001). *Logical Investigations* (Vol. 1). New York: Routledge.
- Knauff, M., Spohn, W. (2021). Psychological and Philosophical Frameworks of Rationality—A Systematic Introduction. In M. Knauff, W. Spohn, *The Handbook of Rationality* (pp. 18–84). Cambridge, MA; London: The MIT Press.
- Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61(4), 343–377.
- Lewis, D. (1984). Putnam's Paradox. *Australasian Journal of Philosophy*, 62(3), 221–236.
- Matthews, R. J. (2013). Belief and Belief's Penumbra. In N. Nottelmann (Eds.), *New Essays on Belief* (pp. 100–123). London: Palgrave Macmillan.
- McKittrick, J. (2005). Are Dispositions Causally Relevant? *Synthese*, 144, 357–371.
- Nozick, R. (1993). *The Nature of Rationality*. Princeton, NJ: PUP.
- Oaksford, M., Chater, N. (2007). *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: Oxford University Press.
- Posłajko, K. (2024). *Unreal Beliefs: An Anti-Realist Approach in the Metaphysics of Mind*. London: Bloomsbury.
- Putnam, H. (1979). What is Mathematical Truth. In *Mathematics Matter and Method: Philosophical Papers* [Volume 1, 2nd Edition, pp. 60–78]. Cambridge: Cambridge University Press.
- Quine, W. V. O. (1980a). On What There Is. In *From a Logical Point of View* [2nd Edition, pp. 1–19]. Cambridge, MA: Harvard University Press.
- Quine, W. V. O. (1980b). Two Dogmas of Empiricism. In *From a Logical Point of View* [2nd Edition, pp. 20–46]. Cambridge, MA: Harvard University Press.
- Schurz, G., Hertwig, R. (2019). Cognitive Success: A Consequentialist Account of Rationality in Cognition. *Topics in Cognitive Science*, 11(1), 1–30.
- Shah, N., Velleman, J. D. (2005). Doxastic Deliberation. *The Philosophical Review*, 114(4), 497–534.
- Simon, H. A. (1990). Invariants of Human Behavior. *Annual Review of Psychology*, 41, 1–19.
- Stanovich, K. E. (1993). Dysrationalia: A New Specific Learning Disability. *Journal of Learning Disabilities*, 26(8), 501–515.
- Stanovich, K. E., West, R. F. (2000). Individual Differences in Reasoning: Implications for the Rationality Debate? *The Behavioral and Brain Sciences*, 23(5), 645–726.
- Stenning, K., Lambalgen, M. van. (2008). *Human Reasoning and Cognitive Science*. Boston: MIT Press.
- Sturm, T. (2012). The “Rationality Wars” in Psychology: Where They Are and Where They Could Go. *Inquiry*, 55(1), 66–81.
- Tversky, A., Kahneman, D. (1974). Judgments Under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131.

- Wason, P. C. (1968). Reasoning About a Rule. *Quarterly Journal of Experimental Psychology*, 20(3), 273–281.
- Wundt, W. (1880/1883). *Logik. Eine Untersuchung der Prinzipien der Erkenntnis und der Methoden wissenschaftlicher Forschung* [Vols. 1, 2; Logic: An Investigation Into the Principles of Knowledge and the Methods of Scientific Research]. Stuttgart: Enke.
- Zangwill, N. (2005). The Normativity of the Mental. *Philosophical Explorations*, 8(1), 1–19.

ANNEMARIE KALIS \*

## THE FOLK ARE DOING FINE (BUT THE PHILOSOPHERS ARE IN TROUBLE)

**SUMMARY:** In his book *Unreal Beliefs: An Anti-Realist Approach in the Metaphysics of Mind* (2024b), Krzysztof Posłajko proposes to separate the question whether beliefs exist from the question whether beliefs are real, and defends the view that whereas beliefs exist, they should not be considered real. He goes on to argue that his anti-realist position implies that our folk psychological understanding of belief is in need of revision, as this common sense understanding at least implicitly builds on an unfounded commitment to realism. In this commentary I raise some objections to the latter claim. Specifically, I question the assumption that common sense notions of believing in Western folk psychology is grounded in specific metaphysical assumptions about mental states. Also, I will review anthropological work in various cultural and linguistic communities showing that there exist many practices of interpersonal understanding in which the ascription of inner mental states hardly plays a role. My claim is that the folk are doing fine: folk psychological practices are multifaceted and fluid phenomena that serve a wide variety of goals. I conclude by arguing that the ones in trouble might be the philosophers, as Posłajko's arguments have a lot of critical potential against causalist accounts in the philosophy of mind.

**KEY WORDS:** folk psychology, anti-realism, mental states, Ryle, anthropology.

---

\* Utrecht University, Department of Philosophy and Religious Studies. E-mail: A.Kalis@uu.nl. ORCID: 0000-0001-8574-492X.

## 1. Introduction

When someone says: “I believe that a new pandemic is just around the corner”, is that person thereby reporting the existence of a specific inner entity, namely a belief state? Are there reasons to presuppose that such things as beliefs really exist—and if they do, how should we understand them? Questions like these have gained renewed traction in recent years due to the rise of various novel and creative philosophical approaches to the metaphysics of mental states (Demeter, Parent, Toon, 2022; Fernández Castro, 2023; 2024; Gozzano, 2019; Kalis, Ghijsen, 2022; Mölder, 2010; Pośłajko, 2022; 2024a; Schwitzgebel, 2002; Toon, 2023). For years, Krzysztof Pośłajko’s work has been an important voice in this debate, and it is a pleasure to see that his contributions have now culminated in his thought-provoking and carefully argued book *Unreal Beliefs: An Anti-Realist Approach in the Metaphysics of Mind* (2024b).

Pośłajko’s main claim in the book is that beliefs exist because the application criteria for the term “belief” are superficial, and can therefore easily be met. However, according to Pośłajko the question whether beliefs exist should be separated from the question whether beliefs are real. He argues that the latter question should be answered in the negative. So, whereas beliefs exist, they should not be considered real: this makes his position anti-realist. His main argument for anti-realism is that belief ascriptions do not track properties that can be understood as (sufficiently) natural, using David Lewis’ distinction between natural and non-natural properties (Lewis, 1983). Subsequently, Pośłajko claims that his anti-realist position implies that our folk psychological understanding of belief is in need of at least partial revision, as our common sense understanding at least implicitly builds on an unfounded commitment to realism.

Pośłajko builds a solid and innovative argument for anti-realism. His account will add much needed impetus and weight to the broader anti-realist perspective, the presence of which in the literature has so far remained somewhat fragmented. Distinguishing the question of existence from the question of realism is an interesting move that creates conceptual room for a novel understanding of mental states, and Pośłajko does a great job in showing the potential of such an understanding. At some points, the distinction also creates difficulties. Specifically, it is not always entirely clear what, in Pośłajko’s framework, the question of realism is about precisely. Early in the book Pośłajko proposes that “the distinction between realism and moderate anti-realism about beliefs should be read in terms of the distinction between natural and non-natural properties” (2024b, p. 64). But later he claims that

if the folk consider the notion “*X*” to denote a superficial/insubstantial category, and the property *X* that is denoted by this notion turned out to be a non-natural one, then the only thing we should conclude is that the folk concept adequately renders reality. This looks very much like realism. (2024b, p. 116)

But why would this count as realism? If we follow Pośłajko's own suggestion that realists take mental states to have natural properties, then everyone who holds that the notion of belief denotes non-natural properties, should count as an anti-realist. In response to this objection, Pośłajko claims that

adherence to the claim that belief-properties are not natural in Lewis's sense is not in itself enough to classify a position as being a specimen of belief anti-realism. What is also needed is some sort of commitment to the idea that the folk notion of belief is in the wrong with respect to mental reality. (2024b, p. 117)

But why would this be an additional requirement? Whether or not beliefs are real seems a question that is distinct from the question whether or not the folk are correct about the reality of mental states.

This latter question is the one I would like to discuss in more detail in the remainder of this paper. Specifically, I will take a closer look at Pośłajko's claim that "there is a serious mistake ingrained in the folk-psychological concept of belief" (p. 136), a mistake which he specifies as the mistake to at least implicitly attribute "internal, causally active states to subjects of belief" (p. 127). In the next section I will question the assumption that our common sense notion of believing does, even if only implicitly, assume a causalist theory of action explanation. In contrast, I will argue that in so far as the folk psychological discourse most Western academics will be familiar with embraces causalism, this seems a highly minimal and pragmatic form of causalism which is not grounded in specific metaphysical assumptions. I argue that more "substantive" interpretations of folk discourse actually smuggle quite some philosophical baggage into the story. Building on this suggestion, I will then take a step back and show that there are actually many different folk psychological practices, and that these are multifaceted and dynamic phenomena regulated by a wide variety of goals and norms. By reviewing anthropological work in various cultural and linguistic communities I will show that there exist many practices of interpersonal understanding in which the ascription of inner mental states hardly plays a role. Folk psychology is thus not a uniform practice, and the folk are not always nor mainly concerned with making metaphysical statements. I will conclude that the folk are doing fine: the ones in trouble might be the philosophers, as Pośłajko's arguments actually have a lot of critical potential against causalist accounts in the philosophy of mind.

The suggestions and questions I raise therefore do absolutely nothing to downplay the importance, quality and creativity of Pośłajko's views. Studying his book has been a highly inspiring and rewarding experience, and I consider it to be one of the most important contributions to the philosophy of mental states that have appeared in the last decade.

## 2. A Causalist Folk Psychology?

Posłajko's arguments for the claim that our folk psychological concept of belief is at least partly mistaken, are mostly developed in Chapter Seven (*Minimal Non-Realism and Common Sense*). At the beginning of that chapter, he acknowledges that his notion of "folk" is a parochial one: what he refers to is "the common sense of the contemporary, educated, western person" (p. 118). In this section, I will follow him in focusing on the common sense understandings of mental states that Western philosophers will generally most familiar with (in the next section, I will take a step back from the parochiality). Posłajko's claim is that the Western folk notion of believing is implicitly causalist, and mistakenly so. As a method Posłajko relies on conceptual analysis, integrating two sources of evidence: firstly, the observation of ordinary language, and secondly the mindreading literature in experimental psychology.

Regarding the observation of ordinary language, he builds on the analysis of Frances Egan (1995). Egan defends a minimalist construal of folk psychology, arguing that our common sense understanding of mental states involves the assumption that our mental states cause our behaviour:

We often say things such as "He believed he was about to be fired because he saw a confidential memo that criticized his job performance" and "She quit smoking because she believed it was affecting her health". There is no reason to suppose that "because" here functions any differently than in locutions which are clearly causal, such as "The fire started because the electrical system was overloaded". (Egan, 1995, p. 187)

According to Posłajko, this supports his claim that our common sense understanding of beliefs is up for revision: after all, in the preceding chapters he has argued that beliefs should not be understood as causally active states. However, I think this conclusion is too quick. In fact, Egan also argues that this commitment to causalism does not presuppose nor rely on any specific metaphysical justification, nor does it entail a commitment to any specific cognitive architecture such as a language of thought or a connectionist model (pp. 186–187). If Egan is right about our folk psychological discourse not being grounded in identifiable metaphysical commitments, then an anti-realist metaphysical position will do nothing to undermine the legitimacy of that discourse. My proposal (which I think comes quite close to Egan's) is to consider folk discourse as a discourse that operates largely independently of, but is in fact compatible with, a wide variety of scientific and metaphysical positions.<sup>1</sup>

---

<sup>1</sup> Of course, any statement that claims the existence of a causal relation is, in a sense, a metaphysical statement. However, my point is when the folk make such statements, they generally do not make such claims on the basis of any identifiable metaphysical theory or definition of causation.

But, one might ask, how could such a common sense “minimalist causalism” be compatible with Pośłajko’s anti-realism? The most plausible answer is that the folk use causation terms in a much looser and less specific way than philosophers do, a suggestion for example also made by Curry (2018). Pośłajko considers this possibility and discusses Curry’s arguments (2024b, p. 120). However, he retorts by saying that he is not convinced: “No matter how we spin it, it seems there is no deep difference in the way causal vocabulary functions in the context of psychological explanations and other types of explanations” (p. 120). I do not share Pośłajko’s intuitions here. Given that even philosophers amongst themselves already disagree substantially on how to understand causation, and given that causation does not seem to be a uniform concept in the first place (Cartwright 2004; 2006), do we really have grounds to suppose that the folk all have a similar understanding of causation, which is moreover the same understanding that philosophers have? Terms like “cause”, “because”, etc. are indeed *used* in everyday life to explain behaviour—but this is not the same as saying that the users of these terms commit themselves to any particular metaphysical claim or theory. In contrast, I want to maintain that even if folk psychological discourse is used to predict and explain, and even if notions like “because” and “inner cause” play a role in such predictions and explanations, they often seem to do so in a very “light” and pragmatic sense.

This interpretation of folk psychological discourse has been extensively defended by Gilbert Ryle. For example, in his paper *On Systematically Misleading Expressions* Ryle argues that if people say things like “honesty does so and so” or “equality is such and such”, we should not take them to hold, even implicitly, that honesty and equality are entities. As he states:

the plain man who uses such quasi-Platonic expressions is not making a philosophical mistake. He is not philosophizing at all. [...] He knows what he means and will, very likely, accept our more formally proper restatement of what he means as a fair paraphrase, but he will not have any motive for desiring the more proper form of expression, nor even any grounds for holding that it is more proper. (Ryle, 1971, p. 48)

I hold that this is exactly the attitude with which people use expressions like “She quit smoking because she believed it was affecting her health”. In his book, Pośłajko does acknowledge that the folk use belief terminology in a looser way than a causalist conception of belief seems to require. As explanation he points to Haslanger’s distinction between manifest and operative concepts: whereas the folk report a causalist, “metaphysically demanding” understanding of belief (this is their manifest concept), their actual use of the concept (their operative concept) can be much looser and less “metaphysical” (Pośłajko, 2024b, p. 132). According to Pośłajko, this implies that at least the folk’s manifest concept is up for revision. Even though I agree that applying Haslanger’s distinction here raises interesting questions, I am not really convinced that most folk have a metaphysically robust manifest concept of belief. Instead, I suspect that the causalist

conception is primarily held by philosophers and scientists, something which becomes more clear when looking at Posłajko's second source of evidence.

Here Posłajko focuses on the mindreading literature in the empirical sciences of the mind: a lot of experimental psychology and cognitive science literature implicitly assumes a causalist framework. This is certainly true, but I am not convinced that this fact actually strengthens his revisionist claim about folk psychology. Causalism as a background assumption in psychology and cognitive science is part of an academic, theoretical discourse, and as such it can easily be explained (as Curry also points out) as a form of "internalized philosophical dogma" (Curry, 2018, p. 122): psychology and cognitive science share their functionalist cognitivist theoretical background with the analytic philosophy of mind. Egan makes a similar suggestion:

In the twenty years since the publication of Fodor's seminal book (1975) the classical computational model of the mind has become the received view in the philosophy of mind. It is not surprising that integral components of this conception [...] have seeped into the collective philosophical consciousness to the point where it may be difficult for philosophers to separate these components from the body of theory that they share with ordinary folk. But they should be separated. (Egan, 1995, p. 195)

I agree with Egan here. For philosophers who have been born and raised with causalist intuitions about the mental, it might seem obvious that this is how people understand the mind. Moreover, of course insights and dogmas from philosophy and psychology do really find their way into folk psychological discourse: it is now for example quite common to say that one is feeling sad "because one's serotonin levels are low", whereas a hundred years ago that would have been unintelligible. However, even if academic and non-academic discourses cannot be neatly separated, this does not mean that folk psychological expressions about causation entail a commitment to any of the theoretical assumptions that are at stake in the academic discourses from which such expressions are derived.

So, to conclude, even though some folk in Western cultures might, under the influence of philosophical and scientific discourses, have causalist intuitions, this cannot be considered a core commitment of "Western folk" in general. I will now take a step back from the folk psychological discourse most Western academics will be familiar with, and show that far from being a uniform phenomenon, folk psychological practices actually come in all shapes and sizes, and are regulated by a wide variety of goals and norms.

### **3. Folk Practices of Believing**

As said, Posłajko is laudably transparent in acknowledging that he uses a somewhat "parochial" notion of folk psychology; moreover, he is well aware that "folk-psychological categories vary cross-culturally, and that the category of belief is not a universal one" (2024b, p. 96). In this section, I will argue that such

insights on cultural variation might have more substantial impact on his arguments than Poślajko is aware of. I will raise doubts about the idea that folk psychology is a uniform phenomenon; if it is not, this entails it might be impossible to make any meaningful claims about folk psychology as such. In support of my argument, I will discuss some examples of recent empirical work that investigates cultural and linguistic differences in practices of mental state ascription. I will argue that this work suggests that a) there are various different folk-psychological conceptions of believing and b) that several folk conceptions of believing do not assume that mental states cause behaviour, not even in the minimal sense described above.

Firstly, studies in comparative linguistics have investigated the use of mental state terms across different languages. Wierzbicka (1997; 2018) did pioneering work in this area, and proposed a so-called *natural semantic metalanguage*, consisting of a list of semantic primes that were found in every language studied. Amongst other things, this showed that whereas some notion of for example knowledge and desire might be universal (or at least were found in all the languages included in the study), the notions of belief and intention clearly are not. Gladkova (2007; 2016) builds on Wierzbicka's work in studying the use of what we know as propositional attitudes in various languages. More specifically, she further develops Wierzbicka's comparison between the English verb *to believe that* and the Russian related verb *ščitat'* (Gladkova, 2016). Even though both terms roughly refer to a mental state verb that expresses an opinion to which the speaker is committed and for which they have grounds, the meaning of both terms differs substantially. For example, Gladkova argues that whereas *to believe that* involves some willingness to provide arguments for one's opinion and openness to change if others provide good counterarguments, this is not part of the meaning of *ščitat'*. On the other hand, whereas *ščitat'* refers to opinions that one has taken substantial time to develop, this temporal feature is not part of the meaning of *to believe that*. Gladkova argues that examples like these show that our use of propositional attitude terms is infused by what she calls cultural scripts. Folk psychological practices of mental ascription should not be seen as isolated activities, but as being part of broader cultural and communicative practices.

This brings me to the proposal that if we want to understand folk psychology in their broader cultural and social context, this might require anthropological methods. In fact, there is a recent surge in anthropological literature on the question how different communities talk about, and ascribe, mental states to each other and themselves (Duranti, 2015; Luhrmann, Astuti, Robbins, 2011; Nuckolls 2017; Robbins, Rumsey, 2008). More specifically, there is an interesting body of work on the notion of belief. Already in 1973, anthropologist Rodney Needham (1973) published a book in which he argued that it is problematic to ascribe belief states to people studied in the context of anthropological fieldwork. His objections were that so-called belief states do not have any "bodily concomitants", no clear phenomenological quality nor are easily translatable into other lan-

guages. All this, according to Needham, showed that the concept of belief is a “cultural artifact” and cannot be used to provide information on the nature of the human mind. Needham’s conclusion nicely resonates with Pośłajko’s perspective on mental state ascription. Recent contributions in anthropological literature picking up on Needham’s work (Risjord, 2020; Streeter, 2020) argue however that Needham failed to see that the notion of belief as a representational inner state has its source in philosophy and not in everyday life: “Anthropology inherited its commitment to representationalism from the 19th-century debates over the possibility of a human science” (Risjord, 2020, p. 589). The lesson, according to these authors, is that it might be impossible to do anthropology without ascribing belief states to others in some minimal sense (Streeter, 2020), but that anthropologists should refrain from projecting a philosophical conception of belief and instead be open to the many possible ways in which cultures and practices might understand minds.

As examples of such an open investigation, I want to briefly reflect on some observations on the role of mental state ascription in various practices as discussed in the ethnographic studies of Rosen (1995) and Carey (2017). Both point out that there are communities that largely refrain from ascribing inner mental states to one another, a phenomenon referred to as *opacity of mind*. Rosen’s edited volume explores the various roles played by the concept of intention in different cultural and communicative practices. In his introduction, he observes that certain communities that live under conditions of political oppression

have restricted the attribution of inner selves to the domain of intracommunity communication. Others, more poignantly, act as if they have no inner lives of their own as a way of preserving their culture against the intrusion of hostile powers. (Rosen, 1995, p. 8)

In a similar vein, Carey (in his ethnography of a group of communities in the Moroccan High Atlas) observes that these communities generally explain or interpret each other’s behaviour by referring to observations of other behaviour or bodily appearances (“he is unshaven”), and not to inner mental states such as beliefs or intentions. According to Carey, this does not mean that these communities do not have a notion of an inner mental life: “it is quite impossible to interpret the meaning of even quite simple sentences unless we in some way impute intention to the speaker” (2017, p. 47). However, the important message of his study is that such imputation need not (and in this case, does not) take the form of the presupposing of the existence of representational inner states.

These examples show that folk psychological practices come in many forms, and in many cases do not match the representationalist picture Western cultures have inherited from philosophical discourse. The same seems to hold for the picture that our folk psychological explanations are causalist in nature. For example, Henrich suggests that:

The pressure for an acceptable, clear, and explicit reason for doing things is merely a social norm common in Western populations, which creates the illusion (among Westerners) that humans generally do things based on explicit causal models and clear reasons. They often do not. (Henrich, 2016, p. 129)

Such observations raise the question: which aspects of folk psychology are actually in need of revision? Is it the practice itself, or might it be merely the representationalist, causalist inheritance that can be found in some Western practices? I want to defend the latter claim. All in all, the idea that beliefs are causally efficacious internal mental states might have to be seen as a background dogma in Western analytic philosophy (Strijbos, de Bruin, 2013). As mentioned, Pośłajko is well aware of this work, and actually uses evidence on cultural variation himself to support his position that the properties of belief are non-natural (pp. 96–97). My point is that such evidence *also* raises doubts about his own claim that the folk have mistaken assumptions about the metaphysics of believing. Instead, I want to claim that the folk are doing fine: folk psychological practices are multifaceted and fluid phenomena that serve a wide variety of goals and that are governed by a wide variety of norms (Curry, 2020; McGeer, 2007; 2020). As mentioned in the previous section, Pośłajko is correct in pointing out that certain Western folk psychological practices seem influenced by representationalist and causalist intuitions. But this does not mean that “the Western folk” commit to causalism. Therefore, the fact that causalist intuitions might be philosophically objectionable does not show that there is anything wrong with folk psychological practices. What it does show is that there is something wrong with the philosophical positions that gave rise to causalist intuitions: the ones in trouble might thus be the philosophers themselves.

#### 4. Conclusions

To summarize, Pośłajko has made a convincing case for the idea that beliefs are not real, and at the same time can be said to exist because the application criteria for the term “belief” are superficial. However, if the folk are not mistaken, according to his own requirements (pp. 116–117) this is a problem for his anti-realist case. After all, Pośłajko claims that anti-realism involves the claim that the folk are somehow mistaken about the reality of mental states. In the introduction I already raised some doubts regarding this criterion, but if we follow Pośłajko here then my claim that “the folk are doing fine” might have the implication that Pośłajko’s position becomes a lot closer to the kind of deflationary realism defended by neo-Ryleans and some interpretivists (pp. 37–38).

However, a maybe unexpected bonus is that Pośłajko’s book makes a highly convincing case for revisionism about causalism in the philosophy of mind: philosophers embracing the position that beliefs are causally effective representational states *are* in trouble. Even though the philosophy of mind as a discipline has become much more diverse and dynamic over the last decades, the representationalist-causalist framework is still highly influential, and classic notions like

the language of thought hypothesis seem to experience a revival in the cognitive sciences (Kazanina, Poeppel, 2023; Quilty-Dunn, Porot, Mandelbaum, 2023). What Posłajko's solid arguments therefore call for, in my view, is not so much conceptual engineering of folk psychological discourse, but renewed philosophical debate about the metaphysics of mind.

## REFERENCES

- Carey, M. (2017). *Mistrust: An Ethnographic Theory*. Hau Books.
- Cartwright, N. (2004). Causation: One Word, Many Things. *Philosophy of Science*, 71(5), 805–819.
- Cartwright, N. (2006). Where Is the Theory in Our “Theories” of Causality? *The Journal of Philosophy*, 103(2), 55–66.
- Curry, D. S. (2018). Beliefs as Inner Causes: The (Lack Of) Evidence. *Philosophical Psychology*, 31(6), 850–877.
- Curry, D. S. (2020). Interpretivism and Norms. *Philosophical Studies*, 177(4), 905–930. <https://doi.org/10.1007/s11098-018-1212-6>
- Demeter, T., Parent, T., Toon, A. (2022). *Mental Fictionalism: Philosophical Explorations*. Routledge.
- Duranti, A. (2015). *The Anthropology of Intentions*. Cambridge University Press.
- Egan, F. (1995). Folk Psychology and Cognitive Architecture. *Philosophy of Science*, 62(2), 179–196.
- Fernández Castro, V. (2023). Folk Psychology Without Metaphysics: An Expressivist Approach. *South African Journal of Philosophy*, 42(2), 128–143. <https://doi.org/10.1080/02580136.2023.2273122>
- Fernández Castro, V. (2024). An Expressivist Approach to Folk Psychological Ascriptions. *Philosophical Explorations*, 27(1), 86–105. <https://doi.org/10.1080/13869795.2023.2251491>
- Gladkova, A. (2007). Universal and Language-Specific Aspects of “Propositional Attitudes”. In A. C. Schalley, D. Khlentzos (Eds.), *Mental States: Volume 2: Language and Cognitive Structure* (pp. 61–84). John Benjamins B. V.
- Gladkova, A. (2016). Propositional Attitudes and Cultural Scripts. In A. Capone, J. L. Mey (Eds.), *Interdisciplinary Studies in Pragmatics, Culture and Society* (Vol. 4, pp. 329–352). Springer International Publishing. [https://doi.org/10.1007/978-3-319-12616-6\\_12](https://doi.org/10.1007/978-3-319-12616-6_12)
- Gozzano, S. (2019). Beliefs as Mental States: Steps Toward a Dispositional Approach. *Rivista Di Filosofia*, 110(2), 205–224.
- Haslanger S. (2012). *Resisting Reality: Social Construction and Social Critique*. New York: Oxford University Press.
- Henrich, J. (2016). *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton University Press.

- Kalis, A., Ghijsen, H. (2022). Understanding Implicit Bias: A Case for Regulative Dispositionalism. *Philosophical Psychology*, 35(8), 1212–1233.  
<https://doi.org/10.1080/09515089.2022.2046261>
- Kazanina, N., Poeppel, D. (2023). The Neural Ingredients for a Language of Thought Are Available. *Trends in Cognitive Sciences*, 27(11), 996–1007.  
<https://doi.org/10.1016/j.tics.2023.07.012>
- Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61(4), 343–377. <https://doi.org/10.1080/00048408312341131>
- Luhrmann, T., Astuti, R., Robbins, J. (2011). Toward an Anthropological Theory of Mind. *Suomen Antropologi: Journal of the Finnish Anthropological Society*, 36(4), 5–69.
- McGeer, V. (2007). The Regulative Dimension of Folk Psychology. In D. D. Hutto, M. Ratcliffe (Eds.), *Folk Psychology Re-Assessed* (pp. 137–156). Springer.
- McGeer, V. (2020). Enculturating Folk Psychologists. *Synthese*, 199(1/2), 1–25.
- Mölder, B. (2010). *Mind Ascribed: An Elaboration and Defence of Interpretivism* (Vol. 80). John Benjamins Publishing.
- Needham, R. (1973). *Belief, Language and Experience* (First Edition). Univ. of Chicago Pr.
- Nuckolls, C. W. (2017). Cognitive Anthropology. In W. Bechtel, G. Graham (Eds.), *A Companion to Cognitive Science* (pp. 140–145). Blackwell Publishing.
- Posłajko, K. (2022). How to Think about the Debate over the Reality of Beliefs. *Review of Philosophy and Psychology*, 13(1), 85–107.  
<https://doi.org/10.1007/s13164-021-00551-8>
- Posłajko, K. (2024a). Expressivism, Inferentialism, and the Status of Attitudes. *Inquiry*, 67(1), 171–198. <https://doi.org/10.1080/0020174X.2021.1886982>
- Posłajko, K. (2024b). *Unreal Beliefs: An Anti-Realist Approach in the Metaphysics of Mind*. Bloomsbury Publishing.
- Quilty-Dunn, J., Porot, N., Mandelbaum, E. (2023). The Best Game in Town: The Reemergence of the Language-Of-Thought Hypothesis Across the Cognitive Sciences. *Behavioral and Brain Sciences*, 46.
- Risjord, M. (2020). Anthropology Without Belief: An Anti-Representationalist Ontological Turn. *Philosophy of the Social Sciences*, 50(6), 586–609.
- Robbins, J., Rumsey, A. (2008). Introduction: Cultural and Linguistic Anthropology and the Opacity of Other Minds. *Anthropological Quarterly*, 81(2), 407–420.
- Rosen, L. (1995). *Other Intentions: Cultural Contexts and the Attribution of Inner States*. School of American Research Press.
- Ryle, G. (1971). *Collected papers: Vol. Volume II: Collected Essays 1929–1968* (Ed. 2009). Routledge.
- Schwitzgebel, E. (2002). A Phenomenal, Dispositional Account of Belief. *Noûs*, 36(2), 249–275.
- Streeter, J. (2020). Should We Worry About Belief? *Anthropological Theory*, 20(2), 133–156.

- Stribos, D. W., de Bruin, L. C. (2013). Universal Belief-Desire Psychology? A Dilemma for Theory Theory and Simulation Theory. *Philosophical Psychology*, 26(5), 744–764. <https://doi.org/10.1080/09515089.2012.711034>
- Toon, A. (2023). *Mind as Metaphor: A Defence of Mental Fictionalism*. Oxford University Press.
- Wierzbicka, A. (1997). *Understanding Cultures Through Their Key Words: English, Russian, Polish, German, and Japanese*. Oxford University Press.
- Wierzbicka, A. (2018). I Know: A Human Universal. In S. Stich, M. Mizumoto, E. McCready (Eds.), *Epistemology for the Rest of the World* (pp. 215–250). Oxford University Press.

JOANNA KOMOROWSKA-MACH \*

## RETHINKING FOLK REALISM ABOUT BELIEFS

**SUMMARY:** Poślajko's *Unreal Beliefs* (2024) provides a lucid overview of competing ontological views on belief, and his minimal anti-realist proposal is well-argued and persuasive. This article calls into question a less central thesis of the book: that ordinary speakers see beliefs as concrete, real, internal causes of behaviour. I offer a plausible alternative explanation of why everyday discourse gives this impression. While endorsing Poślajko's core anti-realist insights, I show that folk attributions of belief arise from a communicative need to translate holistic, nonverbal mindreading into discrete linguistic units. This translation does not reflect a genuine ontological commitment; it simply facilitates the exchange of information about one's own and others' perspectives. Philosophical reconstruction then mistakes these linguistic shortcuts for evidence of folk realism, leading to a theoretical ontology that exceeds the minimal metaphysical stance of everyday speakers. By distinguishing between holistic mindreading, language-driven segmentation, and academic reification, I propose a stance aligned with minimal anti-realism that does not aim to decide the question so much as to move the burden of proof.

**KEYWORDS:** belief, folk psychology, anti-realism, mindreading, belief attribution.

### 1. Introduction

I read Krzysztof Poślajko's book *Unreal Beliefs* (2024) with great interest. In a clear and well-structured manner, the author systematises the various positions on the ontological status of beliefs and carves out a space for a compromise

---

\* University of Warsaw, Faculty of Philosophy. E-mail: j.komorowska-mach@uw.edu.pl.  
ORCID: 0000-0002-8287-6668.

anti-realist approach that, in my view, holds strong appeal for those with a naturalistic outlook on the mind. Since I agree with Pośłajko on the fundamental issues, including his claim that beliefs are not real in the sense characterised in the book, I would like to draw attention to a less central point: the way Pośłajko interprets the implicit metaphysical assumptions embedded in folk psychology's concept of belief.

In Chapter 7, Pośłajko addresses a potential objection to his overall approach: if one accepts that belief properties are not natural, this does not necessarily make one an anti-realist. Instead, one might claim that beliefs are real but merely superficial, along the lines of Schwitzgebel's dispositionalism (2013) or Dennett's interpretationism (2009). If we also accept that the folk concept of belief is superficial, then folk psychology would be correct about beliefs, and there would be no need to adopt anti-realism. (Pośłajko, 2024, p. 116) To respond to this objection, Pośłajko seeks to show that folk speakers do treat beliefs as "concrete causes of behaviour". He argues that both ordinary language practice and psychological research support this view (Pośłajko, 2024, p. 118).

I agree with both Pośłajko's minimal anti-realism and his claim that folk psychology should not be eliminated. I also share his scepticism about treating the folk concept of belief as merely superficial. However, I argue that, at the same time, we are not obliged to accept the claim that the folk treat beliefs as "concrete causes of behaviour" or the conclusion that they are genuinely realists about beliefs. I aim to show that a different, non-realist reading of folk practice is sufficiently plausible.

My criticism is not rooted in suspicion of the folk commitment to mental causation. It is, rather, that we must be cautious in inferring robust ontological commitments from linguistic packaging and explanatory convenience. The folk need not be read as treating beliefs as real, concrete entities, even though the language of folk psychology, as philosophers reconstruct it, suggests otherwise. I explain how such a stance both differs from classical belief realism and escapes superficialism. I also show that adopting such a perspective does not necessarily undermine the minimalist anti-realism developed in Pośłajko's book.

## 2. Are We Realists About Beliefs?

What is beyond dispute is that we all assume people can believe and that believing has some causal connections with a person's behaviour and other mental states. We readily produce sentences like, "She quit smoking because she believed it was hurting her health," or "He turned left because he believed that the other road was closed". If you ask people, "Are beliefs real?" they will answer "Yes" without hesitation. It seems natural to interpret this as folk-belief realism, and such an interpretation appears to be widespread among philosophers. Pośłajko aligns with this trend in arguing that the folk notion of belief is "substantial, at least in the respect that, according to our folk notion, beliefs should be seen as concrete causes of behaviour" (Pośłajko, 2024, p. 118).

Curry (2018), following Ryle (1949), develops one line of objection to this view. He argues that explanations in terms of beliefs appeal to reasons rather than to causes of behaviour. From this perspective, people do not think of beliefs as genuine causes of behaviour; they use belief-talk to contextualise behaviour in terms of behavioural dispositions, that is, on a superficial level. Belief-talk is supposed to give “teleological explanations of actions and insights into the personalities of actors”, not to provide mechanistic explanations of behaviour (Curry, 2018, p. 20). Pośłajko disagrees with this position. He claims that, although some argue that folk-psychological explanations invoke reasons rather than causes, there is no substantive difference in how causal language functions in folk psychology compared to other explanatory domains and no credible alternative to the view that belief attributions often serve straightforward causal purposes (Pośłajko, 2024, p. 120).

Without attempting to resolve their debate, I want to outline a different line of objection to the claim that people regard beliefs as concrete internal causes of behaviour, one that lies somewhat outside the exchange between Curry and Pośłajko. I agree with Pośłajko that the folk are realists about the mental: they do refer to mental states and processes by means of folk psychological language, and they genuinely believe that the mental sphere plays a causal role (which sets my position apart from dispositionalism and other forms of superficialism). At the same time, I question Pośłajko’s claim that the folk are belief-realists in the sense that they treat beliefs as “concrete causes” of behaviour. Such a stance requires countenancing the possibility that ordinary usage of folk-psychological vocabulary (e.g., “belief”) does not entail a metaphysical commitment to the concrete ontology that this framework appears to suggest. In the next section, I clarify and defend this possibility.

The arguments below should be read as exploring a live possibility rather than as establishing a consensus-defeating thesis. I aim to demonstrate that the question of whether the folk are realists about beliefs is far from settled and that there is conceptual space between folk realism and superficialism about beliefs that deserves consideration.

### 3. Three Levels of Folk Psychology

Many discussions on folk psychology begin by distinguishing between two concepts, or levels, of folk psychology (Dennett, 1989; Stich, Ravenscroft, 1994). To show that it is possible to hold a view on which ordinary people are not realists about beliefs without thereby casting them as superficialists, I will propose a more fine-grained distinction between different levels of folk psychology and argue that it is, at minimum, a plausible model. First, there are holistic, context-sensitive mindreading capacities that emerge prior to language acquisition. Second, folk-psychological discourse is employed to communicate about mental states, including beliefs, and this practice can give the impression of a robust realist commitment. Finally, philosophical reconstruction systematises

that appearance into an explicit theoretical framework with a realist ontology. To be clear, in introducing this distinction I am not claiming that there are three disjoint systems or methods for attributing minds. I take them to be three salient elements or aspects of a potentially unified capacity, each building on the previous one, and distinguishable only to a limited extent for conceptual and explanatory purposes.

### **3.1. Level 1: Mindreading as a Craft**

At its most fundamental level, folk psychology is our basic ability to explain and predict others' mental states and behaviour, a craft, as Dennett puts it (1998). It integrates perceptual, affective, motivational, and contextual cues into an understanding of another's stance and appears to be automatic and universal, with evolutionary roots. There is no persuasive evidence that, at this level, we operate with theoretical concepts of folk psychology. In particular, there is no need to assume that we postulate beliefs as discrete, causally efficacious inner objects rather than more holistically understood informational states.

Positions such as Gallagher's (2006) interaction theory, Hutto and Myin's (2012) radical enactivism, and the broader 4E framework (Noë, 2004; Varela, Thompson, Rosch, 2017) have grown in influence by showing that social understanding emerges from embodied, embedded, enactive, and extended interactions rather than from the manipulation of isolated belief-tokens. Together, these accounts suggest that mindreading operates as a unified, context-sensitive capacity rather than through the attribution of "concrete" belief entities.

Cross-cultural evidence reinforces this holistic view: Strijbos and de Bruin (2013) report communities whose languages lack explicit terms for "belief" or "desire", yet whose members navigate social interactions effectively by integrating norms, roles, and embodied cues rather than by invoking propositional attitudes.

Developmental studies also support the conclusion that early mind-reading can be characterised as holistic and goal-centred, rather than dependent on the attribution of discrete belief and desire states. Classic findings show that infants encode others' behaviour in terms of goals, for example, by selectively representing the object toward which an actor repeatedly reaches, well before there is evidence that they parse psychology into separable propositional attitudes (Woodward, 1998). On what Gergely and Csibra (2003) call the teleological stance, by around 12 months, infants interpret actions as means to ends under situational constraints and evaluate efficiency without positing adult-like belief and desire representations. Even psychological interpretations of the false-belief task, which philosophers often cite as evidence for beliefs being fundamental elements of the early ontology of the mind, are often much more nuanced. Baillargeon, Scott, and He (2010) distinguish two subsystems of mindreading. On their account, Subsystem-1 enables infants to attribute two kinds of states to agents: motivational states, which specify the agent's motivation in the scene, and reality-congruent informational states, which specify what accurate information the

agent can gather through perception, memory, or inference. Subsystem-2, which becomes operational in the second year of life, extends this capacity by enabling infants to attribute reality-incongruent informational states to agents (Baillargeon, Scott, He, 2010, p. 111). Although the authors use the term “false beliefs”, their own description of SS2 indicates that the relevant attribution concerns the agent’s structured representation of the scene, rather than discrete propositional attitudes. This supports the more general point that even sophisticated forms of early mindreading need not be interpreted as involving beliefs understood as discrete, concrete entities.

Acknowledging that mindreading as a craft does not have to rely on positing beliefs as concrete objects need not imply rejecting some version of theory-theory. One can still debate whether mindreading operates more like simulation, hypothesis testing, or model-building. However, even if we employ certain theoretical concepts prelinguistically and prereflectively, there is no evidence of metaphysical commitment to beliefs as discrete entities at this level.

### 3.2. Level 2: Folk Psychological Talk and Apparent Realism

At the second level, folk psychology includes categories, generalisations, and predictions that can be formulated in natural language with its fine-grained, socially shaped concepts. The linear, reifying structure of language, stemming from how we describe macroscopic physical objects and processes, necessarily constrains how we conceptualise minds. When we say, “he believes that they do not know he cheated”, we are putting into words something that could also be grasped nonverbally (even chimpanzees are at least to some extent capable of such conceptualisation; Hare, Call, Tomasello, 2001). In communicating such observations, we simplify the complex attribution of someone’s state and perspective into a single dimension. (After saying, “he thinks that they do not know he cheated”, we might, for example, mention his emotions in a subsequent sentence, giving the impression that these are two separate “entities” in the observed scene: a belief and an emotion. We package a richer, holistic understanding of someone’s perspective, their “informational states”, into the noun-and-clause structure our language provides.

However, saying that someone “believes”, “worries”, or “considers” is not the same as assuming that there is a concrete object: “a belief”, “a worry”, or “a consideration” inside that person. The fact that we use the noun “belief” need not carry more ontological weight than the fact that we use nouns like “marathon” or “observation”. These words refer not to concrete objects but to aspects of states people may be in. Thus, we have no difficulty speaking of degrees of belief, changes of belief, or even ascribing contradictory beliefs. We may refer to several similar beliefs someone has about something, but our belief talk never addresses the question of exactly how many beliefs a person has about a particular thing. Such indeterminacy in belief-attribution supports the view that Level 2 belief-talk does not carve the mind into discrete belief-objects. Rather, it trans-

lates a more fluid, holistic grasp of another person's perspective—the kind of understanding operative at Level 1—into the available linguistic form.

To be clear, this does not entail that our belief-attribution sentences are false. Rather, they are the best linguistically available approximations of the situation we are trying to describe, akin to, for example, meteorological statements. When we say, “Today is windy”, we refer to the actual state of the weather, not merely to its superficial manifestations (e.g., gusts on the cheek). We construct causal explanations such as “branches break because it is windy” or “I have a migraine because it is windy”, yet we do not thereby postulate a concrete causal entity called “windiness”, a concrete entity that is a cause of my suffering. “Windy” serves at least partially as a placeholder: a name for whatever there is that causes trees to move, my cheeks to feel cold, and my head to hurt.

The sentence “A windy day is a cause of my headache” may be true in much the same way to the sentence “My belief that I would lose my job is a cause of my headache”. I can say it without positing concrete objects such as windiness or beliefs interacting with branches or heads.

That we so readily invoke belief-talk to explain the behaviour of systems without human minds shows how natural these attributions are and reveals the role of “belief” as a convenient placeholder whose referent is ontologically underdetermined. When we say that a moth flies toward a bulb because it “believes” it is the moon, we are not positing a discrete inner object; we are gesturing at a structured constellation of informational relations and sensitivities that guides its behaviour. The same usage extends to large language models, where talk of “beliefs” tracks stable, action-guiding regularities without committing us to belief tokens in the underlying architecture. In both cases, the notion of belief tracks a system's holistic internal organisation rather than a concrete inner item. Its referent remains underdetermined. In the absence of pressure to take sides, most speakers are neither realists nor anti-realists about beliefs and simply leave the matter undecided.

At the second level, it may seem that the folk are realists about beliefs, but this is merely an artefact of the constraints of natural language, which is geared toward talking about macroscopic, discrete objects. As Dennett puts it, the beliefs and desires “are best viewed as abstracta—more like centres of gravity or vectors than as individualizable concrete states of a mechanism” (Dennett, 1998, p. 85). The burden of proof lies with anyone who claims that these linguistic units express genuine ontological commitments. Absent further evidence, I take it to be entirely possible that the folk are neither realists nor anti-realists about beliefs, at least until philosophers prompt them to take a stance.

### 3.3. Level 3: Philosophical Reconstruction

At this third level, we philosophers and cognitive scientists take folk belief-talk at face value and reify beliefs as theoretical entities. It is only at this level that the grammar of language, combined with the powerful and, in my view,

partly justified impression that people routinely posit invisible “forces” governing human behaviour in everyday practice, leads us to metaphysical conclusions. When we encounter everyday statements such as “Her beliefs made her act like this”, we infer that folk speakers regard beliefs as postulated inner causal objects akin to theoretical objects in science. Although this reconstruction can be useful within academic discourse, there is no reason to assume that it reflects how ordinary speakers conceptualise their own or others’ mental lives. We construct our scientific (philosophical, psychological) ontology of the mind by drawing on the concepts of folk psychology—as they have developed in broadly Western cultures. We formulate and test hypotheses, and then interpret the results within this conceptual framework (for example, by seeking arguments for or against the claim that nonhuman primates attribute beliefs to their conspecifics). Our scientific picture of folk psychology is realist, a stance rightly challenged by philosophers such as Posłajko, who argues against it so effectively in his book that there is no need for me to repeat the argument here. What matters is that, unless philosophers or psychologists attempt to portray and systematise folk psychology in a realist way, ordinary people do not seem to treat these categories as seriously or as rigidly.

#### 4. Success Argument Revisited

If we consider the three-level model of folk psychology as plausible, it casts a somewhat different light on, for example, the success argument. Posłajko reconstructs it as follows (2024, p. 148):

1. Folk psychology is extremely predictively successful.
2. Beliefs are posits of folk psychology.
3. Posits of extremely predictively successful theories are real.

Therefore, beliefs are real.

In my view, this argument rests on a fallacy of equivocation. Folk psychology is indeed predictively successful as a practice (Level 1), but “beliefs” become posits of folk psychology only once they are framed in linguistic form and then also reconstructed philosophically (Levels 2 and 3). In this latter sense, folk psychology is not genuinely predictively successful; over the past hundred years, we have been unable to derive reliable psychological laws—or even stable generalisations—on its basis. Nisbett and Wilson’s classic studies (1977) provide compelling evidence that people struggle even to explain their behaviour in terms of their professed beliefs. Our actual competence lies in participating in a community in which we continuously construct and adjust fluid models of one another’s minds or, more broadly, mental states. We only occasionally attempt to capture those models in language, enabling us to communicate and coordinate, but when we do, our descriptions remain partial and approximate.

Posłajko's reply to this argument is, in some respect, similar to mine. He distinguishes between folk psychology as a craft and folk psychology as a theory (following Dennett, 1987). Unlike the account proposed here, however, he continues to attribute to the folk a realist stance about belief, whereas I take the equivocation to lie in the very concept of belief itself.

Someone who adopts anti-realism about beliefs while viewing the folk as naive realists faces the challenge of explaining how it is possible that folk-psychological concepts, if they do not reflect the metaphysical reality of the mind, can nonetheless underlie our explanatory success in understanding it. However, if we reject the idea that the so-called concepts of folk psychology, that is, the conceptual framework reconstructed from the language we use, play a significant role in that success, the problem dissolves.

### **5. Is This Superficialism?**

The position sketched here does not endorse a superficialist reduction of folk belief-talk. In my view, folk mindreading genuinely involves attributing inner states or processes to agents rather than merely labelling surface patterns. I also do not think that the folk are fictionalists about mental states such as beliefs. Non-philosophers are neither full-blown realists nor anti-realists about beliefs. However, I take them to be realists about minds and mental states understood holistically. When people engage in belief-talk, they presuppose genuine inner causes of behaviour, yet without any specific ontological commitment to beliefs as discrete entities. While Posłajko claims there is "a significant mismatch between the folk concept of belief and the 'metaphysical reality' of beliefs" (Posłajko, 2024, p. 154), I disagree. There is no significant gap between folk belief-talk and mental reality; instead, the fundamental mismatch exists between philosophical interpretations of folk belief-talk (Level 3) and the minimal metaphysical commitments that the folk might hold.

While I agree that "beliefs" function as linguistic shorthand, partial, constrained representations of deeper attributions, I do not think the folk build their theory of mind on mere behaviour and dispositions. In other words, beliefs in folk psychology are neither robust internal objects nor mere superficial dispositions; they are language-driven labels for subtler categorisations and inferences about others' minds.

### **6. Austere Conceptions of Folk Psychology**

The position advanced here may be seen as falling under the category of approaches that emphasise the austere character of folk psychology (Horgan, 1993), at least in the part where I agree that the structure of folk-psychological language is not supposed to correspond to any actual "architecture" guiding our modelling of others' minds and behaviour. However, there is a difference in that proponents of austere approaches still seem to assume that beliefs are "functionally discrete"

states that play distinct causal roles in explaining behaviour (Bogdan, 1993). By contrast, I maintain that genuine mindreading does not depend on ascribing functionally discrete beliefs, even though our language later imposes this category. More interesting from my point of view Gauker's proposal (2021), which treats attributing beliefs as a fundamentally communicative act. Gauker challenges the idea that we use belief-talk to predict behaviour: most real-world predictions arise rather from induction, social norms, or demonstrated skills, not from attributing specific beliefs. He also defends some form of belief-realism: beliefs are real only insofar as they are governed by objective normative rules (akin to money or promises), but not as causal, propositional "inner objects". This is also the point at which Posłajko disagrees with him (Posłajko, 2024, p. 154).

Gauker's position emphasises different aspects of the problem from my approach. However, some of his insights can help explain why the belief-concept in folk psychology remains firmly entrenched even if (1) beliefs are not real in Posłajko's sense and (2) ordinary people are not belief realists. At Level 1, people possess nonverbal mindreading skills that model others' perspectives holistically without segmenting their mental lives into discrete beliefs. At Level 2, belief-talk serves primarily as a communicative act, asserting or commanding on another's behalf, rather than as a straightforward explanatory or predictive tool. This rich communicative role reinforces and stabilises the folk concept of "belief", giving it a practical force in everyday interaction despite the absence of any literal, architecturally distinct belief states. In other words, the power of belief language in folk psychology stems not from cognitive architecture but from its indispensability in conveying information, corroborating testimony, and coordinating social action.

Posłajko sees austere accounts as implying that "it is the philosophers who went astray in their interpretation of the folk discourse" (Posłajko, 2024, p. 154), a sentiment I wholeheartedly share. He adds that if austere theorists "deny that the idea of mental causation is involved in our folk conception of belief, they might be accused of underestimating the strength of the folk commitments" (p. 155). I disagree: in my opinion, philosophers rather overestimate those commitments. There is no reason to assume that people base their predictions about behaviour on the postulation of concrete, discrete entities as folk-psychological language suggests.

## 7. Conclusion

I have argued that the folk do not necessarily treat beliefs as discrete, concrete causes of behaviour. Instead, I have sketched an alternative, three-level model of capacities underlying the folk ontology of belief. From this perspective, mindreading relies, at least for the most part, on modelling a person's broader perspective in a way independent of linguistic categories (Level 1). Language forces us to break our mental ascriptions into discrete chunks, including what we call beliefs; however, it does not compel us to adopt a realist ontology of beliefs

(Level 2). Philosophers and psychologists then reify this language, postulating beliefs as theoretical objects (Level 3), but that move does not reflect any genuine folk commitment. Recognising these three levels allows us to interpret apparent folk belief realism as an artefact of grammar and theory rather than an authentic stance.

Like Pośłajko, I do not offer these arguments as a “master argument” against folk belief realism; rather, my goal is to demonstrate that claims that the folk are belief-realists “lack adequate support” (Pośłajko, 2024, p. 112). My argument is intended to shift the burden of proof. I have sketched an account of the folk ontology of beliefs, which, in my view, is a sufficiently plausible alternative to Pośłajko’s view. It thereby obliges advocates of folk belief realism, who take ordinary interpretation to involve commitment to discrete inner belief entities, to supply fuller and more careful justification.

I do not rule out the possibility that the question of whether the folk are realists about beliefs may yield an empirical answer. However, conventional experimental-philosophy surveys would likely impose a philosophical conceptual framework on participants either during the experiment or when interpreting the results. Any conclusive research would, therefore, demand greater methodological creativity. Whatever studies we may consider, the burden of proof lies with those who assert that the folk are belief-realists, not with those who contend that ordinary language users make no explicit ontological commitments when speaking of beliefs.

Pośłajko’s minimal anti-realism serves as a helpful corrective, warning philosophers not to fall under the deceptive spell of ordinary language. However, this therapy must go deeper: it should also caution philosophers against developing a scientific concept of belief that exceeds anything folk speakers are committed to. In other words, we should not look down on ordinary people as if they were hopelessly trapped in belief realism. Perhaps, unlike us philosophers, they have never been there.

## REFERENCES

- Baillargeon, R., Scott, R. M., He, Z. (2010). False-Belief Understanding in Infants. *Trends in Cognitive Sciences*, 14(3), 110–118.
- Bogdan, R. J. (1993). The Architectural Nonchalance of Commonsense Psychology. *Mind and Language*, 8(2), 189–205.
- Curry, D. S. (2018). Beliefs as Inner Causes: The (Lack Of) Evidence. *Philosophical Psychology*, 31(6), 850–877.
- Dennett, D. C. (1987). *The Intentional Stance*. MIT Press.
- 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.

- Dennett, D. C. (2009). *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. MIT Press.
- Gallagher, S. (2006). *How the Body Shapes the Mind*. Clarendon Press.
- Gauker, C. (2021). Belief Attribution as Indirect Communication. In: L. Koreň, H. B. Schmid, P. Stovall, L. Townsend (Eds), *Groups, Norms and Practices. Studies in the Philosophy of Sociality* [Vol 13, pp. 173–187]. Springer.
- Gergely, G., Csibra, G. (2003). Teleological Reasoning in Infancy: The Naïve Theory of Rational Action. *Trends in Cognitive Sciences*, 7(7), 287–292.
- Hare, B., Call, J., Tomasello, M. (2001). Do Chimpanzees Know What Conspecifics Know? *Animal Behaviour*, 61(1), 139–151.
- Horgan, T. (1993). The Austere Ideology of Folk Psychology. *Mind & Language*, 8(1–2), 13–34.
- Hutto, D. D., Myin, E. (2012). *Radicalising Enactivism: Basic Minds Without Content*. MIT Press.
- Nisbett, R. E., Wilson, T. D. (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review*, 84(3), 231–259.
- Noč, A. (2004). *Action in Perception*. MIT Press.
- Posłajko, K. (2024). *Unreal Beliefs: An Anti-Realist Approach in the Metaphysics of Mind*. Bloomsbury Academic.
- Ryle, G. (1949). *The Concept of Mind*. University of Chicago Press.
- Schwitzgebel, E. (2013). A Dispositional Approach to Attitudes: Thinking Outside of the Belief Box. In *New Essays on Belief: Constitution, Content and Structure* (pp. 75–99). London: Palgrave Macmillan UK.
- Stich, S., Ravenscroft, I. (1994). What Is Folk Psychology? *Cognition*, 50(1–3), 447–468.
- Strijbos, D. W., de Bruin, L. C. (2013). Universal Belief-Desire Psychology? A Dilemma for Theory Theory and Simulation Theory. *Philosophical Psychology*, 26(5), 744–764.
- Varela, F. J., Thompson, E., Rosch, E. (2017). *The Embodied Mind, Revised Edition: Cognitive Science and Human Experience*. MIT Press.
- Woodward, A. L. (1998). Infants Selectively Encode the Goal Object of an Actor's Reach. *Cognition*, 69(1–3), 1–34.



DEVIN SANCHEZ CURRY \*

## BELIEFS ARE REAL(LY FOR SOCIAL ANIMALS)<sup>1</sup>

**SUMMARY:** In *Unreal Beliefs*, Krzysztof Pośłajko argues that socially constructed beliefs exist but are not real. In this commentary, I argue that the grounds for taking socially constructed beliefs to exist are also grounds for taking them to be real (per Pośłajko's own definition of reality).

**KEYWORDS:** belief, realism, anti-realism, interpretivism, social construction.

### 1. Introduction: Why Buy Reality When You Can Get Existence for Free?

Krzysztof Pośłajko and I have previously exchanged arguments about the metaphysics of belief, with me (Curry, 2021a, 2023a) defending interpretivism against Pośłajko's (2020, 2023a) critiques. Until recently, the last two printed lines of that exchange read as follows:

So far as I can tell, there is no coherent metaphysics of belief that is superficial yet realist without also being a variety of interpretivism. Given that Pośłajko already joins me in thinking that some realist superficial metaphysics of belief or other must be true, I cordially invite him to celebrate—perhaps even participate in!—the union of dispositionalism with interpretivism. (Curry, 2023a, p. 2145)

---

\* West Virginia University, Department of Philosophy. E-mail: devin.curry@mail.wvu.edu. ORCID: 0000-0002-9678-5353.

<sup>1</sup> Thanks to two anonymous reviewers and especially to Krzysztof Pośłajko. It is a pleasure to disagree so much with someone with whom I disagree so little.

The title of Pośłajko's new book, *Unreal Beliefs*, immediately makes it clear that I was too hasty in presuming that Pośłajko was committed to some form of realism about belief. In fact, Pośłajko (2024, pp. 112–113) argues that “belief realism is a bold and interesting hypothesis [which] does not have adequate empirical support”, and that betting against realism “is the safer option” given the present state of the empirical evidence. The contents of the book do go on to make it equally clear that I was right that Pośłajko and I share a great many relevant commitments. Pośłajko does join me in thinking that a superficial metaphysics of belief is true,<sup>2</sup> he joins me in thinking that the beliefs described by that metaphysics exist,<sup>3</sup> he joins me in thinking that ascriptions of extant beliefs are true descriptive statements,<sup>4</sup> and he joins me in thinking that those truly described extant beliefs are nonfictional.<sup>5</sup> But he emphatically does not join me in thinking that they are real.

In what follows, I will do my best to explain why not, with special attention to Pośłajko's anticipation of the objection that I am indeed inclined to make—namely, that beliefs can be both real and socially constructed. I will then argue that Pośłajko's preemptive response to this objection fails. We have good reason to countenance the reality of beliefs (not in spite of but) in virtue of the social and metacognitive processes in relation to which they exist.

## 2. Pośłajko's General Argument Against Realism

Here are several cases of belief attribution, which will serve both as stage-setting examples of the phenomenon of interest and as fodder for the ensuing discussion:

YOU—Michael Jordan to Bryon Russell (paraphrased): “You think you can guard me. You're about to get your chance”.

I—Michael Jordan: “I've always believed that if you put in the work, the results will come”.

---

<sup>2</sup> This superficialism puts us both at odds with psychofunctionalists (e.g., Porot, Mandelbaum, 2022) and others who take beliefs to be deep cognitive kinds (e.g., Mugg, 2025), with the caveat that I (Curry, 2021b; 2023b) take my superficial account of what I call “attitudes of belief” to be compatible with various deep accounts of what I call “cognitive states of belief” (so long as the two kinds of belief phenomena are not conflated with each other). In this commentary, except where explicitly noted, I will use “belief” and “beliefs” to refer to attitudes of belief.

<sup>3</sup> This commitment to the existence of beliefs puts us both at odds with eliminativists (e.g., Rosenberg, 2015).

<sup>4</sup> This commitment to the claim that belief ascriptions are descriptive statements puts us both at odds with expressivists (e.g., Fernández Castro, 2023).

<sup>5</sup> This nonfictionalism puts us both at odds with fictionalists (e.g., Parent, Toon, Demeter, 2026; for my arguments against fictionalism, see Curry, 2025a).

THEY—Alicja: “Jan and Lena think Poland will qualify for the World Cup”.

Y’ALL—Yankee: “Down here you lot think it’s bitterly cold when the temperature hits like 60 degrees” (Fahrenheit).

WE—yard sign: “In this house, we believe: black lives matter, women’s rights are human rights, no human is illegal, science is real, love is love, kindness is everything”. (I knew a guy who would vocally affirm these beliefs, like a mantra, whenever he came home and saw the sign in his front lawn).<sup>6</sup>

Posłajko’s view is that each of these belief ascriptions may well be true, and, if so, that the beliefs ascribed in each of these cases exist in the real world (as opposed to a fictional realm).<sup>7</sup> Russell did believe he could guard Jordan, and Jor-

---

<sup>6</sup> We Believe yard signs might reasonably be interpreted as expressing a group belief (ascribed to a group agent—We Who Live in This House—as opposed to each individual agent who is a member of the group in question—each person who lives in the house). Y’ALL might reasonably be interpreted the same way. Posłajko (2023b; 2025) has written insightfully about the (purported) phenomenon of group belief. Nevertheless, for the purposes of this commentary, I will ignore the possibility of group belief. My discussion will keep open two other possible interpretations. The declaration that We Believe may involve the attribution of belief to each individual member of the group (including the guy making the declaration). Or it may be a generic generalization (Leslie, Lerner, 2022)—it may be generically true of the members of the household, even though one rebelling teen does not actually believe. Likewise, the declaration in Y’ALL may involve the attribution of belief to each individual member of the group (which does not include the person making the declaration). Or, more likely, it involves a generic generalization about Southerners.

<sup>7</sup> This is to say, using Posłajko’s terminology, that the operative contents of belief attributions are often veridical. Following Haslanger (2012), Posłajko (2024, p. 132) distinguishes between the “manifest and operative concepts” of belief at play in belief attributions. Manifest concepts are what we think we mean by words. Operative concepts are how we actually use words. Posłajko argues that folks employ a manifest concept of beliefs as deep cognitive kinds, but that they also employ an operative concept of belief that is compatible with beliefs being superficial. This particular nuance in Posłajko’s view complicates his intimations (pp. 119–122) that (a) he and I (along with other realist “superficial theorists”) offer incompatible accounts of folk psychology and that (b) “there is a significant theoretical conflict between the way the superficial theories conceptualize beliefs and the way we conceive of this category in our folk schema” (p. 122). Posłajko’s distinction between manifest and operative concepts as applied to belief attribution bears strong similarities to Dennett’s (1991) distinction between folk theory and folk craft (which Posłajko himself later invokes in a slightly different context; 2024, p. 149). Since I have always stressed that my denial that folks construe beliefs as inner causes applies only to folk craft and not folk theory (e.g., Curry, 2018, p. 853)—and thus only to operative concepts and not manifest concepts—it is not clear that Posłajko and I have any deep disagreement on that score. Similarly, if superficial theories narrowly target beliefs *as attributed in folk craft*, and folk craft functions mainly via operative concepts (whereas

dan does believe that good results follow hard work. Jan and Lena do believe Poland will qualify, Southerners do believe it is cold outside when Yankees are still wearing shorts, and the members of the relevant household do believe in equality, science, and kindness. Nevertheless, Pośłajko argues that we have no good reason to accept the reality of these truth-aptly ascribed, extant beliefs. It may be true that each member of the household believes that women's rights are human rights. If so, then their beliefs that women's rights are human rights exist. Nevertheless, according to Pośłajko, their beliefs that women's rights are human rights are not real.

Pośłajko's anti-realism is grounded in a technical definition of reality. He defines reality in terms of naturalness (which is in turn to be understood along the lines of David Lewis's distinction between natural and non-natural properties; Lewis, 1983), and identifies realism about belief with the thesis that beliefs are causally relevant natural kinds that are explanatorily linked to more metaphysically fundamental properties.<sup>8</sup> The basic idea is that the world (and its properties) described by fundamental physics is perfectly natural and hence indisputably real. ("Science is real", if you like). Everything that exists but that is not part of a complete fundamental physical description of the world is less than perfectly natural. However, some nonfundamental properties of the world are way *more* natural than others, by virtue of being members of natural kinds (i.e., kinds of properties that share objective similarities with one another) that play roles in causal explanations and which, in turn, "might be somehow explained (rather than strictly defined) by their relation to more fundamental" properties (Pośłajko, 2024, p. 60). So naturalness is a graded metaproperty. For a property to be real, on Pośłajko's account, is for it to be "reasonably natural" (p. 48): reasonably objectively similar to other properties of the same kind, reasonably causally explanatory, and reasonably well explained by its relation to more fundamental properties.

Pośłajko advocates a "mild anti-realism" about belief because he thinks there is ample reason to doubt that beliefs are reasonably natural: ample reason to doubt that beliefs are natural kinds (since there are reasons to doubt that there are deep commonalities between all of the believers in any particular proposition), ample reason to doubt that beliefs have causal powers (or that they are "genuine, ineliminable, higher-level causes"; Pośłajko, 2024, p. 59), and ample reason to doubt that beliefs can be explained by their relation to more fundamental proper-

---

folk theory hinges on manifest concepts), then there need be no significant theoretical conflict between the way superficial theories conceptualize beliefs and the way we (folks) conceive of beliefs in our (crafty) folk schemas, even if Pośłajko is right (as he may well be) that most folks also sometimes employ a manifest concept that casts beliefs as deep causes of behavior.

<sup>8</sup> As Pośłajko (2024, p. 45) notes, Lewis did not himself define reality in terms of naturalness; nor did he identify natural properties with natural kinds. So, while Pośłajko more or less faithfully adopts Lewis's account of natural properties, he also puts that account to new work.

ties (since there are reasons to doubt that the representational content of beliefs can be explained in terms of underlying physical properties). By Pośłajko's lights, the fact that Jan and Lena both believe that Poland will qualify for the World Cup might tell us very little about what else they have in common; Jan and Lena's apparently relevant behaviors (cheering, betting, making travel plans) are not best explained as having been causally produced by their belief that Poland will qualify for the World Cup; and the fact that Jan and Lena's beliefs are both about Poland's World Cup chances is not well explained by more fundamental facts about the functioning of their brains or cognitive systems.

For the purposes of this commentary, I will not quibble with Pośłajko's technical definition of reality in terms of (Lewisian) naturalness. Nor will I object to his skepticism about the project of naturalizing intentionality by explaining the contents of beliefs in terms of more fundamental, non-semantic properties of believers (but see Footnote 17 below). Moreover, I am happy to admit that beliefs are not producing causes, especially in the senses that concern Pośłajko: they are not "internal, causally active states that bear [non-derived] semantic content" (2024, p. 183).<sup>9</sup> And, finally, I will not argue that there are deep intrinsic (i.e., non-relational) commonalities between all of the believers in any particular proposition. I will, nevertheless, argue that Pośłajko should countenance beliefs as reasonably natural and thus real. My strategy will be to shift our focus from believers to belief attributors.

According to interpretivists like myself, beliefs are not really for believers; real beliefs are for belief attributors. Beliefs are (relational) natural kinds, not because of deep intrinsic commonalities between believers, but because believers are veridically interpreted *as believers* by belief attributors in objectively similar ways. Beliefs are causally relevant, despite lacking causal powers, because they play causal-explanatory functions as objects in the social environments of belief attributors. (Or, in first-personal cases like Michael Jordan self-ascribing the belief that hard work pays off, as objects of metacognition). And beliefs are well-explained, not solely by their relations to more fundamental properties of the brains or cognitive systems of believers, but also by their relations to more fundamental properties of the brains or cognitive systems of belief attributors.

The relevant similarity between Jan and Lena is that they both behave, think, and feel in ways that fit how Alicja—the belief attributor—models the belief that Poland will qualify for the World Cup. And this shared belief of Jan and Lena's functions to help Alicja predict, explain, evaluate, and attempt to shape Jan and Lena's future actions, thoughts, and feelings—their sports betting, their thought that it is okay if they decline Alicja's invitation to watch this Sunday's match on her big TV since there will be more opportunities come summer, their bewildered disappointment when Poland fails to qualify. Jan and Lena's belief has the potential to pay huge dividends for Alicja—metaphorical dividends insofar as

---

<sup>9</sup> As Pośłajko readily allows (p. 59), beliefs are undeniably causes in some weaker senses of "causation" (Curry, 2018, p. 852).

veridically attributing the belief helps her navigate her friendship with them, and more literal dividends when she decides to place a bet on Saturday (with Jan and Lena’s own enterprising bookie) about whether Jan and Lena will each independently place bets on the same team winning on Sunday. Finally, the intentionality of Jan and Lena’s beliefs—the fact that their beliefs are *about* Poland’s World Cup chances—is explained by the fact that Alicja represents them as believing that Poland will qualify for the World Cup.

I will unpack this set of claims in Section 4. But Pośłajko has already anticipated the main thrust of my strategy. So I will first discuss his preemptive parry in Section 3.

### 3. Pośłajko’s Particular Argument Against Social Constructionist Realism

Towards the end of his book, Pośłajko considers “possible charges” that might be brought against his mild anti-realism. The last charge he considers, in Chapter 8 Section 6 (2024, pp. 159–163), is that the social construction of beliefs is compatible with their being natural kinds—and hence with their reality. My main aim in this commentary is to make that possible charge actual. In Section 4, I am going to argue that beliefs are real, socially constructed natural kinds. Before getting there, however, it is worth examining Pośłajko’s preemptive attempt to evade my charge in detail.

Pośłajko recognizes two related features of socially constructed kinds that provide good reason to think they are also natural kinds. First, natural kinds—including socially constructed ones like money—are projectible: they figure in surprising empirical generalizations. Second, the reason that natural kinds figure in surprising empirical generalizations is that there are robust causal links among their properties, and between their properties and other properties of the world—as per either Boyd’s (1991) homeostatic property cluster theory or Khalidi’s (2013) more structurally flexible simple causal theory. In the case of socially constructed natural kinds like money, many of these causal links may be forged through processes of social construction. If beliefs are socially constructed natural kinds, then the causal links between their properties that render them projectible are plausibly due to the folk psychological processes (including practices of belief modelling and attribution) in relation to which they are constructed.

It is worth noting that Pośłajko recognizes these features of socially constructed natural kinds with a caveat, “for the sake of the discussion”. Before doing so, he favorably cites Amie Thomasson “for an account of human kinds which seems to preclude the possibility of them being natural” (2024, p. 161). Thomasson argues in the cited article that some socially constructed kinds—artifactual and institutional kinds, which depend for their existence on “certain people accepting principles *about the nature of the kind itself*”<sup>10</sup>—“do not

---

<sup>10</sup> Given the nuances of Pośłajko’s view—and in particular his distinction between manifest and operative concepts of belief (Footnote 7 above)—it is not at all clear that

meet the realist paradigm of entities entirely independent of us for their existence and essence” (2003, p. 605). Given the particular type of realist paradigm that is Thomasson’s exclusive focus, according to which natural kinds necessarily (a) could exist even if there were no mental states and (b) are individuated by boundaries that are not imposed on the world by human concepts (p. 582), natural kindhood and artifactual/institutional kindhood do seem difficult to square. Strikingly, though, the definition of realism that Pośłajko develops at length in *Unreal Beliefs* does not obviously belong to the type of realist paradigm that concerns Thomasson. Pośłajko’s Lewisian account of naturalness/reality is silent on the question of whether the boundaries of natural kinds could be imposed on the world by human concepts, as well as on the question of whether natural kinds could rely on mental states for their existence. On Pośłajko’s account, if a kind is projectible, causally explanatory, and well explained by its relation to more fundamental kinds, then it is reasonably natural and thus real. It would seem undermotivated for Pośłajko to tack on the provision that the kind must be mind-independent (in the specific senses that concern Thomasson) solely in order to stave off the possibility of socially constructed natural kinds that otherwise meet each of Pośłajko’s criteria. It would seem much more in keeping with Pośłajko’s account to allow—and not only for the sake of discussion—that kinds can be both socially constructed and real.

The question, then, is whether folk psychological processes do construct beliefs such that those beliefs go on to play a distinctive role in the causal nexus of the world (rendering them projectible). As Pośłajko notes, the mindshaping hypothesis, which has become increasingly influential in the literature on social cognition over the last decade (Zawidzki, Tison, 2025), suggests an affirmative answer to this question. According to the mindshaping hypothesis, social practices of imitation, pedagogy, norm-enforcement, and storytelling allow people to coordinate with each other largely by shaping how everybody behaves, thinks, and feels. These mindshaping practices are well-understood as processes of social construction (Curry, 2025b; Westfall, 2025). As Pośłajko himself glosses the hypothesis, “mental states, including beliefs, are socio-culturally created by the processes of their attribution” (2024, p. 161). And these socially constructed beliefs do apparently go on to play a distinctive role in the causal nexus of the world (rendering them projectible), since mindshaped people are more predictable and intelligible (and thus more effective coordinators).

---

socially constructed beliefs do depend for their existence on people “accepting principles about the nature of the kind itself”. After all, Pośłajko (wisely) takes the beliefs that exist to be captured by operative concepts, not manifest concepts—but it is only the latter that obviously rely on the explicit (“declarative”, in Thomasson’s terminology) acceptance of principles about the nature of belief. So Thomasson’s argument may be quite irrelevant to Pośłajko’s concerns, even putting aside the difference between their definitions of realism that I spell out in the main text. As Thomasson points out, “although the above results do apply to artifactual and institutional kinds, it should not be inferred that the same follows for all ordinary, social, or (partially) mind-dependent objects” (2003, p. 605).

Posłajko thinks this last move is too fast. His argument on this score—which turns out to be his sole argument against the thesis that beliefs are socially constructed natural kinds (his nod to Thomasson notwithstanding)—is worth quoting at some length.

In order to justify the idea that a certain social kind *X* is a natural one, we must show some proof of the claim that the people/objects to which the category *X* is ascribed share important projectible properties and are a good target for inductive generalizations. It is not enough to show that the people who do the process of ascription share important properties and that the process of ascription itself can be naturalistically explained [...].

In our case, what would be needed is a story about what properties are shared by beliefs as a general category, and by people to whom we ascribe a belief with a specific content. Only by providing such a story could we claim to have positive reasons to think that beliefs are socially constructed natural kinds. However, this kind of evidence is precisely what we are lacking, and this claim was argued for in Chapter 6 Section 1 [where Posłajko argues that beliefs are probably not natural kinds since there are no objective similarities between all believers in any particular proposition]. It is not that we know that there are no generalizations about beliefs; instead, we lack evidence to think that there are any, and this lack of evidence might be taken to support the view that beliefs should not, at the present moment, be conceptualized as being natural kinds. Obviously, this need not be true about mindreading [i.e., belief attribution], mindshaping, mindreaders and mindshapers: there might be important empirical discoveries about these processes and people who attribute mental states. (Posłajko, 2024, pp. 161–162)

Posłajko grants, in other words, that belief *attribution* processes might well be natural kinds, since the mindshaping hypothesis suggests that they play an important causal role in shaping how people behave. But the causal potency of belief attribution does not guarantee the reality of belief. (After all, accusations of witchhood are causally potent—they cause women to be burned at the stake—even though witches are not real). If all of the predictive and explanatory power of folk psychology can be explained in terms of the causal influences of belief attribution (and other mindshaping practices), then the fact that the social construction of belief makes people more effective coordinators simply does not entail that beliefs are real. For socially constructed beliefs to be real (and not merely extant), they would have to themselves play distinctive causal roles (and thus be projectible). And, Posłajko says, we have no good reason to think they are.

But in Chapter 6, where Posłajko argues for that claim—that we have no good reason to think beliefs play distinctive causal roles—he is not yet considering the question of whether beliefs are *socially constructed* natural kinds. He does not entertain that possibility at any length until Chapter 8. In Chapter 6, he is still considering the question of whether beliefs are non-constructed cognitive kinds—psychofunctional and representational kinds à la Quilty-Dunn and Mandelbaum (2018), for instance. His argument against considering beliefs to be real cognitive kinds rests on the claims that, “importantly, there seems to be no con-

vergence in the criteria for detecting beliefs” across the cognitive sciences and that “the category of belief (as defined in cognitive psychology) seems to play many distinct theoretical roles” (Posłajko, 2024, p. 96). He concludes Chapter 6 Section 1 as follows:

To sum up, the reasons to claim that the category of belief is a cognitive natural kind seem to be outweighed by arguments to the effect that it is not. The best interpretation of the evidence we have in this respect is to consider this category as just one of the many possible ways of categorizing the human mental sphere and that we should have no presumption that by using it we are referring to any joint in cognitive nature. The preferred alternative to the idea that beliefs constitute a natural kind seems to be the view that they constitute a human kind. The claim has been recently put forward by, for example, Dewhurst (2017), according to whom the “belief” kind should be treated as, in a way, socially constructed and not necessarily matching the deep cognitive ontology. In my opinion, such a view corresponds nicely with the idea that belief properties are non-natural, in Lewis’s sense, which is the central claim of the minimal non-realist view I develop in this book (but see Chapter 8 Section 6 for an important objection to the idea that construction goes hand in hand with rejection of realism). (Posłajko, 2024, p. 97)

Chapter 8 Section 6 is, of course, where Posłajko presents the argument against beliefs being socially constructed natural kinds quoted at length above. In that argument, he punts the central move—denying that there is any reason to think that there are projectible, causally linked properties shared by beliefs and believers—back to Chapter 6 Section 1. But in Chapter 6 Section 1 he is interrogating the idea that beliefs exist as cognitive kinds, within a “deep cognitive ontology”, rather than as socially constructed kinds. As such, in Chapter 6, he is focused on examining (and finding wanting) claims about the causal roles that beliefs play as cogs in cognitive systems. Upon acknowledging the possibility of belief being a socially constructed but nevertheless natural kind—which is the possibility of a kind of belief that, if it plays any distinctive causal role at all, presumably plays a quite different causal role than beliefs considered as theoretical posits in cognitive science—Posłajko refers the reader to Chapter 8 Section 6. Unfortunately, as we have seen, Chapter 8 offers no new arguments against the causal relevance of the socially constructed kind of the belief in particular. It just sends the reader back to Chapter 6.

In sum, there is a lacuna in Posłajko’s argument against the socially constructed reality of beliefs. Nevertheless, at this juncture in the dialectic, Posłajko is likely to recommend mild anti-realism all the same: if we have no good reason to think that socially constructed beliefs are causally relevant and projectible, then the absence of a good argument for anti-realism might not be enough to take the leap and countenance beliefs’ reality. So, in Section 4, I will argue that we do have good reason to think that socially constructed beliefs are causally relevant and projectible in manners that allow them to meet all three of Posłajko’s criteria for reality.

#### 4. Interpretivist Realism Defended

As I have argued elsewhere (Curry, 2021b), an interpretivist approach to countenancing beliefs as social constructs suggests that beliefs are superficial patterns of dispositions that are individuated in terms of the functional roles they play in belief attributors' social environments rather than the functional roles they play in believers' cognitive systems. Pośłajko is, as we have seen, content to allow that beliefs exist qua socially constructed superficial patterns of dispositions. The question is whether they are real. And that depends on whether they are projectible, causally explanatory, and themselves explained in terms of more fundamental properties.

As Pośłajko rightly points out, to answer those questions in the affirmative, "it is not enough to show that the people who do the process of ascription share important properties and that the process of ascription itself can be naturalistically explained" (2024, p. 162). Belief attribution capacities may be projectible, causally potent natural kinds insofar as they really do help belief attributors predict, explain, judge, and shape people. But witch accusation capacities are likewise projectible, causally potent natural kinds insofar as they really do help accusers burn poor women at the stake. That does not mean witches are real.

There is, however, a crucial difference between witches and beliefs: witches do not exist. The major problem with including witches in one's realist ontology is not that their magic falls short of being causally potent enough to qualify as natural and hence real. The problem is that magic does not exist. Beliefs do exist, even by Pośłajko's anti-realist lights. Whereas witch accusations are never veridical, belief attributions are sometimes veridical (and sometimes not).<sup>11</sup> When they are veridical, it is because the attributed beliefs exist: the believer genuinely does take the world to be the way the belief attributor represents them as taking the world to be. Bryon Russell genuinely does think he can guard MJ. His behavior on the basketball court is more readily predictable, explainable, evaluable, and manipulable in light of that belief. To belabor the simple but pivotal contrast: nobody is a witch; nobody's behavior is more readily predictable, explainable, evaluable, and manipulable in light of their magical powers. (Though, regrettably, many people's behavior is more readily predictable, explainable, evaluable, and manipulable in light of their *belief in* magical powers).

To be fair, Pośłajko does not draw the analogy between belief attribution and witch accusation. Instead, he draws an analogy with debates about the metaphysics of race, writing that

---

<sup>11</sup> Veridicality was the subject of Pośłajko's (2020) initial critique of interpretivism, as well as my initial reply to Pośłajko (Curry, 2021a), in which I explain how (according to my version of interpretivism) belief attribution can often fail to be veridical even though beliefs are constituted in relation to attributors' models. In brief, an attributor can be mistaken about their target possessing dispositions that sufficiently fit the attributor's model of belief.

if one wanted to claim that races are socially constructed natural kinds, one would have to provide positive reasons that the people being racialized share important properties and that we can make surprising empirical generalizations about them. It would simply not do if we tried to show this by pointing to the fact that people who racialize others share some important properties and that there is an important naturalistic story about the origins of the racialization process. (Posłajko, 2024, p. 162)

Posłajko is right about this. For instance, as Kareem Khalifa and Richard Lauer have argued, the fact that the social sciences make surprising empirical generalizations using race-concepts does not provide evidence for social constructionist realism about race, since the causal-explanatory success of race-concepts is just as well accounted for by the anti-realist view that “some groups are racialized, but races do not exist” (2021, p. 9). But note that the standard anti-realist accounts on offer in the literature on the metaphysics of race (Appiah, 1985; Fields, 2012; Glasgow, 2009) are accounts on which races do not even exist, not accounts on which races exist but merely are not natural or real. If racialization can exhaustively explain all of the surprising empirical generalizations that use race concepts, then we are not merely left with doubts about the naturalness of race: we are left without any good reason to take race to exist at all.

By contrast, Posłajko grants that we have plenty good reason to take beliefs to exist. It is true that Bryon Russell thinks he can guard Jordan, and it is false that Greg Ostertag thinks he can guard Jordan. Why? Because Russell’s belief exists (qua superficial pattern of dispositions) whereas Ostertag lacks any sufficiently similar pattern of dispositions. So there is something there—an extant belief—to do extra causal work in the case of belief attribution, whereas there is nothing there—no extant witchcraft or racial essence—to do extra causal work in the cases of witch accusation or racialization (assuming anti-realism about race is true). Does that extant belief do *enough* causal work, over and above the causal work done by the belief attribution alone, to qualify as real?

I reckon it does. I do not have knockdown proof. Airtight evidence of causality is notoriously evasive in the social sciences, and this is an empirical, paradigmatically social scientific question (or set of three questions) if ever there was one. Insofar as Posłajko is dedicated to his stance that, in the absence of strong scientific evidence of projectability, we should lean towards anti-realism, it is possible that nothing I have to say can persuade him. For my own part, though, I am far from persuaded that it is sensible to lean anti-realist by default. After all, belief attributors do very often have good reasons for attributing belief, and those reasons very often seem to be predicated on their being able to track something real about believers.

Beliefs play a diversity of epistemic and symbolic functional roles for belief attributors: they function to help belief attributors describe, predict, explain, contextualize, evaluate, regulate, and manage impressions of believers’ behav-

ioral and mental lives.<sup>12</sup> In order to play many (though perhaps not all) of these roles, it sure seems to be crucially important whether the belief attribution is veridical. Alicja will not be able to predict Jan and Lena's betting behavior if she is wrong that they believe Poland will qualify for the World Cup. If she is right about what they believe, then her predictions will be much more accurate. A belief veridically attributed to oneself, meanwhile, will often lead to action on the basis of that belief. Michael Jordan putting in the work may not be directly causally produced by his belief that hard work pays off, but an extra trip to the gym is very likely to be spurred by his self-knowledge that that is what he truly believes. The guy who displays his We Believe yard sign will not successfully signal the political identities of his housemates—and will not be able to leverage those political identities to strengthen the social bonds between them—if he is wrong to presume that they share his feminist and anti-racist commitments. If his belief attributions are veridical, however, then his yard sign might help generate the solidarity he is seeking, especially when paired with the attribution of opposing beliefs to an out-group. The Yankee's attribution of the belief that 60°F is cold to an out-group might similarly serve to accentuate differences.<sup>13</sup> Alternatively, though, the Yankee's veridical belief ascription might serve to help him better appreciate the perspectives that his Southern acquaintances genuinely have, bridging the personal and cultural divide between them. If so, his social success will not be adequately explained by his act of belief attribution alone; it will have been caused in large part by the fact that he got the Southerners' beliefs right. In that case, as in the case where Alicja gets rich by predicting how Jan and Lena will bet, as well as the case where Jordan becomes the best basketball player in the world, the beliefs themselves—qua patterns of dispositions that render

---

<sup>12</sup> The labels “epistemic” and “symbolic” as used here come from Evan Westra (2023), who suggests that they apply to two distinct concepts of belief. When belief attributors want to predict, explain, or contextualize behavior, or simply want to keep track of what others take to be true, they use the epistemic concept of belief (and typically speak of what people “think”). When belief attributors want to signal social identities or enforce social norms, they use the symbolic concept of belief (and typically speak of what people “believe”). I am not convinced that these epistemic and symbolic functions should be thought of dichotomously—I suspect that almost every belief attribution serves both epistemic and symbolic functions (weighted differently)—so in the main text I write of epistemic and symbolic functional roles played by beliefs rather than epistemic and symbolic concepts of belief *per se*.

<sup>13</sup> For this particular, otherizing symbolic purpose, it is admittedly not obvious that veridicality is indispensable (for roughly the same reason that it is not obvious that veridicality is indispensable in the case of racialization). But it cannot hurt! Indeed, I suspect that the causal roles played by veridical out-group belief attributions are quite distinct from the causal roles played by non-veridical out-group belief attributions, since the former plausibly unlock predictive and explanatory power that the latter cannot.

belief attributions veridical—are doing significant causal-explanatory work in helping belief attributors achieve their goals.<sup>14</sup>

If (a standard version of) anti-realism is the correct metaphysics of race, then the relevance of veridicality marks a significant difference between belief attribution and racialization. If anti-realism about race is true, then races do not exist, racialization is never veridical, and the (non)veridicality of racialization is of no causal-explanatory importance. Anti-realists about race are thus well-positioned to explain the causal-explanatory power of race-concepts solely in terms of the properties of racializers, without needing to pay any heed to the properties of the racialized.<sup>15</sup> But by (rightly!) refusing to deny the existence of beliefs, Pośłajko severely limits the relevance of his analogy to racialization. Belief attributions are often veridical, and their veridicality is of causal-explanatory importance. Pośłajko is thus not well-positioned to explain the causal-explanatory power of belief attribution solely in terms of the properties of belief attributors, without needing to pay any heed to the properties of believers. On the contrary, in order to adequately capture the causal links between veridical belief attributions and the further actions they license, Pośłajko must pay heed to the properties of believers—the beliefs—that make those belief attributions veridical.

So I think Pośłajko should countenance the reality of beliefs as socially constructed natural kinds. He should turn his back on mild anti-realism and embrace the realist marriage of dispositionalism and interpretivism after all. As I have put the view,

to believe is to live—roughly, to be disposed to act, react, think, and feel—in a pattern that fits an actual belief attributor’s interpretive model of how somebody would tend to act, react, think, and feel if they took the world to be some particular way. (Curry, 2023a, p. 2139)

For a property to be real (per Pośłajko’s definition) is for it to be reasonably natural: reasonably objectively similar to other properties of the same kind, reasonably causally explanatory, and reasonably well explained by its relation to more fundamental properties. In order to motivate his mild anti-realism, Pośłajko notes that beliefs are neither *intrinsically* objectively similar to each other nor

---

<sup>14</sup> I take these examples of veridicality mattering to motivate an account of beliefs as natural kinds at least on Khalidi’s (2013) simple causal theory of natural kinds. I take it to be an interesting open question whether beliefs also meet the more stringent criteria for natural kindhood set by Boyd’s (1991) homeostatic property cluster theory. I suspect they do, since the near-universality (and apparently cross-culturally stable aspects) of belief attribution practices lead me to suspect that mindshaping serves as a mechanism for the requisite homeostasis. But making good on that suspicion would require further work.

<sup>15</sup> If, on the other hand, there is good reason (as I suppose there may be) to think racialization is like belief attribution (and unlike witch accusation) in needing to be veridical to function as it does actually function in human social life—and thus to causally explain everything that it does causally explain—then that is reason to reject anti-realism and accept a realist social constructionism about race.

*intrinsically* causally explanatory. But if my interpretivism is on the right track, then this emphasis on the intrinsic properties of beliefs is neither here nor there. Rather, per interpretivism, beliefs are real since they are reasonably objectively similar to each other in virtue of each fitting a belief attributor's model, and reasonably causally explanatory in virtue of the causal-explanatory roles they play in belief attributors' social environments. These are ineliminably relational—non-intrinsic—similarities and causal roles. But there is nothing in Pośłajko's account of reality that prohibits reasonably natural (as opposed to perfectly natural) properties from being relational.<sup>16</sup>

Consider an analogy with poison. Going by Pośłajko's account of reality, I would judge that poison does not merely exist—it is a real natural kind. Why? Not because of an intrinsic chemical property that the various poisons all share in common. The hodgepodge set of chemical formulas that describe the various substances that each really are poisons might seem totally arbitrary from the perspective of the chemist. But the set of poisons is not arbitrary, certainly not from the perspectives of the medical practitioner or the assassin. Rather, poison is real because poisons are reasonably objectively similar to each other in virtue of being harmful or lethal to organisms, and reasonably causally explanatory in virtue of the causal-explanatory roles they play in organisms' environments. These too are ineliminably relational—non-intrinsic—similarities and causal roles. That relationality does nothing to make me doubt poison's reality. (Indeed, I would strongly recommend against anyone doubting the causal relevance of poison).

Highlighting that beliefs exist in relation to belief attributors—like poisons exist in relation to vulnerable organisms—also sheds light on how beliefs meet Pośłajko's third criterion: explicability in terms of more fundamental properties. Real beliefs—extant patterns of dispositions that fit interpreters' models of belief—will plausibly be adequately explained by scientific research into two sets of plainly more fundamental properties: the subpersonal cognitive capacities of believers, on the one hand, and the subpersonal cognitive capacities of belief attributors, on the other hand. (Compare: poisons are adequately explained by scientific research into the chemical properties of poisonous substances, on the one hand, and the physiology of vulnerable organisms, on the other hand). The

---

<sup>16</sup> Pośłajko might be thought to prohibit relational properties from being reasonably natural via his invocation of Thomasson (discussed above), along with his repeated (2024, pp. 45, 83–84, 134) suggestion that Chomsky's (2000) anti-realism about desks and cities (as well as beliefs) is similar in spirit to his approach. (Chomsky is a proponent of the variety of mind-independence-centric realist paradigm that concerns Thomasson). However, I do not think Chomsky's line of reasoning fits Pośłajko's Lewisian framework as neatly as Pośłajko suggests—the three criteria for naturalness strike me as perfectly compatible with mind-independent phenomena being (reasonably) natural. In any event, I think Chomsky is mistaken. Desks and cities are real. They are relationally real—desks and cities are similar to one another, and causally explanatory, only in relation to people who use them as work surfaces and complex social institutions (respectively)—but they are no less real for that. (I am at one and in one right now!).

underlying psychology of believers—in tandem with broader mindshaping processes—will explain how believers come to have the relevant dispositions. And the underlying psychology of belief attributors—in tandem with broader mindshaping processes—will explain how the relevant patterns of dispositions come to be modelled as beliefs. On this interpretivist picture, the intentionality of beliefs is also derived from belief attributors' models, rather than from the underlying cognitive states of believers that generate the dispositions that make up their beliefs. Russell's belief is about guarding Jordan because it is represented as having that intentional content by the belief attributor (whether that be Jordan, Russell himself, or anybody else positioned to make the ascription).<sup>17</sup>

Why not instead say, with Pośłajko, that these more fundamental properties—the cognitive capacities of believers and belief attributors—are real, whereas beliefs themselves, being less fundamental, are not? Because beliefs, like poisons, are projectible and play causal-explanatory roles that are not well captured at the more fundamental level of analysis. We have already seen why attention to belief attribution alone (without regard for veridicality) does not suffice to account for the causal-explanatory roles played by extant beliefs. Moreover, in his Chapter 6 argument for mild anti-realism, Pośłajko accidentally draws attention to another reason that beliefs should be countenanced as real relational properties. If the cognitive kinds underpinning beliefs are not reasonably objectively similar to each other, as Pośłajko argues that they are not, then the objective similarities that veridically attributed beliefs do share in common—for example, the fact that they all count as beliefs and play the same functional roles in the social environments of belief attributors—must be similarities they boast *qua* socially constructed beliefs. The deeper, more explanatorily fundamental kinds lack the projectability and causal links that the superficial kind boasts—and which qualify the superficial kind as real.

Pośłajko writes that “belief attributions might be true but they add very little to our understanding of what is really, ‘on a fundamental level’, going on within the subjects” (2024, p. 66). I agree. That is not their job. Poison attributions add very little to our understanding of what is really, “on a fundamental level”, going on within chemical substances. Poison is nevertheless real because poison attributions (when veridical) tell us a great deal about what will go on if we ingest the chemical substance in question. Similarly, belief attributions tell us who takes the world to be one way and not another. This adds very little to our understanding of what is really, “on a fundamental level”, going on with those people. But it can be extremely helpful when we desire—as humans often do—to distinguish

---

<sup>17</sup> I should admit straight away that this picture pushes off rather than solving the issues with naturalizing intentionality that worry Pośłajko, since I am assuming that belief attribution capacities are representational capacities. Nevertheless, pushing off the issue is enough for present purposes, since it allows me to highlight how beliefs fulfill Pośłajko's third criterion by being explainable in terms of more fundamental cognitive capacities. All the better that I can do so without having to tackle the problem of intentionality. (I would like to do other things with my life).

between the believers and the non-believers. It behooves Jordan to know who thinks they can guard him and who is afraid to try. (He will cross Russell up a few times to chip away at that belief, but he will also call many pick and rolls in order to get Ostertag, who emphatically does *not* believe he can guard Jordan, switched onto him). Alicja may not be tracking a deep intrinsic similarity between Jan and Lena; nevertheless, if she is interested in making a big wager—not on the outcome of a soccer match, but on what Jan and Lena are going to bet will be the outcome of a soccer match—then she would rightly balk at Pośłajko's assertion that "attributing beliefs adds little to our predictive abilities". The Yankee will strengthen his friendship with his Southern guests when he brings extra sweaters to the cookout on the basis of his veridical ascription of their belief that 60°F is cold. Their beliefs are plenty real for him. And that is reality enough, since social (and self-reflective) animals like him are whom beliefs are really for.

### 5. Conclusion

So far as I could tell before reading Pośłajko's book, there is no coherent metaphysics of belief that is superficial yet realist without also being a variety of interpretivism. So far as I can tell now, there is no plausible metaphysics of belief that is ontologically committed to the existence of beliefs without also being a variety of realism. Given that Pośłajko already joins me in thinking that beliefs exist, and that some superficial metaphysics of belief or other must be true, I cordially invite him to celebrate—perhaps even officiate!—the union of dispositionalism with interpretivism in the service of realism.

Pośłajko is well-qualified to officiate. The account of realism qua (reasonable) naturalism at the heart of *Unreal Beliefs* does provide a compelling basis for doubting the reality of the cognitive states of belief posited by psychofunctionalists and other cognitive scientists. But it simultaneously provides a compelling basis for accepting the reality—the projectability, causal-explanatory relevance, and reductive explainability—of the sorts of (socially constructed) beliefs that people ascribe to each other in everyday life. Since those are the beliefs that Pośłajko already countenances as existing, my simple recommendation is that he convert to realism about the beliefs that exist. I am happy to join him in being an anti-realist about any and all beliefs that do not exist.

### REFERENCES

- Appiah, K. A. (1985). The Uncompleted Argument: Du Bois and the Illusion of Race. *Critical Inquiry*, 12(1), 21–37.
- Boyd, R. (1991). Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds. *Philosophical Studies*, 61(1/2), 127–148.

- Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge University Press.
- Curry, D. S. (2018). Beliefs as Inner Causes: The (Lack Of) Evidence. *Philosophical Psychology*, 31(6), 850–877.
- Curry, D. S. (2021a). Interpretivism Without Judgement-Dependence. *Philosophia*, 49(2), 611–615.
- Curry, D. S. (2021b). How Beliefs Are Like Colors. *Synthese*, 199(3–4), 7889–7918.
- Curry, D. S. (2023a). Why Dispositionalism Needs Interpretivism: A Reply to Pośłajko. *Philosophia*, 51(4), 2139–2145.
- Curry, D. S. (2023b). Morgan’s Quaker Gun and the Species of Belief. *Philosophical Perspectives*, 37(1), 119–144.
- Curry, D. S. (2025a). Nonfiction Stories About Minds. *Topoi*, 44, 1291–1302.
- Curry, D. S. (2025b). How Scientific Psychology Shapes Minds. In T. Zawidzki, R. Tison (Eds.), *The Routledge Handbook of Mindshaping* (pp. 330–341). Routledge.
- Dennett, D. C. (1991). Two Contrasts: Folk Craft vs Folk Theory and Belief vs Opinion. In J. Greenwood (Ed.), *The Future of Folk Psychology* (pp. 135–148). Cambridge University Press.
- Dewhurst, J. (2017). *From Folk Psychology to Cognitive Ontology* [Doctoral dissertation, University of Edinburgh]. <http://hdl.handle.net/1842/25916>
- Fernández Castro, V. (2023). Folk Psychology Without Metaphysics: An Expressivist Approach. *South African Journal of Philosophy*, 42(2), 128–143.
- Fields, B., Fields, K. (2012). *Racecraft: The Soul of Inequality in American Life*. Verso.
- Glasgow, J. (2009). *A Theory of Race*. Routledge.
- Haslanger, S. (2012). *Resisting Reality: Social Construction and Social Critique*. Oxford University Press.
- Khalidi, M. A. (2013). *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*. Cambridge University Press.
- Khalifa, K., Lauer, R. (2021). Do the Social Sciences Vindicate Race’s Reality? *Philosophers’ Imprint*, 21, 1–17.
- Leslie, S. J., Lerner, A. (2022). Generic Generalizations. In E. Zalta, U. Nodelman (Eds.), *Stanford Encyclopedia of Philosophy*. Retrieved April 12, 2026, from <https://plato.stanford.edu/archives/fall2022/entries/generics/>
- Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61(4), 343–377.
- Mugg, J. (2025). *From Human Reasoning to Belief: An Empirical Account*. Routledge.
- Parent, T., Toon, A., Demeter, T. (2026). *Mental Fictionalism*. Cambridge Elements.
- Pośłajko, K. (2020). Can Deflationism Save Interpretivism? *Philosophia*, 48(2), 709–725.
- Pośłajko, K. (2023a). Folk Psychological Models and the Metaphysics of Belief. A Reply to Curry. *Philosophia*, 51, 919–931.

- Posłajko, K. (2023b). The Meta-Metaphysics of Group Beliefs: In Search of Alternatives. *Synthese*, 113(4), 1–18.
- Posłajko, K. (2024). *Unreal Beliefs: An Anti-Realist Approach in the Metaphysics of Mind*. Bloomsbury.
- Posłajko, K. (2025). Legislative Intent and the Hard Problem of Content. *Law and Philosophy*, 44, 327–356.
- Porot, N., Mandelbaum, E. (2022). The Science of Belief: A Progress Report. In J. Musolini, J. Sommer, P. Hemmer (Eds.), *The Cognitive Science of Belief: A Multidisciplinary Approach* (pp. 55–91). Cambridge University Press.
- Rosenberg, A. (2015). The Genealogy of Content or the Future of an Illusion. *Philosophia*, 43(3), 537–547.
- Thomasson, A. (2003). Realism and Human Kinds. *Philosophy and Phenomenological Research*, 67(3), 580–609.
- Westfall, M. (2025). Mindshaping and Constructing Kinds. In T. Zawidzki, R. Tison (Eds.), *The Routledge Handbook of Mindshaping* (pp. 316–329). Routledge.
- Westra, E. (2023). Symbolic Belief in Social Cognition. *Philosophical Perspectives*, 37(1), 388–408.
- Zawidzki, T., Tison, R. (2025). *The Routledge Handbook of Mindshaping*. Routledge.

BRUNO MÖLDER \*

## BELIEF: MINIMAL NON-REALISM AND THE CONSTITUTIVE ROLE OF INTERPRETATION

**SUMMARY:** Pośłajko's minimal non-realism about belief combines a deflationary account of belief's existence, a denial that beliefs are natural properties, and a revisionary stance toward the current folk concept of belief. This paper argues that rejecting or revising the folk concept is not necessary for anti-realism and develops a framework that allows for a more nuanced articulation of realist and anti-realist options about belief. The second half of the paper defends interpretivism, another deflationary position alongside minimal non-realism. In particular, it clarifies the constitutive role of interpretation in what it takes to have mental states, and the relationship between canonical ascribability and the possession of mental properties.

**KEYWORDS:** belief, minimal non-realism, anti-realism, interpretivism, causation, constitution, folk psychology.

### 1. Introduction

Krzysztof Pośłajko's *Unreal Beliefs: An Anti-Realist Approach in the Philosophy of Mind* is a meticulously argued defence of anti-realism about belief. It is clearly written and largely congenial to my own philosophical outlook. In my own work, however, I have generally tried to avoid using broad labels such as "realism" and "anti-realism" in order to minimise confusion (Mölder, 2010, p. 14). Since these terms can be understood in multiple ways, I have been wary

---

\* University of Tartu, Department of Philosophy. E-mail: bruno.moelder@ut.ee. ORCID: 0000-0002-0136-9729.

of the risk that once one adopts, say, the ominous label of “anti-realism”, people assume that one incurs certain commitments that may come with the label even if the actual position is more intricate and nuanced. For this reason, I have found it more productive to work with related but more specific notions, such as “objectivity” or “recognition-dependence”. Of course, it is possible to define the term “realism” more narrowly from the outset, and this is what Pośłajko does. However, this approach carries the risk that certain intuitive aspects commonly associated with realist or anti-realist positions might be left out, and some formulations can then sound paradoxical, such as the claim that beliefs exist but are not real.

Let us begin with a brief summary of how Pośłajko construes the debate over beliefs. As already hinted, he distinguishes the question of the existence of beliefs from the question of the reality of beliefs. On the standard Quinean understanding, the reality and existence of something are the same: things that exist are real, and things that do not exist are not real. In contrast, Pośłajko adopts the deflationary view that existence questions are not substantial. Following Thomasson (2015), he regards answering existence questions as an easy matter—one only has to assess whether the application criteria for the relevant term are met. In this way, philosophical questions about the existence of beliefs can be set aside as trivial. The existence of beliefs no longer poses a deep philosophical problem. In fact, it is difficult to find anti-realist philosophers who would deny the existence of beliefs in this deflationary sense. How, then, should we distinguish between belief realism and belief anti-realism?

For Pośłajko, the main divide between realists and anti-realists concerning belief lies not in the question of whether beliefs exist but in whether they are real. He takes real properties to be natural in David Lewis’s (1983) sense of “sparse” properties. For Pośłajko (2024, p. 51), there are three hallmarks of natural properties: they form natural kinds, since they are objectively similar; they are referred to in causal explanations; and they stand in relations to fundamental or “elite” natural properties.<sup>1</sup> Pośłajko argues that beliefs fail to be natural in all three respects. The position he proposes with regard to beliefs is “minimal non-realism”. He articulates this position through the following five points (pp. 64–65):

1. Belief attributions can be true and often are; there are beliefs.
2. Attributed belief states share no objectively similar properties.
3. Beliefs are not natural kinds.
4. Beliefs do not play a causal role, and belief attributions do not have “genuine explanatory power in the context of causal explanations”.
5. Belief contents cannot be naturalised.

---

<sup>1</sup> Henceforth, I refer to Pośłajko’s (2024) only by page numbers.

I find myself in agreement with almost every item on this list. In Mölder (2010), I also adopt a deflationary approach to truth and regard the existence of beliefs *qua* pleonastic entities as ontologically unproblematic (as Posłajko also acknowledges on p. 36). I do not think that there are natural belief properties and have argued that mental states (including beliefs) do not form natural kinds (Mölder, 2019; 2024). My views on mental causation are a bit more nuanced, as I distinguish between causal efficacy and causal relevance. I do think that there is no efficient mental causation, but explanations in terms of mental states could be causally relevant in the sense of Jackson and Pettit (1990): if successful, causally relevant explanations indicate that there are some causally efficacious non-mental processes taking place in the brain. I do not find that the objections to mental causation that Posłajko develops in sect. 6.2 apply to this very light notion of causal relevance. Finally, I have also never truly believed in the prospects of a naturalistic account of mental content.

Given these broad agreements on non-realism about belief, are there any differences between us worth discussing? There are two such issues that I will focus on in this paper. First, I argue that Posłajko's more specific way of setting up his position, and particularly the revisionary character of his account of belief, creates problems for how his account should be labelled. The second point of difference concerns interpretivism, a view that Posłajko has criticised on various counts. He does not see the need to regard beliefs as being constitutively dependent on interpretation and prefers minimal non-realism, which is arguably simpler than interpretivism. In what follows, I address both points.

## 2. The Varieties of Realism and Anti-Realism

In more general terms, Posłajko (p. 1) presents his minimal non-realism about belief as a conjunction of three statements: (1) beliefs exist in a deflationary sense; (2) beliefs are not real; and (3) the folk-psychological concept of belief is mistaken and requires amelioration. While the first two claims are also covered in the five points above, the third clause is somewhat surprising and may be seen as in tension with the other claims. It is not entirely clear why denying that beliefs are real is not already sufficient for anti-realism about belief. Why bring the issue of the correctness of folk psychology to bear on this question? Let us explore this further.

My contention is that Posłajko's notion of realism (and, respectively, anti-realism) collapses two notions that should be kept separate for the sake of clarity. One is realism as a metaphysical view about the nature of an entity or a class of entities; the other is realism as a view about a particular discourse (cf. Demeter, 2009). Thus, to be a realist about beliefs is to take beliefs as real entities—if one is a physicalist, then beliefs would be physical entities, but one could also be

a non-natural realist about beliefs and regard them as, say, non-physical states.<sup>2</sup> But surely one way to be an anti-realist about beliefs is to deny that beliefs are natural properties in Lewis's sense. Note that nothing has been said about folk psychology yet. This is because realism/anti-realism about folk psychology can be separated from the question of realism/anti-realism about belief. Realism about the folk-psychological discourse is foremost a semantic thesis, which states that (i) mentalistic ascriptions in folk-psychological terms are truth-apt statements capable of representing mental facts (in some sense or other), and (ii) when we make these ascriptions, our aim is to represent these facts successfully.<sup>3</sup> One way of being an anti-realist about folk-psychological discourse is to accept (i), but deny (ii). Mental fictionalism is such a position, as it takes folk ascriptions to be true or false, but denies that our aim in making these ascriptions is representational. Another kind of anti-realism about folk-psychological discourse would deny (i), namely, that folk-psychological sentences are truth-apt. This would amount to a kind of non-cognitivism about folk psychology.<sup>4</sup>

Metaphysical views on entities can be combined with various views on discourse. While entity-realism about beliefs is most naturally combined with realism about folk-psychological discourse (a Fodorian position), an entity-realist about beliefs can also be an anti-realist about folk psychology. This would be the position of someone who takes beliefs to be a natural kind and thinks that folk psychology is not a good guide to the nature of belief. She could hold that folk psychology mainly fulfils regulative or evaluative functions instead of a descriptive one. Realism about folk psychology combined with anti-realism about beliefs would amount to an error theory about folk psychology, which in the philosophy of mind, is exemplified by eliminativism. On that view, folk-psychological discourse is representational, but fails, as folk is wrong about the nature of belief. Another way to combine realism about folk-psychological discourse with belief anti-realism would be to read (i)–(ii) in a deflationary manner and deny the natural kindhood of beliefs. Anti-realism on both counts would

---

<sup>2</sup> In principle, entities could be “sparse” or “natural” in Lewis's sense without being naturalistic or physical.

<sup>3</sup> These correspond to the first two elements of realism in Wright (1995, p. 198), who also adds a metaphysical clause (see also Drayson, 2022, p. 148; Mölder, 2022, p. 321). However, the metaphysical view on entities and the semantic view on discourses are separable. This is made clear by Demeter (2009), who has served as an obvious inspiration here. Yet, while Demeter distinguishes between entity realism and discourse realism, he understands entity-realism in terms of ontological commitment, thus linking it to the question of existence. In the context of Poślajko's book, entity realism can be given more substantial content in terms of natural properties.

<sup>4</sup> For a recent work on mental fictionalism, see Demeter, Parent, and Toon's (2022). Poślajko discusses non-cognitivism about belief on pp. 11–14. A recent and more sophisticated non-cognitivist approach, which allows belief ascriptions to be truth-conditional, is presented by Fernández Castro (2023a; 2023b). He defends an expressivist view of folk psychology, according to which ascribing mental states is not an attempt to describe the subject's mind but rather to express an evaluative attitude.

deny that beliefs are natural entities and would be either fictionalist or non-cognitivist about the folk-psychological discourse, depending on whether one rejects (ii) or (i).

As this discussion hopefully shows, one gets a more nuanced elaboration of the realism issue by distinguishing between realism about entities and realism about discourse. Given that these stances are independent, a realist about beliefs *qua* entities need not take a stance on the status of the folk-psychological concept of belief. A realist who takes beliefs to be natural kinds usually assumes a distinction between belief as an entity and our ordinary concept of belief, and does not care much about the latter. It is enough that the folk concept of belief sets the process of empirical study in motion. Demeter (2009, p. 66) describes this perspective on the folk notion in this way: “Folk psychology may provide initial contingent, and thus possibly false, descriptions of entities to start with, but it has no special importance beyond that”. Analogously, an anti-realist about belief who does not think that beliefs are natural properties is not bound to assume that the folk concept of belief is flawed (from a certain angle, my own position could be characterised in such terms).

Posłajko’s treatment of folk-psychological concepts actually includes both semantic (or epistemic) and metaphysical aspects. He distinguishes between the application criteria of a notion and its metaphysical commitments (pp. 123–128). The notion of belief has superficial application criteria: beliefs are normally attributed on the basis of information about how people behave and what they say. Information about processes in their internal cognitive system or brain is not needed. On the inferentialist conception that Posłajko adopts, the metaphysical commitments of a term are fixed by its inferential relations, and according to him, these commitments are substantial. Namely, when we attribute a belief to someone, we are thereby licensed to infer that the person has an internal state that causes her behaviour.

Posłajko finds these commitments to internal causal states flawed and proposes revising our ordinary concept of belief. This revision should replace the current substantial concept of belief with a superficial one (p. 131). However, Posłajko does not elaborate on how exactly this revision should be effected. He notes, incorporating Haslanger’s (2012) distinction between manifest and operative concepts, that only our manifest concept of belief—that is, the concept we can explicitly articulate—involves metaphysically substantial commitments, whereas the operative concept of belief, the one we use implicitly, is shallow (p. 132). Yet even if we assume that operatively people are already on the right track, it is still unclear how to get from a philosophical critique to actually getting people to change the way they think about beliefs (and other mental states with similar substantial commitments).

Given that Posłajko incorporates the revision of the ordinary concept of belief into his approach, his theory acquires a dual and shifting character—his stance on the overall concept of belief is very different before and after the revision. Before the revision, the notion of belief carries substantial commitments,

and thus current belief attributions must be false. There are no such things that would satisfy the substantial commitments of the folk concept. Since the revision has not yet taken place, Pośłajko's stance currently amounts to a version of eliminativism.<sup>5</sup> After the revision, however, when only the superficial notion of belief is retained, belief ascriptions can be true in this minimal sense. He writes:

Thus, it might indeed be the case that belief attributions might be seen as substantially false if we treat them as employing the substantial notion of belief; however, the same belief attributions might be treated as being minimally true if we saw them as employing the new thinner notion of belief. (p. 141)

The claim that the folk concept of belief carries substantial but flawed metaphysical commitments provides Pośłajko with a way to justify why revising the folk concept is part of his construal of minimal non-realism. He can be read as arguing that non-realism about belief requires not only rejecting the metaphysical reality of beliefs but also recognising that the concept itself misrepresents the world (cf. p. 136).

While acknowledging this, I still find that separating entity realism from discourse realism allows for a more nuanced representation of positions, including those that do not burden folk psychology with strong commitments. From this point of view, one could reject belief realism without having to say that the folk concept is mistaken or needs fixing, since its commitments are already minimal and it might still serve its role even if the world does not contain beliefs as natural entities.

Pośłajko seems to assume that a match with the folk concept is required for realism, and a mismatch for anti-realism. For example, he notes that, in addition to showing that beliefs are not natural properties, one also needs "some sort of commitment to the idea that the folk notion of belief is in the wrong with respect to mental reality" (p. 117). In contrast, from the opposing point of view, if it turns out that "the folk concept of belief matches mental reality" (p. 152), then the position should be regarded as the realist one.

However, it seems to me that a match between the folk concept and reality is not always required for belief realism, and that anti-realism does not always imply a mismatch. Let us presume that belief could have either substantial metaphysics, that is, be a natural property, or shallow metaphysics, that is, not be a natural property. Folk psychology, or more particularly the folk concept of belief, could carry either substantial or minimal commitments. Then we get a two-dimensional matrix with four positions.<sup>6</sup> (1) Belief is a natural property

---

<sup>5</sup> Pośłajko would no doubt object to the "eliminativist" label, since he wants to preserve the term "belief" and does not claim that beliefs do not exist. However, what is preserved after the revision is only the shallow shell of a concept, while its inferential core has been eliminated (see also pp. 136, 172).

<sup>6</sup> A structurally similar two-dimensional matrix is used by Mun (2016) to distinguish positions in theories of emotion according to their metaphysical commitments and views

and folk psychology of belief is substantial. There is a match between folk-psychological notion and the metaphysics of belief. This is a position of robust realism about belief. (2) Belief is not a natural property, but folk psychology carries substantial commitments. In that case, folk psychology is mistaken about belief. This corresponds to Pośłajko's view of belief before the conceptual revision and amounts to a kind of error-theory or anti-realism about belief of an eliminativist variety. (3) Belief is a natural property, but folk psychology has minimal commitments. As folk psychology is superficialist, the folk concept of belief does not capture the rich metaphysics of belief—there is a mismatch between the concept and mental reality. (4) Belief is not a natural property, and folk psychology is austere. This would be Pośłajko's view of belief after the conceptual revision.<sup>7</sup> In this case, the minimal folk concept matches the minimal reality of belief.

As we can see from this list of positions, there are two ways in which folk psychology could be in error given a mismatch between the folk concept of belief and reality, but not both cases are versions of anti-realism. While (2) is anti-realist about belief, (3) could still count as a version of realism about belief, since it takes beliefs as natural properties not fully captured by the folk concept. Also, of the two cases in which the folk concept matches reality, only (1) would count as realism in the straightforward sense, while (4) could be labelled either as anti-realism (if one puts weight on the idea that belief is not a natural property) or as a version of minimal realism (if one follows Pośłajko and proceeds from the “match” idea).

Compared with the earlier distinction between entity realism about beliefs and realism about folk psychology as discourse, this two-dimensional matrix focuses on the match between the metaphysics of belief and folk commitments. It contests Pośłajko's assumption that realism requires a match between the folk concept and reality, while anti-realism implies a mismatch. The matrix shows that this is not necessarily the case. This provides further support for the point that Pośłajko's inclusion of the third clause in his minimal non-realism is not strictly required by the rejection of belief realism.

Here, then, lies the tension in Pośłajko's position. The third clause (that the folk concept of belief embodies false assumptions) is the reason why Pośłajko's account is revisionary about belief. But in being so, he holds different views on the concept of belief at different times (pre- and post-revision). Basically, he moves from a version of eliminativism (position 2) to a version of realism (posi-

---

on the reference of the ordinary term “emotion”. The present matrix applies a comparable strategy to belief and folk psychology in order to distinguish between types of realism and anti-realism.

<sup>7</sup> Pośłajko discusses this option on p. 152, and concedes that it would be “straightforwardly realistic”, but dismisses it, since he regards the minimal commitments (as assumed by Bogdan, 1993; Horgan, 1993) as still too strong. However, if we assume that the minimal folk notion of belief is exactly as superficial as Pośłajko takes it to be, then his post-revision view falls into category (4).

tion 4—“realism” in his terms). As things stand, he is still a kind of eliminativist about the metaphysically loaded concept of belief, and will be a realist in the future, provided the revision succeeds.<sup>8</sup> In fact, his overall view is not easy to name, as its parts are in tension. Poślajko calls his position “minimal non-realism”, and the label “non-realism” is warranted by his stance that the current conception of belief is faulty (p. 152). However, if we proceed from his assumption that a match with reality is required for realism, then post-revision he could no longer be called a “non-realist”, as the future revised concept of belief would match reality and would no longer be faulty. Conversely, if the rejection of beliefs as natural properties is enough for anti-realism, as I suggested, then he could keep the “anti-realist” or “non-realist” label throughout. But in that case, the third clause about the folk concept being faulty is not a necessary part of anti-realism.

It is not entirely clear what commitments are left for the concept of belief after the revision, according to Poślajko. Recall that he assumes a version of the inferentialist theory of meaning, on which the meaning of a term includes epistemic application criteria and a set of licensed inferences that embody its metaphysical commitments (pp. 126–127). Having set up the distinction between these two components of meaning, Poślajko describes an alternative linguistic community that employs *only* the superficial concept of belief and remarks that “perhaps it is the one we should be using” (p. 128). If this is his suggestion, then the revised concept of belief would incur not just fewer metaphysical commitments, but no commitments at all. In that case, the notion of belief would lack an inferential component, that is, applying the term “belief” would not license any inferences. One half of the meaning of the term “belief” goes missing after the revision. What remains is just the superficial shell of the application criterion, which “specifies that we are justified to ascribe beliefs to a given subject if the subject in question satisfies fairly weak behavioural criteria” (p. 127).

However, this leaves it rather mysterious why we ascribe beliefs at all. What purpose does belief ascription serve, even if the behavioural criteria are satisfied? When discussing why the wholesale elimination of folk-psychological vocabulary would not work, Poślajko writes (p. 169): “We need a way of describing humans and making sense of their actions that is easily acquired and not too demanding of our cognitive capacities”. Indeed—but making sense of people’s actions by ascribing beliefs to them presupposes that there is some kind of connection between those actions and beliefs. If this relationship were merely justificatory and non-causal, a deeper explanation of what made the action happen would still be required. If making sense of an action involves explaining why it occurred by reference to the agent’s beliefs, then, given the widespread view of

---

<sup>8</sup> Poślajko could respond that since the operative concept of belief is already shallow, it need not be changed, and so his view on that concept remains stable. However, this does not resolve the tension. First, as we saw before, his stance towards the *overall* concept of belief clearly changes, as he moves from regarding belief ascriptions as false to regarding them as true. Second, a more convincing case needs to be made for showing that, in the case of the notion of belief, the manifest and operative concepts do really come apart.

folk-psychological explanation, the relationship would have a causal character. In fact, Pośłajko (pp. 118–123) stresses that, pace Curry (2018), on the current folk conception, beliefs really are causes. It is precisely this causal character that he proposes to remove from the concept of belief. He is then faced with the challenging task of explaining how the revised, superficial concept of belief can be compatible with the explanatory aim of belief ascription, which is to make sense of our actions.

Ascertaining the commitments that people actually incur when they ascribe beliefs might well be a task that cannot be settled from the armchair, so the following remarks are just my opinion. I think that the commitments that Pośłajko saddles folk psychology with are too strong. As already noted, he assumes that the folk are committed to beliefs being internal representational states that cause behaviour. In particular, he writes that the inferential norms comprising the concept of belief license us “to infer from the claim ‘*x* believes that *p*’ to the claim that ‘*x* is in some definite internal state which is causally responsible for *x*’s behaviour” (p. 127). It is hard to believe that a quotidian concept involves so specific inferential licences to internal mechanisms. If there are inferential links in folk psychology, those usually connect the belief to other mental states and to the state of affairs that the belief is about. Pośłajko concedes that people cannot articulate this rule but can follow it implicitly. However, this latter claim requires independent evidence—evidence that would support such strong commitments rather than merely minimal ones.

It is more plausible to assume that folk commitments about belief are rather minimal—that people have beliefs, which are about something, and that they play a causal role in relation to actions.<sup>9</sup> There is no need to assume that the folk are committed to the existence of mental representations or concrete internal mental states. It is the task of philosophy to provide an account that respects folk commitments while explaining them in an ontologically unproblematic way, rather than dismissing them or treating them as misguided. On the interpretivist view that I favour, the possession of beliefs, understood in this minimal sense, can be accounted for in terms of people being interpretable as having these contentful beliefs, and these beliefs could be causally *relevant* even if they are not causally *efficacious* internal states (Mölder, 2010, Chapter 7). That would also explain why we attribute beliefs to people—it is because we find them causally relevant to what they do.

---

<sup>9</sup> Given the availability of an interpretivist gloss, this could be understood even more minimally than Horgan’s (1993, p. 289) austere folk psychology, which holds that “the states it posits are realized in humans, somehow or other, by states that causally interact in ways that render the generalizations of FP true”.

### 3. The Bearable Lightness of Interpretation

This brings us to interpretivism, the view that having mental states is a matter of interpretation.<sup>10</sup> In its classic forms, the view was developed by Dennett (1987) and Davidson (1973; 1974). In the remaining part of the paper, I focus on Poślajko's critique of my own version of interpretivism and offer a long-overdue response to it.

While interpretivism is briefly discussed in Poślajko (2024), his stance in the book rests on a more thorough analysis of the view and of my own specific position in his earlier work (Poślajko, 2020). In that paper, he construes interpretivism as the position that interpretation is constitutive of propositional attitudes. Following Byrne (1998) and Kriegel (2010), he unpacks this as the claim that facts about mental states are judgement-dependent, that is, dependent on what mental states the interpreter would interpret one as having.<sup>11</sup> Poślajko also notes that adopting a deflationary take on the existence of mental states is the best way for an interpretivist to avoid succumbing to mental fictionalism. Interpretivism is thus seen as consisting of two main claims that are independent: "deflationary ontological realism about the existence of mental states, and an additional metaphysical assumption that mental states are judgement-dependent" (Poślajko, 2020, p. 717). He agrees with the deflationary approach but finds interpretivism "an unnecessarily convoluted version" of deflationism (Poślajko, 2020, p. 715).<sup>12</sup>

Poślajko (2020, pp. 718–719) argues that subscribing to judgement-dependence leads to troubles with the epistemology of mental states. Specifically, the issue concerns the capacities of the interpreter whose judgements are constitutive of the possession of mental states. Specifying these capacities gives rise to a dilemma. If the interpreter has ordinary capacities, then the account cannot accommodate mistaken attributions—for if ordinary interpretations are constitutive, there is no room for error. If, however, the interpreter is idealised, then she knows too much, and her judgements will be remote from those that people would routinely give. Both horns of this dilemma are rather unattractive. Poślajko (2020, pp. 719–720) acknowledges that my version of interpretivism does not encounter the dilemma; however, he argues that my account rejects

---

<sup>10</sup> In this section, I write broadly about mental states and set aside the differences between propositional attitudes (such as belief) and other types of mental states.

<sup>11</sup> I should note that Byrne (1998, p. 204; see also p. 206) does not take the judgement-dependent analysis to express the idea of one set of facts being constituted by another. He formulates the analysis as a biconditional and points out that both sides stand on equal footing, with neither being more basic.

<sup>12</sup> In another paper, Poślajko (2023, p. 928) repeats this point, claiming that interpretivism incorporates "a problematic component in the form of the claim about attitudes being constituted by the process of interpretation". He regards this component as "an unnecessary complication which makes the interpretivist version of the general deflationary/superficial approach far less palatable than the original generic dish".

judgement-dependence and thus becomes indistinguishable from a general deflationist approach, losing its interpretivist flavour.

Let us examine this claim in more detail. To begin with, I will briefly summarise my own position, the ascription theory (Mölder, 2010; 2017). It is an interpretivist account of the possession conditions of mental states, that is, of what it takes to have them. My claim is that people have those mental states that are canonically ascribable to them. Canonical ascription sets a standard for correctness, so not every ordinary ascription would count as true. A canonical ascription is one that integrates information from various ascription sources in a coherent way and would not require revision if it were actually made. The second clause relaxes the condition to a range of acceptable ascriptions that fall short of the ascription an ideal interpreter would make. In this way, the canonical ascription is, in principle, within the reach of ordinary people. The mental properties whose possession is fixed by canonical ascription are conceived in a deflationary manner. These properties are pleonastic: metaphysically, to have them amounts to nothing more than being describable by mental predicates.

Avoiding the aforementioned epistemological issues does not hang on the rejection of judgement-dependence. The notion of canonical ascription does the work of evading the dilemma. Nevertheless, Pośłajko argues that my account fails to make mental properties dependent on interpretation, since I reject judgement-dependence, which he defines as the view that “it is both necessary and sufficient that one is interpreted as possessing mental states in order to possess them” (Pośłajko, 2020, pp. 719–720). However, I do not think that I reject judgement-dependence in the relevant sense, once it is suitably clarified.

First, I never rejected the view that interpretability is necessary for the possession of mental states, as I regard interpretability as ineliminable from the possession condition. Second, the answer to the question of whether interpretation is sufficient for the possession of mental states depends on how much is packed into the notion of interpretation or ascription (Mölder, 2010, pp. 77–78). The view I reject is the one according to which the act of ascription alone suffices to constitute mental state possession, as such an act could be whimsical, showing no regard for the evidence. I hold that the contribution of the world and of the subject of ascription is also important. This contribution enters the picture through the ascription sources, which include information about the subject’s behaviour, the stimuli affecting her, her personal background, and other mental states ascribable to her. But when the notion of interpretation is understood more broadly, as including the ascription sources and the ways of making them coherent in accordance with folk principles of interpretation, then such interpretation would be both necessary and sufficient for the possession of mental states. On this broader reading, the conditions that make interpretation appropriate, such as coherence with the evidence, must also be included in order for interpretation to be sufficient. The way I phrased the connection between possession and interpretation (for ascription is a kind of interpretation)—“to have a belief that *p* is to have it canonically ascribable”—makes it clear that canonical ascribability is

both necessary and sufficient for belief possession (Mölder, 2010, p. 171; also quoted by Pośłajko, 2020, p. 720).

I have also argued that mental states are recognition-dependent, and this can be taken as a reason for offering an interpretivist account of their possession (Mölder, 2010, pp. 158, 202–204; Mölder, 2024). The idea is that a state qualifies as mental only when it is specifiable in mental terms. However, this specification is not intrinsic to the state itself. Rather, the applicability of mental terms depends on various external factors (as captured by the ascription sources), and interpretation is required to determine which mental terms are appropriate. Thus, possessing a mental state depends on the interpreter's ability to recognise someone as being in a state with a specific mental description. Recognition-dependence can be seen as a form of judgement-dependence.

Still, I have not yet answered the concerns raised by Pośłajko (2020). There are two worries. One is the “no-distinction challenge”, as interpretivism seems not to differ from other deflationary accounts of the mind. The other is the worry that my account does not explain how exactly mental states such as beliefs depend on interpretation. I will address these concerns in turn.

The no-distinction challenge asks how the ascription theory differs from generic deflationism, which states that an object *O* exists if and only if the application conditions of the term “*O*” are met. What is the distinctive role for interpretation to play in the case of the mind? Pośłajko (2020, p. 721) also points out that one could be a deflationist about boulders. What is peculiar about boulders is that they do not form a natural kind, but are just a category of large rocks, whose diameter exceeds 25,6 cm. One could thus claim that boulders exist, since the application conditions of “boulder” are fulfilled, and that the property of being a boulder is pleonastic. This does not commit the boulder-deflationist to the claim that boulders are constituted by our acts of classification. Could an interpretivist maintain that beliefs (and other mental states) differ from boulders? If not, then it cannot be said that either is dependent on or constituted by interpretation.

Indeed, to avoid collapsing into general deflationism, interpretivism needs to maintain its distinctive interpretive element. It is also not a good idea to be a global interpretivist. Instead, its range of application should be kept more restricted. For if everything depends on interpretation, then interpretation-dependence would not be distinctive anymore. I do not object to applying interpretivism to certain other social, non-mental entities that may require some form of canonical determination, provided we care enough about clearly delineating their boundaries. This does not mean that interpretivism generalises to boulders. A distinction can be drawn between fixing the application conditions of a term and fulfilling those conditions. In some cases, fixing the application conditions involves conventional elements, but once established, their application is automatic and depends solely on empirical facts. In other cases, however, even after the application conditions are fixed, fulfilling them requires interpretation. Thus, even if being a belief and being a boulder are both pleonastic properties, the relevant distinction lies between cases where meeting the application conditions

of a term is straightforwardly empirical (as in the case of boulders) and cases where meeting them is sensitive to interpretation, as with mental states.

Hence, I would stick to the claim that the possession of mental states is constituted by interpretation, and this is what makes interpretivism distinct from other deflationary views. The crucial point is that interpretation is required to fulfil the application conditions. A canonical interpretation seeks the most coherent ascription (within a reasonable range) given the ascription sources and their relevance. This is based on the ascriber's competence, which is uncodifiable. It cannot be made fully explicit, and this is why the interpreter is ineliminable. In other words, a proper application of a mental term *is* already an interpretation.

It is crucial to stress that mental state ascription is not simply a matter of applying fixed criteria in a straightforward manner. Making an ascription requires constructing a coherent arrangement of information from various ascription sources. This arrangement is unique in each case of applying a mental term, such as "belief". This contrasts with cases that do not involve interpretation—cases in which a term applies automatically once its application conditions are fixed and empirically satisfied.

Posłajko's second challenge takes the form of an invitation to explain the sense of dependence that holds between interpretation and mental states in my view: "what is the sense of the notion of metaphysical dependency that is adopted by the proponent of interpretivism once the idea of judgement-dependence, in the sense which requires an actual interpretation to take place, is rejected?" (Posłajko, 2020, p. 722). This challenge is motivated by the fact that my account does not require actual ascription but instead settles for interpretability or ascribability. Posłajko's worry is that a merely possible interpretation is metaphysically too light to support the constitution claim. I have already explained that I do not reject the idea that interpretability (considered broadly) is necessary and sufficient for the possession of mental states. The question now is whether it is strong enough for constitution.

I believe it is. Remember that mental properties are pleonastic. Having pleonastic properties is a lightweight affair, and ascribability, although light, is strong enough to secure what it takes to have pleonastic properties. Talk in terms of having pleonastic properties can be replaced by talk about the applicability of predicates. Thus, the issue boils down to the applicability of mental predicates. In my reply to the first challenge, I already explained the constitutive role of interpretation in this. To ensure that mental predicates are applicable to someone, (mere) canonical ascribability is exactly at the right level of strength.

Pleonastic facts, such as having a pleonastic property, do not require constitution in any more substantial metaphysical sense. What is needed is simply that it is possible for an individual to fall under a specific mental description, given the right conditions. Not just any mental description will do; the conditions of canonicity must be met. In every case where someone is describable in mental terms, it is in virtue of the pleonastic fact that a relevant mental state is canon-

ically ascribable to them. In short: lightweight facts call for a lightweight form of constitution.

If my responses to Pośłajko's challenges are successful, this will have implications for his overall argumentative strategy. He rejects the deflationary version of interpretivism in favour of his minimal non-realism, on the grounds that the former incorporates the claim of constitutive interpretation-dependence. But if I am right that any ascription of mental states or application of mental terms already involves interpretation, then interpretability—although lightweight—must also be an inescapable component of minimal non-realism.

#### 4. Conclusion

This paper has examined Pośłajko's minimal non-realism about belief, which combines three claims: that beliefs exist in a deflationary sense, that beliefs are not natural properties, and that the folk concept of belief is mistaken and requires revision. I argued that the third claim introduces unnecessary complications into his account, to the extent that it is unclear whether the position still qualifies as anti-realist. The more nuanced framework provided in the paper shows that one can reject belief realism without committing to the view that the folk concept is mistaken. This alternative view does not regard beliefs as natural properties but preserves the explanatory usefulness of folk psychology.

I disputed Pośłajko's assumption that the folk concept of belief involves strong metaphysical commitments to internal causal states. When he rejects these commitments, it becomes mysterious why people attribute beliefs at all if they bear no causal relevance. Instead, I propose that the shallow application criteria for the ordinary concept of belief are largely consistent with its minimal metaphysical commitments. This speaks in favour of a deflationary account of belief that retains its explanatory role in everyday discourse without invoking substantial metaphysics.

In the second part of the paper, I responded to Pośłajko's challenges to interpretivism. I clarified my stance on the judgement-dependence of mental states on interpretation. I argued that interpretivism does not lose its distinctive flavour even within the deflationary framework. Interpretability is constitutive of mental state possession, because only through interpretation can one fix which pleonastic mental properties the subject has. The condition for possessing a pleonastic property is precisely the condition under which it can be ascribed. Finally, I pointed out that canonical ascribability is not too lightweight to ground the constitution claim; on the contrary, it is exactly right for securing that someone is describable in such-and-such mental terms. So I think that interpretivism as a position is doing quite well—any reports of its death (as in Pośłajko, 2023, p. 929) are exaggerated, and the use of obituary discourse is premature.

## REFERENCES

- Bogdan, R. J. (1993). The Architectural Nonchalance of Commonsense Psychology. *Mind and Language*, 8(2), 189–205.
- Byrne, A. (1998). Interpretivism. *European Review of Philosophy*, 3, 199–223.
- Curry, D. S. (2018). Beliefs as Inner Causes: The (Lack of) Evidence. *Philosophical Psychology*, 31(6), 850–877.
- Davidson, D. (1973). Radical Interpretation. *Dialectica*, 27, 313–328.
- Davidson, D. (1974). Belief and the Basis of Meaning. *Synthese*, 27, 309–323.
- Demeter, T. (2009). Two Kinds of Mental Realism. *Journal for General Philosophy of Science*, 40(1), 59–71.
- Demeter, T., Parent, T., Toon, A. (Eds.). (2022). *Mental Fictionalism: Philosophical Explorations*. London; New York: Routledge.
- Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.
- Drayson, Z. (2022). What We Talk About When We Talk About Mental States. In T. Demeter, T. Parent, A. Toon (Eds.), *Mental Fictionalism: Philosophical Explorations* (pp. 147–159). London; New York: Routledge.
- Fernández Castro, V. (2023a). An Expressivist Approach to Folk Psychological Ascriptions. *Philosophical Explorations*, 27(1), 86–105.
- Fernández Castro, V. (2023b). Folk Psychology Without Metaphysics: An Expressivist Approach. *South African Journal of Philosophy*, 42(2), 128–143.
- Haslanger, S. (2012). *Resisting Reality: Social Construction and Social Critique*. New York: Oxford University Press.
- Horgan, T. (1993). The Austere Ideology of Folk Psychology. *Mind and Language*, 8(2), 282–297.
- Jackson, F., Pettit, P. (1990). Program Explanation: A General Perspective. *Analysis*, 50(2), 107–117.
- Kriegel, U. (2010). Interpretation: Its Scope and Limits. In A. Hazlett (Ed.), *New Waves in Metaphysics* (pp. 111–135). London: Palgrave Macmillan.
- Lewis, D. K. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61(4), 343–377.
- Mölder, B. (2010). *Mind Ascribed: An Elaboration and Defence of Interpretivism*. Amsterdam: John Benjamins.
- Mölder, B. (2017). Mind Re-Ascribed. *Studia Philosophica Estonica*, 10(2), 55–104.
- Mölder, B. (2019). Vaimuseisundid pole loomulikud liigid. [Mental States Are Not Natural Kinds]. In R. Altnurme (Ed.), *Humanitaarteadused ja kunstid 100-aastases rahvuslikoolis* (pp. 173–185). Tartu: Tartu University Press.
- Mölder, B. (2022). Interpretivism and Mental Fictionalism. In T. Demeter, T. Parent, A. Toon (Eds.), *Mental Fictionalism: Philosophical Explorations* (pp. 320–339). London and New York: Routledge.
- Mölder, B. (2024). Mental Kinds and Practical Realism. In A. Mets, E. Lõhkivi, P. Mürsepp, J. Eigi-Watkin (Eds.), *Practical Realist Philosophy of Science: Reflecting on Rein Vihtalemm's Ideas* (pp. 143–159). Lanham, MD: Rowman & Littlefield (Lexington Books).

- Mun, C. (2016). Natural Kinds, Social Constructions, and Ordinary Language: Clarifying the Crisis in the Science of Emotion. *Journal of Social Ontology*, 2(2), 247–269.
- Posłajko, K. (2020). Can Deflationism Save Interpretivism? *Philosophia*, 48(2), 709–725.
- Posłajko, K. (2023). Folk Psychological Models and the Metaphysics of Belief. A Reply to Curry. *Philosophia*, 51(2), 919–931.
- Posłajko, K. (2024). *Unreal Beliefs. An Anti-Realist Approach in the Metaphysics of Mind*. London: Bloomsbury Academic.
- Thomasson, A. L. (2015). *Ontology Made Easy*. Oxford: OUP.
- Wright, C. (1995). Can There Be a Rationally Compelling Argument for Anti-Realism About Ordinary ('Folk') Psychology? *Philosophical Issues*, 6, 197–221.

BRICE BANTEGNIE \*

## DISPOSITIONALISM AND REALISM

**SUMMARY:** In this commentary, I invite Krzysztof to adopt a dispositionalist analysis of the belief concept and to draw the distinction between what is real and what is not in a way that does not beg the question against dispositionalism. Because his arguments against the dispositionalist analysis of the belief concept fall short, he certainly could do so.

**KEY WORDS:** dispositionalism, eliminativism, beliefs, mental causation.

### 1. Introduction

Having been a card-carrying eliminativist in the past, I know how hard it is just to get philosophers to give the view a fair hearing. I am therefore impressed not only by Krzysztof's persistence in pursuing these questions but also by his accomplishments, which I certainly cannot claim, the most recent of which is this book. For this reason alone, it would be both an honor and a pleasure to comment on this book. But, in addition to that, in writing this book, Krzysztof has done philosophers of mind a great service. The slim book we have in our hands is an up-to-date investigation into the reality of beliefs, or lack thereof, which should and most probably will serve as the starting point for any serious work on the metaphysics of beliefs. Therefore, if I happen to make some interest-

---

\* University of Southampton, Department of Philosophy. University of California, Riverside, Department of Philosophy. E-mail: bricebantegnie@gmail.com. ORCID: 0000-0003-0902-4829.

ing and, maybe, some true claims in this commentary, it will be due in no small part to Krzysztuf's work.

Krzysztuf tells us that his book has three aims. These are, roughly: (i) to provide a new logical space of positions on the metaphysics of beliefs, (ii) to argue for a specific position in this logical space, minimal non-realism, and (iii) to state the relationship between this position and common sense. Unfortunately, I cannot cover all these points in depth. I will focus on (i) but I will also draw some of the consequences of what I will say for (iii). And even with this limited scope I will keep many of my questions and worries for myself. I should also say that given the nature of this contribution, I will allow myself to be less guarded when making claims, and more speculative. Hopefully this will make for a more lively read.

I favor a dispositionalist view of our mental state concepts, that is, I take mental states to be best conceptualized as dispositions to behave in specific ways in specific circumstances. Two comments on this. First, I do not say "behaviorist". Behaviorism is a dispositionalist view of a reductive kind, that is, to wit, according to behaviorism, mental states concepts are best conceptualized without using mental terms. I think, like virtually everyone else, that this kind of dispositionalism is a dead end. When I write "dispositionalism" to denote the view I favor, one should then hear "non-reductive dispositionalism". Second, if I say "best conceptualized" it is not because I intend to do any kind of conceptual engineering. I am interested in understanding what we ordinarily mean by "belief", not in reforming our uses. It is just that, given that the search for necessary and sufficient conditions is a goose chase, some choices have to be made if one wants to arrive at a reasonably precise characterization of beliefs. And these choices are going to be driven mainly by one's philosophical position on other topics, i.e., there is going to be a best way to conceptualize beliefs depending on one's philosophical commitments.

Given that I favor dispositionalism, my reader will not be surprised to hear that I am going to focus on what Krzysztuf writes about dispositionalism in this book. I think, though, that this will allow me to uncover a number of problems with the ideas Krzysztuf defends.

I take it that the first and still to this day best articulation of dispositionalism ever published has been given by Ryle in *The Concept of Mind*. I will not, however, discuss Ryle here.<sup>1</sup> So when I talk about dispositionalism, I talk about the view as just defined and I express my vision of it.

The way I will proceed is that I will follow the book and pick up on the passages that are relevant for my purpose. I could have chosen, instead, to go directly to the passage(s) which I think is (are) central to what I have to say. But I think this way of proceeding is better for several reasons. First, it will hopefully make for a more pleasant read (more than what precedes, hopefully). Second,

---

<sup>1</sup> I should note though, that, unfortunately, Krzysztuf endorses the old saw that Ryle was an analytic behaviorist (6) (cf. Graham, 2023). Even a cursory reading of *The Concept of Mind* will show that this is not the case (for recent scholarship on this, see Crawford, 2013; Kremer, 2017).

I hope this will show my fellow dispositionalists that, though, from a dispositionalist perspective, things start getting problematic on page 74 at the latest, they should read on, as there is a lot to get out of reading the book in its entirety. Third, this should make it easier for the reader of this commentary to understand what I am driving at. Without further ado, then, let us start.

## 2. That Dispositionalism Is Not (Necessarily) a Non-Realist View

In the first four chapters of the book Krzysztof describes a new logical space of positions on the metaphysics of beliefs. The idea, in a nutshell, is to say that even if one agrees that some of the claims we make about beliefs are true, i.e., beliefs exist, one can still claim that they are not real. (It is therefore a non-realism but it is minimal; hence “minimal non-realism”). And this turns, roughly, on whether beliefs are natural properties—that is, whether (i) they constitute a natural kind, (ii) they are causally efficacious and (iii) there is a naturalistic story to tell about their content. Then, before arguing for negative answers to these questions in Chapter 6 (and therefore for minimal non-realism), Krzysztof, in Chapter 5, looks at recent history to find precedents to his view. His view, he tells us, is “modelled on some of the existing positions”.

One of these positions is dispositionalism. Here, his reasoning seems to be as follows. Dispositionalists take it that beliefs are dispositions. This means two things. First, beliefs are superficial features whose existence is easily ascertainable. And indeed, they exist. Second, beliefs are not “cogs in a cognitive architecture” (to use Devin Curry’s helpful turn of phrase; Curry, 2018) and are therefore not explanatory. It follows from this that they are not natural entities. As a consequence, they are not real.

To my mind, something went wrong in this chain of reasoning. Dispositionalists say that beliefs are dispositions. Krzysztof contends that, *just because* they say this, they are then committed to saying that beliefs are not real. Now, needless to say, he is free to draw the line between the natural and the non-natural where he wants to draw it, but then, if he goes on saying that dispositionalists are therefore, ipso facto, anti-realists about belief, it seems to me that he is using the word “anti-realist” in such a way that, I think, dispositionalists should leave the boat. Why would beliefs not be real *just because* they are dispositions? Why would dispositionalists, just because they are dispositionalists, be considered anti-realists?

If one wants to continue characterizing the debate between anti-realism and realism in terms of the distinction between the natural and its negation, one does best, it seems to me, to draw this distinction in a way that is sensitive to the kind of things to which it applies and, consequently, in a way that does not immediately put dispositions on the non-natural side. Is there a way to do that? There is. I will not give a complete account because this is not my goal here but here are two (partial) options:

- (i) According to dispositionalists, believing is a multi-track disposition, that is, a disposition composed of distinct dispositions. (Think of dispositions as being specified by counterfactual conditionals and think of a multi-track disposition as being specified by a conjunction of such conditionals).

First option:

- Ask the following question: Do these dispositions cluster together, that is, do they co-occur with the same entity in a law-like way?
- How do you answer this question? Look at (scientific) psychology. If the answer is yes, then the multi-track disposition is a natural property and beliefs are real. If not, then it is not, and beliefs are not real.

- (ii) Dispositions are often taken to have categorical bases. The categorical basis of a disposition is the property or the set of properties such that it is in virtue of having this property/these properties that the entity has this disposition.

Second option:

- Ask the following question: Is this categorical basis the same across believers?
- How do you answer this question? Look at (cognitive) neuroscience. If the answer is yes, then the multi-track disposition is a natural property and beliefs are real.

I invite Krzysztof to revise his conception of what is real in light of these suggestions in order to make room for dispositionalism.

### 3. Looking for the Non-Dispositionalist “Superficial” Analysis of the Belief Concept

The fact that Krzysztof assumes that in order to be real beliefs would have to be internal states is clear from a number of passages in the book. For example, commenting on his negative verdict on the viability of eliminativism (which, in this context, he calls “irrealism”), he writes:

even if we had good reasons to reject both irrealism about beliefs and the negative ontological claim about them, this would not in itself show that there are good reasons to think that there are indeed beliefs *in the sense that we have internal, causally active states which are vehicles of propositional contents*. (Pośtajko, 2024, p. 31, emphasis added)

What makes it especially clear that he makes this assumption, however, is the fact that “the position that can be said to relate to the strong realist view most closely [...] is the position of representationalism about beliefs” according to which—leaving details aside that I do not think are relevant here—beliefs are mental representations.

Thus, in Krzysztof's rendering of the debate over realism about beliefs, his view, minimal non-realism, has two rivals: dispositionalism and representationalism about beliefs. Each view disagrees with a part of Krzysztof's view. Krzysztof's view is close to dispositionalism in that he holds that beliefs are not real (I have just argued that this is not the right way to conceive of the distinction between the real and the non-real as it applies to mental states, but, for the sake of the argument, I now go along with Krzysztof). However, when it comes to analyzing the concept of belief, Krzysztof parts ways with dispositionalism. Krzysztof's view is close to representationalism in that he holds that the belief concept is not a dispositional concept. However, Krzysztof parts ways with representationalism in that he contends that beliefs are not real. The three views can be contrasted along two lines: (i) whether beliefs are real and (ii) whether they analyze beliefs as dispositions or not. Here is what it looks like:

**Table 1**

*"Minimal Non-Realism" Compared to Its Rivals*

|                       | Dispositional Analysis<br>of the Belief Concept | Realism |
|-----------------------|-------------------------------------------------|---------|
| Dispositionalism      | Yes                                             | No      |
| "Minimal Non-Realism" | No                                              | No      |
| Representationalism   | No                                              | Yes     |

*Note.* Table prepared by the author.

Remember that all three views agree on the fact that beliefs exist, which means, in this picture, that, in order to determine whether someone believes something, "we only need to engage in trivial conceptual analysis and empirical observations" (p. 37). Now, it is clear that the content of the dispositionalist concept of beliefs can deliver this: the conceptual analysis is not complicated, and one can check by empirical observation—I suppose—whether someone has the requisite multi-track disposition.<sup>2</sup>

It is unclear, however, how Krzysztof delivers this. He seems to want to distinguish his view from dispositionalism, but, if his analysis of the concept of belief is not the dispositionalist one, what is this analysis? (If, as it seems, he agrees with the representationalist on that point, the latter faces the same ques-

---

<sup>2</sup> If one wants to say, as I have done in the previous section, that the dispositionalist can distinguish between the existence of beliefs and their reality, we have to imagine that the concept of beliefs places very little demand on someone to be deemed a believer. We can imagine for example a disjunction of disposition to assert things or to act in a way that would be successful if the belief was true or something like that. Schwitzgebel speaks of the concept of belief as being a stereotype (2002). That is congenial to this picture.

tion). Maybe I missed the relevant passage, but as far as I can see, he never states what it is. On page 65, Krzysztof does talk about “common-sense criteria” for the ascription of a belief that  $p$ , but the criteria he gives (having claimed that  $p$ , answering affirmatively when asked whether  $p$ , “exhibit[ing] some behavioral dispositions related to this belief”) are all things a dispositionalist could agree with. He concludes by writing “and so on”. It may be, then, that a dispositionalist would disagree with what he has in mind. But in order to make this determination, I would need to know what other criteria he is thinking about.

Answering this question is supposed to be easy. It is “easy ontology”, according to Thomasson (2015), who is Krzysztof’s main source of inspiration. Do tables exist? Just check whether some entities in the world have a plane surface that supports stuff, such that you can put things on it. If there are, then tables exist. If there is a “superficial” analysis of the belief concept, which is not the dispositional analysis, however, it is unclear to me what it could be. Here, things do not seem to be easy. Maybe I put too much stress on this source of inspiration, and I should read Krzysztof differently, but I do not see a better way. Hopefully, he will offer some clarifications that will help. (I make a suggestion in the next section).

#### 4. Packing Concepts is Bad for One’s Back

Up until now, I have only been discussing the new logical space of positions on the metaphysics of beliefs that Krzysztof aims to provide. That is his first objective in the book. His second objective is to argue that beliefs are not natural and therefore not really real. I will not have anything to say about this. I am too much of a dispositionalist to spend time discussing Krzysztof’s argument against the representationalist theory of mind. I also find this theory wrong, and I have arguments of my own, more fundamental than Krzysztof’s, against this theory.<sup>3</sup> What I will now do is take a look at what he says as a way to reach his third aim, namely, “to inquire into the relation between the minimal non-realist approach and commonsense” (2). I follow my thread of discussing what Krzysztof says about dispositionalism. This is going to lead us to that.

As I just said, Krzysztof does not tell us what a superficial analysis of the belief concept, which is not the dispositional analysis, is. That he does not explicitly say what it is, is surprising. What he says about what he calls “common-sense functionalism”, which is the view presented by Jackson and Pettit as a response to eliminativism, might help: it is about people satisfying a number of platitudes, namely asserting that  $p$  and other things (p. 36). Close to dispositionalism, then, but not dispositionalism: functionalism. Krzysztof does not endorse this view. But maybe he should. I will not focus on this, though. Instead, I will focus on what Krzysztof explicitly does, namely criticizing the dispositionalist analysis of

---

<sup>3</sup> They can be found in (Bantegnie, 2026). If reviewer permits, this paper will be published at some point. The interested reader should feel free to reach out to me.

the belief concept. What I will say below may or may not show that the way I have laid out Krzysztuf's view about the logical space of positions should be revised. Given that I am not entirely sure what he says, I will not examine the question, preferring to leave all the interpretive threads loose instead of tying up the loose ones. He will most likely pick them up. This should make for interesting discussion.

In Chapter 7, he argues that part and parcel of our belief concept is the idea that beliefs are concrete internal states entering into causal relations. To me this is the most puzzling part of the book. I have a hard time understanding why Krzysztuf argues for this conclusion because I do not see why he thinks that he needs to do that. Actually, there seems to me to be clear reasons why he should *not* argue for this view (more on this below). I am genuinely looking forward to hearing what he will say about this but, in the meantime, allow me to explain why I think he should not argue for this view, as far as I can see.

Krzysztuf seems to think that dispositionalism is a threat to his view. In Chapter 7, and in contradiction to what he says in Chapter 5, he seems to think that a dispositionalist can say that beliefs are real. I have criticized what he says in Chapter 5 and argued for this very claim above, in Section 1. I therefore agree with him. And yet I find it very hard to understand Krzysztuf's shift. Here is what he seems to think in both chapters.

- In Chapter 5, he seems to take dispositionalists to think that the folk conceive of mental states as internal states entering into causal relations and to argue that this is not so: beliefs are dispositions. Therefore, according to dispositionalists, or so they are presented, the folk are mistaken; hence the fact that dispositionalism is seen as a form of non-realism.
- In Chapter 7, he seems to take dispositionalists to think that the folk conceive of mental states as dispositions. And beliefs are indeed dispositions. Therefore, according to dispositionalists, or so they are presented, the folk are right; hence the fact that dispositionalism is seen as a form of realism.

With this in hand, how are dispositionalists, as presented in Chapter 7, seen as a threat to Krzysztuf's minimal non-realism? Krzysztuf, let's assume, has shown in Chapter 6 that beliefs are not internal states entering into causal relations. But, if this is not how the folk conceptualize beliefs, then the folk are not wrong; ergo, Krzysztuf's non-realism is mistaken.

I do not see what else Krzysztuf could be saying in the first Section of Chapter 7 and several passages point towards this interpretation, the last paragraph of the Section being maybe the clearest example:

Minimal non-realism, in opposition to neo-behaviourism and interpretivism, is intended to be a version of anti-realism about belief. But, as the considerations presented above show, adherence to the claim that belief properties are not natural in Lewis's sense is not in itself enough to classify a position as being a specimen of belief anti-realism. What is also needed is some sort of commitment to the idea that the folk notion of belief is in the wrong with respect to mental reality. (Pośajko, 2024, p. 117)

However, this strikes me as wrong-headed. Krzysztof offers us a criterion for what is real and what is not: it is about the property being natural or not. What the folk think about this is not relevant. Whether this is a thought that they have just in virtue of thinking about beliefs (it is part of their concept of belief), or whether this is a thought that they have about beliefs that is not constitutive of what beliefs are according to them (this is an empirical view), what the folk think is irrelevant either way. To put it otherwise, if we adopt the characterization Krzysztof gives of what is real, whether non-realism is true has nothing to do with what the folk think.

But that is not all: if Krzysztof says that contrary to what people think is constitutive about belief, beliefs are not internal entities entering in causal relationships, if, as he says, “there is some error involved in this concept” (p. 128), how can we still say that there are beliefs in the first place. That is, how does minimal non-realism not collapse into classical eliminativism? I do not see how. And if it does, it is Krzysztof's whole enterprise of offering an alternative form of non-realism that collapses; hence my claim that he should not argue for the view that part and parcel of our belief concept is the idea that beliefs are concrete internal states entering into causal relations.<sup>4</sup> This is not good for him!

## 5. That the Dispositionalist Analysis of Belief Concepts is Still Standing

As I argued in Section 1, a dispositionalist can make perfect sense of the distinction between natural and non-natural mental entities. A dispositionalist can therefore be a minimal realist, or so it seems. Maybe Krzysztof should adopt dispositionalism then. The problem, however, is that he gives arguments against the dispositionalist analysis of the belief concept. Therefore, if I wanted to con-

---

<sup>4</sup> Here I continue to mean by “eliminativism” what Krzysztof means by it himself when he introduces it in Chapter 2. There he writes: “[...] in its purest form, eliminative materialism is committed to the claim that all elementary attributions of beliefs are false and beliefs do not exist” (p. 14). Later, in Section 7.5, titled *Truth in Eliminativism* he seems to be implicitly using a different characterization of eliminativism, as the contrast which he draws with his view, that he calls “revisionist-preservationist” (p. 136), and eliminativism, suggests that eliminativism is the view that we should stop using the belief concept. This is not what I mean by “eliminativism”.

vince him that he should join me in the dispositionalist camp, for his own good, I should address his worries. Unsurprisingly, I think the arguments he gives are weak. Before drawing this commentary to a close, I will show why.

The first argument he takes from Frances Egan, in a paper from the nineties (1995).<sup>5</sup> I might as well cite the passage he cites and whose argument he endorses:

We often say things such as “He believed he was about to be fired because he saw a confidential memo that criticized his job performance” and “She quit smoking because she believed it was affecting her health”. There is no reason to suppose “because” functions any differently here than in locutions which are clearly causal, such as “The fire started because the electrical system was overloaded”. In claiming that common sense explanations of belief fixation and action are not causal, despite appearances, the neo-Rylean assumes a rather heavy burden of proof. (Egan, 1995, pp. 187–88)

Krzysztof concurs, saying that we should take the folk idiom at face value and not attempt to paraphrase it. The first comment I would like to make on Egan’s objection is that she presents a mistaken view of dispositionalism in this passage. Dispositionalism need not and probably should not say that belief fixation is not causal: a dispositionalist has no good reason, just in virtue of being dispositionalist, to say, to paraphrase, that the fixation of the multi-track dispositions that beliefs are is not causal. (Why would not the acquisition of a disposition be caused, I wonder...). The second thing to say is the following. In many contexts, “because” is used to express a non-causal explanation, as when we say, “The temperature in the room has dropped because the kinetic energy of the air molecules has lowered” or when we say, “7 is a prime number because it is divisible only by 1 and itself”. There is a huge literature on explanations of actions by means of reasons which *are not* causal explanations. Obviously, many people do not think that these explanations just appear as causal. Therefore, I would be curious to hear Egan explain to me how she determined that explanations by means of beliefs “appear” causal. Krzysztof does allude to non-causal views of action explanations. Yet he begs the question when he writes:

this putative special character of folk psychology does not provide us with enough evidence for the claim that the word “because” functions differently in the context of explanations in which we use beliefs than it does in other forms of causal explanations. (Pośtajko, 2024, p. 120)

Second, Egan also writes:

---

<sup>5</sup> I do not know whether Egan still subscribes to what she wrote in this paper.

The case for a causal construal of belief-desire attributions does not rest solely on linguistic intuitions [...]. Of present relevance is the fact that a large and influential body of research in empirical psychology is predicated on the claim that beliefs, desire, and more permanent conditions such as character traits are implicated in causal explanations of behavior. (Egan, 1995, pp. 187–188)

Krzysztof concurs. My first comment is of the same nature as before: one of the things Egan says in this passage is completely false. She talks about explanations by means of character traits as causal. However, it is—at the very least—plausible to conceive of character traits as dispositional in nature. And, if so, explanations by means of character traits are not causal. Second, Krzysztof invokes the literature in mind-reading. It is not entirely clear to me what his argument is. He writes the following: “there is a strong consensus among psychologists that the attribution of beliefs to other subjects is used for behavior prediction” (p. 120). And this suggests that he takes it that dispositionalists cannot vindicate this. But they can: if I know that sugar is soluble at room temperature, this helps me predict what will happen to a sugar cube if I put it in water at room temperature and stir the water for a few seconds. Here once again, arguing against dispositionalism would take much more than this, namely, it would require showing that one *has* to interpret psychological findings in a way that constitutively mental states are taken to cause our actions in order to make sense of them. And this, neither Egan, in this paper, nor Krzysztof, have done (cf. Footnote 3 of the current paper).<sup>6</sup>

## 6. Conclusion

In the end, I think that the picture of the logical space of positions on the metaphysics of beliefs Krzysztof offers is mistaken (dispositionalist is not well located) because he draws on a mistaken conception of what it is to be real (this was Section 2). He has not provided us with a non-dispositional analysis of the belief concept (that was Section 3), which, even though that is awkward, is not that big of an issue. But he packs the belief concept with too many things for the good of his minimal non-realism (that was Section 4). I suggest he should go for dispositionalism and, because his arguments against the dispositionalist analysis of the belief concept fall short (that was Section 5), he certainly could. I therefore

---

<sup>6</sup> I should note that both Egan and, following her, Krzysztof assume that dispositions are not causes. I agree with this claim. And this is why I do not dispute it. An action that is the manifestation of, e.g., a belief is not an effect of the belief; it is partly constitutive of what it is to believe. One could say that a belief is a causal power the agent has, that is, a power to cause specific actions in specific circumstances. However, according to dispositionalism as I mean it here, in the wake of Ryle, the right analysis of the concept of belief is not that believers are disposed to cause their actions but that they are disposed to act. Showing as much is irrelevant to my purpose here. I will therefore stop here.

invite Krzysztof (i) to adopt a dispositionalist analysis of the belief concept and (ii) to draw the distinction between what is real and what is not in the way I did.

Krzysztof's book is very much a product of its time. This is a good thing: he updates the debate. This is also less than a good thing. First, the way he conceptualizes the distinction between the real and the non-real seems to reflect a view of science that overstresses its explanatoriness. Second, the weakness of his arguments for the thesis that we conceptualize beliefs as internal states entering into causal relationships suggests that he is beholden to this dogma of contemporary philosophy of mind.<sup>7</sup>

## REFERENCES

- Bantegnie, B. (2026). *Against a Dogma in the Philosophy of Mind* [Unpublished manuscript]. Department of Philosophy, University of Southampton. Department of Philosophy, University of California, Riverside.
- Crawford, S. (2013). The Myth of Logical Behaviourism and the Origins of the Identity Theory. In M. Beaney (Ed.), *The Oxford Handbook of the History of Analytic Philosophy* (pp. 621–655). Oxford University Press.
- Curry, D. S. (2018). Beliefs as Inner Causes: The (Lack Of) Evidence. *Philosophical Psychology*, 31(6), 850–877.
- Egan, F. (1995). Folk Psychology and Cognitive Architecture. *Philosophy of Science*, 62(2), 179–196.
- Graham, G. (2023). Behaviorism. In E. N. Zalta, U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy (Spring 2023)*. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2023/entries/behaviorism/>
- Kremer, M. (2017). “One of My Feet Was Still Pretty Firmly Encased in This Boot”: Behaviorism and The Concept of Mind. In A. Preston (Ed.), *Analytic Philosophy: An Interpretive History* (pp. 175–192). Routledge.
- Posłajko, K. (2024). *Unreal Beliefs. An Anti-Realist Approach in the Metaphysics of Mind*. London: Bloomsbury Academic.
- Schwitzgebel, E. (2002). A Phenomenal, Dispositional Account of Belief. *Noûs*, 36(2), 249–275.
- Thomasson, A. L. (2015). *Ontology Made Easy*. Oxford University Press.

---

<sup>7</sup> Thanks to Eric Schwitzgebel for helpful discussion.



ERIC SCHWITZGEBEL \*

## **SUPERFICIALISM ABOUT BELIEF, AND HOW WE WILL DECIDE THAT ROBOTS “BELIEVE”<sup>1</sup>**

**SUMMARY:** *Superficialism* about property  $X$  treats the possession of property  $X$  as determined entirely by superficial as opposed to deep facts. Belief should be understood superficially. Facts about belief are determined entirely by facts about actual and potential behavior and conscious experience—plus transitional cognitive states ultimately understood in terms of actual and potential behavior and conscious experience. Superficialism about belief excels on both intuitive and pragmatic grounds, compared to accounts of belief in terms of deep cognitive or neural architecture. Behavior-focused superficialism suggests that robots and Large Language Models may soon (perhaps already do) believe. If consciousness is also essential to belief, matters might soon become more complicated, if it becomes reasonable to wonder whether some of our most advanced AI systems are conscious. Regardless, it will be practical to describe some such systems as having what I will call “belief\*”—belief shorn of commitment to any conscious aspect—and people might forgivably forget to pronounce the asterisk. Krzysztof Posłajko should welcome this manner of thinking, though it need not be as “antirealist” as he suggests.

**KEYWORDS:** artificial intelligence, belief, Large Language Models, Krzysztof Posłajko, robots, superficialism.

---

\* University of California, Riverside, Department of Philosophy. E-mail: [eschwitz@ucr.edu](mailto:eschwitz@ucr.edu). ORCID: 0000-0003-4908-7074.

<sup>1</sup> For helpful discussion, thanks to Brice Bantegnie, Sophie Nelson, Adam Pautz, Jeremy Pober, Krzysztof Posłajko, and William Robinson.

## 1. Introduction

I will explain and defend *superficialism* about belief, discuss its application to the question of whether robots and Large Language Models believe, and suggest that Krzysztof Posłajko should agree with these positions, while weakening his commitment to “anti-realism” about belief.

## 2. Superficialism Explained

A superficial judge of character will assess someone’s clothes, bearing, and words interpreted at face value. A superficial reading of Shakespeare will likewise mostly take characters at their word and focus on the obvious aspects of each scene, rather than searching for deeper themes and motives. A superficial historical account will take actions one at a time without exploring nonobvious connections or underlying social forces. A superficial housecleaning does not touch the backsides and undersides of household items. In general, superficial treatments and understandings focus on what is readily observed and disregard or downplay what it requires work to discern. However, what counts as “the surface” can vary for different types of superficiality.

*Superficialism* about property *X* is the view that whether an entity has property *X* is determined (that is, constituted or grounded; covariance or supervenience is not enough) entirely by superficial facts about that entity, where “superficial facts” are some class of readily observed facts, in contrast with some alternative set of “deeper”, or less obvious, less readily observable facts. So, for example, a superficialist about beauty in humans might hold that beauty is skin-deep. On such a view, whether a person is physically beautiful depends entirely on the shape and coloration of their bodily surface; nothing else matters. A metaphysical superficialist about cats might hold that even robotic cats are cats, as long as those robots superficially look and act enough like cats. A different sort of superficialist might define the “surface” to include internal organs, so that a macroscopic copy of a cat is a cat, regardless of its evolutionary history or whether it has catlike DNA. A dissection goes beneath the skin, both literally and metaphorically, but not as deep as a DNA analysis.

Superficialism about belief is the view that whether an entity believes some particular proposition depends entirely on some set of “superficial” facts about that entity. For example, philosophical behaviorism (Graham, 2000/2023; Robinson, 2019) takes the relevant superficial facts to be the entity’s actual and potential outward behavior, including verbal behavior. Dennettian and Davidsonian interpretativism (Davidson, 1984; Dennett, 1987; cf. Curry, 2020; Mölder, 2010; Pautz, 2021a; 2021b) is similarly superficialist: Whether an entity is properly interpretable as believing depends entirely on facts about its behavior and environment, as they guide potential interpretative practices, and not at all (except derivatively) on brain states or underlying cognitive architecture. On such views, an entity made of pure undifferentiated balsa wood, if it would always outwardly

act and react identically to a person who believes that plaid dungarees are hideous, necessarily would also believe that plaid dungarees are hideous. (Such an entity is presumably not physically possible, but the relevant modality is conceptual possibility).

The form of belief superficialism I favor—"phenomenal" or "liberal" dispositionalism (Bantegnie, Schwitzgebel, 2026; Schwitzgebel 2002; 2013; 2021; 2026)—conceptualizes the surface more capaciously, in terms not only of behavioral but also of phenomenal and cognitive dispositions. *Behavioral* dispositions are just the dispositions countenanced by philosophical behaviorism. *Phenomenal* dispositions are dispositions to have particular conscious experiences (e.g., feelings of surprise, inner speech experiences). *Cognitive* dispositions are dispositions to enter cognitive states that are in turn definable in terms of other behavioral, phenomenal, and cognitive states and dispositions (e.g., the disposition to adopt a new belief that logically follows from the belief in question, or the disposition to form a new intention, on the assumption that intentions themselves can be understood in terms of behavioral, phenomenal, and cognitive dispositions). Ideally, on this view, all of the terms for cognitive dispositions could in principle be eliminated en masse by Ramsification (Lewis, 1972): They serve as ontological placeholders in a network characterizable entirely in terms of phenomenal and behavioral dispositions, plus other such transitional "cognitive" placeholders. Some dispositions might mix the behavioral, phenomenal, and cognitive: For example, the disposition to *express the thought* that plaid dungarees are hideous might have behavioral, phenomenal, and cognitive aspects. Crucially, neural states and low-level functional architecture are not part of the surface.

Thus, on this view, to believe that plaid dungarees are hideous is just to be disposed to say they are hideous if asked, to refuse to wear them, to feel revolted upon seeing them, to reduce one's opinion of a friend's sartorial taste upon seeing that friend in plaid dungarees, and so on. All such dispositions hold only *ceteris paribus* (all else being equal or normal, absent countervailing influences) or given implicit conditions. One might not express one's negative opinion to a sensitive friend proud of their new purchase. The relevant dispositions are not a finitely specifiable set, and ordinary folk psychology enables us to devise or recognize new ones on the fly: If the believer in question wishes to revolt a suitor, wearing plaid dungarees might strike them as just the right choice. The disposition to make that choice in those circumstances is not only *consistent* with believing them to be hideous (springing the *ceteris paribus* clause, so to speak) but partly *constitutive* of having the belief (which includes the disposition positively to wear them under certain rare conditions).

Anyone possessing the pattern of dispositions I began to express in the previous paragraph, and which ordinary language users could further articulate with good-enough agreement, if that person has that pattern of dispositions robustly, across a wide variety of situations, thereby believes that plaid dungarees are hideous. There is nothing more to it. It does not matter what low-level cognitive architecture they have, as long as it robustly supports this dispositional pattern. It

does not matter what brain states they are in, or even whether they have a brain at all, as long as their physiology supports this pattern of behavioral, phenomenal, and cognitive dispositions. It does not matter what their causal history is, or whether they even have a causal history (unless having a causal history of the right sort is necessary for having the dispositions, for example, on some versions of externalism about phenomenology and linguistic meaning; e.g., Dretske, 1995; Putnam, 1975).

Beliefs, on this view, are structurally like personality traits as we ordinarily think of personality traits. To be an extravert, or to be narcissistic, is nothing more or less than to be disposed to act and react—behaviorally, phenomenally, and cognitively—in the manner characteristic of an extravert or narcissist. Extraverts and narcissists have brains and deep cognitive architectures that are responsible for those dispositional patterns, but it is the dispositional patterns that constitute the trait, not the brain states or deep architectures, and any entity with the exactly the same dispositional patterns would be extraverted or narcissistic to exactly the same degree, even if their brain or architecture were somehow radically different. I hope this strikes you as an intuitively plausible metaphysics of personality traits. According to my variety of liberal dispositional superficialism about belief, the same metaphysics applies to believing that plaid dungarees are hideous, and believing that your children's happiness is more important than their grades, and believing that Confucius loved the ancient rituals of the early Zhou dynasty.

### 3. Liberal Superficialism Defended

Liberal superficialism about belief is attractive on intuitive and pragmatic grounds, and not unattractive on scientific grounds.

*The Intuitive Defense.* Intuitively, superficialism gives the right answers about cases (Pośłajko calls these “application criteria”). Imagine someone as different as possible from you in their internal architecture and representational structures, consistently with robustly possessing all the phenomenal, behavioral, and cognitive dispositions associated with believing plaid dungarees to be hideous. A science fiction creature, snooping on our internet, sees an advertisement for plaid dungarees. Recoiling in horror, it exclaims “*U y?*!”—where “*U*” is a noun regularly instantiated in the presence of plaid dungarees, “*?*” is an adjective regularly instantiated in the presence of things from which it aesthetically recoils, and “*y*” is a copula. This creature would not purchase a pair, except perhaps as a joke. It would downgrade its opinions of any human or alien who intentionally chose to wear such an item. And so on. Not only externally, but also phenomenally and cognitively, it is disposed to act and react just like a human who finds plaid dungarees hideous, despite as-different-as-possible a low-level architecture. I suggest that, intuitively, most of us would find it natural to describe this alien as believing that plaid dungarees are hideous.

Conversely, imagine an entity as similar as possible to us in low-level internal architecture but who lacks the behavioral, cognitive, and phenomenal dispositions characteristic of (I would say constitutive of) believing that plaid dungarees are hideous. They have no tendency to negatively judge those who wear them, no inclination to avoid purchasing them, no inclination to agree with a friend who calls them hideous, and so on—and further, they have these dispositions robustly. This is not a case in which the dispositions are “masked” or some unusual circumstance obtains (such as preparing a method-acting performance of a character who favors plaid dungarees). To the extent we can imagine such cases, it seems right to say that despite those details of architecture, the entity in question does not find plaid dungarees hideous.

What if the behavioral dispositions are present but the phenomenal dispositions absent or wrong? If there is no phenomenology whatsoever, we might think of it as a “zombie” case, in the sense of Kirk (1974) and Chalmers (1996). Intuitions probably diverge about whether such zombies genuinely have beliefs (Fischer, Sytsma, 2021). However, if the phenomenal dispositions are reversed (e.g., inner delight at seeing dungarees) or irrelevant (e.g., experiencing neither delight nor loathing but instead only the sound of a trumpet), the intuitive inclination to ascribe belief is likely to be seriously undercut. What if the phenomenal dispositions are present and behavioral dispositions absent? If the absence of behavior is “excused”—for example, by paralysis—that is probably enough to make belief ascription intuitive (as in Strawson’s, 1994, “Weather Watchers”). If the absence of behavior is inexplicable—experiencing passionate confidence that P is true but excuselessly having no inclination to assert it or act on it—intuitions probably collapse or diverge.

My suggestion is this: Our intuitive judgments in belief ascription track the patterns of behavioral, phenomenal, and cognitive dispositions. When they are robustly present, we are inclined to ascribe belief. When they are robustly absent, we are inclined to deny belief. When they are mixed, intuitions diverge or we treat it as an in-between case. Thus, the type of superficialism I am calling liberal dispositionalism gets belief ascription extensionally correct by intuitive standards. If we treat possession of the relevant behavioral, phenomenal, and cognitive dispositions as necessary and sufficient for belief, or even better constitutive of belief, we can respect and nicely account for that pattern of intuition.

*The Pragmatic Defense.* Philosophical theory need not always follow ordinary intuition about cases. When it does not, good pragmatic or scientific/theoretical grounds should justify that break from ordinary use. Are there pragmatic grounds for favoring a deep view or a purely behavioral superficialism?

To see why not, consider: Why do we care about what people believe? It is because we care about how they do and would act, what they do and would feel, and what inferences they do and would make—not because we care about their deep architecture (except insofar as that architecture supports what we do care about). You care what your friend thinks of plaid dungarees because you want to

know what birthday purchases would and would not delight her, her fashion preferences, what quiet judgments she might be making about your new plaid-dungaree-wearing boyfriend. You care about what she would outwardly say and do, and also—contra entirely behavioral approaches—about what she might inwardly think and feel. The superficialist account of belief in terms of behavioral, phenomenal, and cognitive dispositions tracks what we care about in belief attribution. It tracks why the word “belief” is important to us in the first place.

Suppose we were to devise a new term, *schmelief*, for people (perhaps a subclass of humans with highly unusual interior architectures, or perhaps alien creatures who have joined human society) whose behavioral, phenomenal, and cognitive dispositional patterns match ours but whose deep architecture is different. Your friend Trina believes that plaid dungarees are hideous, while your other friend Jordan schmelieves that they are hideous. Would not you want some term to capture what Trina and Jordan share in common? You would have to invent a third term—call it *lululieve*—to capture the genus to which belief and schmelief belong. *Lululieve* will turn out to be the more useful and important term for most ordinary purposes. It will also be the one you apply to new acquaintances, if you do not yet know their interior architecture. It is the more important and practical term. It is the one we will shift to using. If I am right, it is what we meant, or should have meant, by “believe” all along.

As I have already mentioned, it is not just the outward behavior that matters pragmatically but the consciousness experiences as well. Experiences of affirmation, surprise, conscious judgment, confidence, and doubt; inner speech and imagination; nervousness about *what might happen if*; facts about what does and does not occupy conscious attention; and so on—these are too central to our interests in belief ascription to warrant excluding them from the stereotype constitutive of believing. They comprise a crucial part of what we care about in ascribing beliefs. We do not want only to predict and explain outward behavior; we want our belief ascriptions also to track the target’s experienced inner life. This point is perhaps more vivid if we imagine belief-like outward behavioral dispositions alongside *contrary* phenomenology—saying that *P* while innerly thinking that not-*P*, etc.—though a pure case of this might be even harder to coherently imagine than an experientially blank philosophical zombie. Phenomenology is central not only to what we do track but to what we want to, and should want to, track in belief ascription. If so, we should agree that an entity with no conscious experiences whatsoever could at best qualify only as a borderline, mixed, or in-between believer—one who conforms to some important portion of the dispositional stereotype while sharply deviating from another important portion. However, my main aim in this paper is to defend superficialism in general, not liberal dispositionalism in particular. We care a lot about behavior—often quite a bit more about that than about inner life—and I have no strong objection against a pragmatic defense of behaviorist superficialism. Indeed, in Section 4, I will defend a close relative of that view.

*The Scientific Defense.* Some philosophers have suggested that superficialism about belief is unscientific (e.g., Quilty-Dunn, Mandelbaum, 2018). A properly scientific approach, they suggest, requires looking at the subpersonal architecture underlying belief and characterizing belief in those terms. While I certainly do not object to exploring underlying cognitive architecture, the result is a cognitive science of *humans*, or *mammals*, or some other specific biological type. If we aim to think about belief in general, as it might potentially apply to future conscious AI systems (if they can exist) or hypothetical (or actual, presumably somewhere in our giant universe) alien systems, our account of belief had better not rest too finely on contingent details of human architecture. The question of AI belief is imminent, and even a skeptic about AI belief should agree that it cuts inquiry precipitously short to rule out AI belief simply on grounds of different underlying architecture, whatever else might be true of them.

Compare again with personality traits. It would be interesting to discover that extraversion in humans is implemented by a particular neural or low-level functional architecture. But what does and should constitute extraversion is not such details of implementation. (Admittedly, there is a metaphysical rabbit-hole here related to the debate between role functionalism and realizer functionalism; Levin, 2004/2023; Pober, 2024). Personality psychology is not unscientific by virtue of staying at the dispositional surface. Neither is the treatment of belief. Indeed, as I have argued elsewhere (Schwitzgebel, 2026), the most obvious way of being a deep-structure scientific realist about belief generates pseudopuzzles, such as attempting to count the number of beliefs that are stored and accessed.

#### **4. How We Will Decide That Robots, and Maybe Language Models, “Believe”**

If behaviorism were the best approach to belief, then robots and language models would believe if (and only if) their behavioral dispositions are the dispositions characteristic of believers. Behaviorist accounts might differ concerning the relevant class of behavioral dispositions. If belief requires dispositions to bodily action in our shared world, then robots might believe, but a purely text-based language model could never believe. If linguistic dispositions are sufficient, then language models could potentially also believe (perhaps if they can pass a sufficiently rigorous Turing test; Turing, 1950). If behavioristically inflected interpretativism were correct, the same result would follow. However, to answer the question of whether robots and language models believe, as I prefer to construe the question, we must determine whether robots and language models have (the right kind of) conscious experiences.

Scientific consensus and popular opinion strongly favors the answer that artificial systems are *not yet* meaningfully conscious. However, that consensus might soon break, with popular opinion quickly shifting and some influential mainstream scientists and philosophers holding that we are on the cusp of designing meaningfully conscious or sentient AI systems (Butlin et al. 2023;

Chalmers, 2023; Colombatto, Fleming, 2024; Dehaene, Lau, Kouider, 2017). Once that break occurs, consensus is likely to elude us for a long time, with some researchers defending restrictive and/or biological views on which no familiar type of AI system could possibly have consciousness (Godfrey-Smith, 2016; Seth, 2024). Our scientific and theoretical tools will fail, their applicability contingent on contentious assumptions about consciousness that advocates of alternative views will justifiably decline to accept (Schwitzgebel, 2024).

Still, we will want to talk in certain ways about robots and language models. In particular, we will want to communicate with each other about how they are likely to respond to certain types of inputs or situations. The language of belief, and more common neighboring folk-psychological terms such as “thinks” and “knows”, will be convenient and tempting. Does this language model think/believe/know that David Lewis had diabetes? Does this nursing assistant robot think/believe/know/remember that the pills are in the second drawer on the left? With cautious neutrality about whether such AI systems really have the phenomenal and phenomenology-involving cognitive dispositions necessary for genuine belief, we might create the new dispositional concept *belief\**, which is present if and only if the system has the right behavioral dispositions, plus cognitive dispositions characterizable in terms of networks of behavioral dispositions.

A language model might then have a *belief\** that *P* (for example *belief\** that Paris is the capital of France or *belief\** that cobalt is two elements right of manganese on the periodic table) if:

- it would consistently output *P* or text strings of similar content consistent with *P*, when directly asked about *P*;
- it would frequently output *P* or text strings of similar content consistent with *P*, when *P* is relevant to other textual outputs it is producing (e.g., when *P* would support an inference to *Q* and it has been asked about *Q*);
- it would rarely output negations of *P*, or text that (if a human were to write it) would be naturally interpreted as claims of ignorance about *P*, and similarly for negations of propositions that straightforwardly follow from *P* given its other beliefs\*;
- when *P*, in combination with other propositions that the language model believes\*, would straightforwardly imply *Q*, and the truth of *Q* is important to the truth or falsity of forthcoming text outputs, it will commonly output *Q* or enter a state that facilitates the future output of *Q*;
- the above dispositions are relatively stable, unless new evidence against the truth of *P* is presented.

These conditions could be refined and further conditions could be added, but this conveys the idea. These conditions are imprecise, but that is a feature, not a bug. The same is true for dispositional characterizations of personality traits and human beliefs. Such concepts are fuzzy-bordered and require expertise to apply.

Current language models do not meet these conditions as a general matter. They “hallucinate” too frequently, change their answers, do not consistently enough “remember” what they earlier committed to, can easily be coaxed into making a statement and then later retracting it, make tangled and inconsistent inferences. If I ask ChatGPT 4.5 whether waffles are better than eggs or vice versa, it will sometimes say (or say\*) one thing, sometimes the opposite, sometimes refuse to express a preference unless pressed, and reverse itself. If I ask Claude 3.7 Sonnet which element is two to the right of manganese on the periodic table, it sometimes gets it right, sometimes gets it wrong, and sometimes reasons in baldly inconsistent ways.<sup>2</sup> On matters such as these, current Large Language Models emit text *suggestive* of belief\*, but it is probably not right to say that they have beliefs\* about the superiority of eggs to waffles or the relative positions of elements in the periodic table.

Still, the situation is changing fast, and for widely shared pieces of common knowledge, such as that Paris is the capital of France, Large Language Models show excellent stability and use of inferential connections (though see Ganapini, Sokol, Chowla, 2026). Also, of course, real human believers are far from consistent and inferentially perfect, so we should not insist on too high a bar. It is a reasonable extension of common usage to say that ChatGPT “says” and “knows” that Paris is the capital of France and thus, perhaps, similarly for “believes”. Even if we want to say that such extensions are not literally correct usages, this is probably enough for belief\*. The practical problem is in determining where these models’ beliefs\* end and their wobbly responding begins. No rigid or bright-line criterion is likely to work. Instead, I recommend a pragmatic approach to belief\*—the ascription need only be good enough, in the relevant respects, for the purposes at hand.

For embodied robots with sensorimotor activity in our shared spaces, belief-talk will likely prove practically unavoidable. The nurse’s assistant robot rolls into a client’s room. Does it “know” where the pills are stored? Maybe yes, may-

---

<sup>2</sup> For an example of an incoherent answer, as of this writing on March 11, 2025, I asked Claude 3.7 Sonnet:

What element is two elements to the right of manganese on the periodic table?

It replied:

The element that is two elements to the right of manganese in the periodic table is iron (Fe).

Manganese has atomic number 25, so two elements to the right would be atomic number 27, which is iron.

The sequence in that part of the periodic table is:

- \* Manganese (Mn); atomic number 25
- \* Iron (Fe): atomic number 26
- \* Cobalt (Co): atomic number 27

So cobalt is the element that is two positions to the right of manganese.

be no. It will be useful to ask this question, giving a different answer for the robot that will straightaway roll to the relevant drawer than for the robot that will need to launch a physical search. We can even try the classic “false belief” task from developmental psychology. Take a robot that we would normally say “knows” where the pills are, then surreptitiously move the pills while the robot is away. It will be natural to say that it “thinks” or “believes” that the pills are in the drawer where it left them yesterday, and we can thereby explain why it moves toward that drawer at pill-taking time. If this robot is sophisticated enough in other ways, for example linguistically, it would meet reasonable behavioral criteria for belief\*. Not only will it search there, it will tell (or tell\*) others that the pills are there; it will say (or say\*) “The pills are supposed to be here; did somebody move them?”; it will initiate a search, or call for help, or at least refrain from the pointless action of offering an empty hand to the client at pill time. Similarly we might say that delivery bots do, or do not, know\* or believe\* that some particular sidewalk is closed for construction, and that companion bots, do or do not, remember\* or believe\* that March 17 is their friend’s birthday, and so on, predicting and explaining their behavior accordingly.

Of course, it is hard to pronounce that asterisk, even if one is scientifically or philosophically convinced that robots are nonconscious and have the wrong underlying cognitive structures and thus cannot literally be said to believe. It is easier to be loose and sloppy, just adapting our comfortable old terms for this new use, until eventually it seems no longer like analogy or extension, but just the literal meaning. You and I might still wonder whether robots of this sort “really believe”, but if technology proceeds in the direction I have described, that will probably come to seem like a philosopher’s quibble.

That is how we will decide that robots “believe”.

Have we fallen back into behaviorist instrumentalism? No, not about belief (hence the scare quotes around “believe” in the title and previous paragraph), only about belief\*. In *our current* language—or, at least, in the language that I hope you and I can share—such systems would be at best in-between cases of believers, despite being paradigm believers\*. What I am imagining is a future evolution of language in which “belief” shifts its meaning. Presumably, people will continue to care about the presence or absence of consciousness and will debate its existence or nonexistence in robots, but we will need words other than “belief” to track such distinctions.

## 5. Realism or Anti-Realism about Belief?

Krzysztof Posłajko (2024) and other “anti-realists” about belief should welcome all this. On anti-realist views, there is no underlying metaphysical fact about what belief “really is” independent of our usage. No metaphysical anchor prevents us from changing our usage in whatever way is convenient. The concept is “plastic” and “apt for serious semantic changes” (p. 176). Neither do our belief-attribution practices need to track the results of cognitive science, from

which they are “autonomous” (p. 182). As long as belief-ascription practice is useful—as Pośłajko thinks it is—and as long as we embrace a pragmatic metaphilosophy—as Pośłajko does—we should welcome its effective adoption to new uses. Fictionalists (e.g., Toon, 2023) and interpretativists (e.g., Curry, 2020), should adopt a similar attitude. However, Pośłajko mentions “artificial systems” only briefly in passing (p. 174), so it is hard to know whether he would indeed reach similar conclusions about belief or belief\*. I invite him to expand upon this issue.

Although Pośłajko cites my phenomenal dispositionalist view as an inspiration for his view (2024, p. 73), he portrays himself as disagreeing with me on a crucial point. I hold, he says, that the ordinary folk psychological concept of belief is superficial, and thus beliefs exist in the ordinary sense as long as superficial patterns of the right sort exist (p. 116). My view then counts as realist, though with a light touch. On his view, the ordinary folk psychological concept of belief is deep and requires the existence of “underlying mental causes” (p. 121) and “stored internal states” (p. 123). If there are no such causes and states, or if no such states exist as robustly real “natural kinds”, then our current concept of belief “is misleading” and “requires conceptual revision” (p. 128). Still, Pośłajko argues, beliefs exist and belief ascriptions can be true and useful. Ordinary people are mistaken only about the metaphysical kind to which beliefs belong and how robustly real they are. Pośłajko’s anti-realism is thus also light-touch.

I suspect the truth about the folk psychology of belief lies somewhere in the middle—containing conflicting strands and metaphysically open-textured. Pośłajko notices a real tendency in the folk to treat beliefs as discretely stored, deep underlying causes of our dispositions. Yet also, as Curry (2019) and Gauker (2021) note, folk psychological thinking has superficialist elements too, especially if we think about how people would apply the concept of belief to hypothetical cases where the deep structure is postulated to be different or absent. By emphasizing different aspects of our informal understanding of belief, we can rigorize the concept in different ways—and we can then appeal to pragmatic grounds for rigorizing in one way or another.

The oft-cited fact that people are widely held to act as they do “because” they believe as they do (Egan, 1995, as cited in Pośłajko, 2024, pp. 118–119) is no evidence at all in favor of the view that ordinary people hold a deep, causal notion of belief. “Because” explanations can be dispositional (the glass broke because it was fragile) and even entirely non-causal (the slope of the curve increases because the equation is exponential).

If Pośłajko is willing to accept that the ordinary conception of belief is multifaceted in a way that makes dispositionalism one acceptable rigorization of it, the gap between Pośłajko’s view and my own narrows. Any remaining gap lies, I suspect, mainly in the metaphysical respect I have for “unnatural” folk kinds (cf. Dupré, 1995 on fish), including belief as real enough, and in my inclination to accept the importance of patterns of phenomenal consciousness to belief.

Still, insisting on the centrality of consciousness to belief might soon begin to look old-fashioned and, perhaps ironically, not superficial enough.

## REFERENCES

- Bantegnie, B., Schwitzgebel, E. (2026). The Metaphysics of Beliefs: Dispositional Theories. In T. Lombrozo, N. Van Leeuwen (Eds.), *The Oxford Handbook of the Cognitive Science of Belief*. Oxford University Press.
- Butlin, P., et al. (2023). Consciousness in Artificial Intelligence: Insights From the Science of Consciousness [ArXiv:2308.08708]. *ArXiv*.  
<https://arxiv.org/abs/2308.08708>
- Chalmers, D. J. (1996). *The Conscious Mind*. Oxford University Press.
- Chalmers, D. J. (2023, August 9). Could a Large Language Model be Conscious? *Boston Review*.  
<https://www.bostonreview.net/articles/could-a-large-language-model-be-conscious>
- Colombatto, C., Fleming, S. M. (2024). Folk Psychological Attributions of Consciousness to Large Language Models. *Neuroscience of Consciousness*, 2024(1).
- Curry, D. S. (2018). Beliefs as Inner Causes: The (Lack Of) Evidence. *Philosophical Psychology*, 31, 850–877.
- Curry, D. S. (2020). Interpretativism and Norms. *Philosophical Studies*, 177, 905–930.
- Davidson, D. (1984). *Inquiries Into Truth and Interpretation*. OUP.
- Dehaene, S., Lau, H., Kouider, S. (2017). What Is Consciousness, and Could Machines Have It? *Science*, 358(6362), 486–492.
- Dennett, D. C. (1987). *The Intentional Stance*. MIT Press.
- Dretske, F. (1995). *Naturalizing the Mind*. MIT Press.
- Dupré, J. (1995). *The Disorder of Things*. Harvard University Press.
- Egan, F. (1995). Folk Psychology and Cognitive Architecture. *Philosophy of Science*, 62, 179–196.
- Fischer, E., Sytsma, J. (2021). Zombie Intuitions. *Cognition*, 215.
- Ganapini, M. B., Sokol, A., Chawla, N. (2026). Do LLMs Have Beliefs? <https://philarchive.org/rec/BERDLH-3>
- Gauker, Ch. (2021). Belief Attribution as Indirect Communication. In L. Koreň, H. B. Schmid, P. Stovall, L. Townsend (Eds.), *Groups, Norms and Practices* (pp. 173–187). Springer.
- Godfrey-Smith, P. (2016). Mind, Matter, and Metabolism. *Journal of Philosophy*, 113, 481–506.
- Graham, G. (2000/2023). Behaviorism. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/spr2023/entries/behaviorism/>
- Kirk, R. (1974). Zombies vs Materialists. *Proceedings of the Aristotelian Society*, 48 (supplementary), 135–152.
- Levin, J. (2004/2023). Functionalism. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/functionalism/>

- Lewis, D. K. (1972). Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy*, 50, 249–258.
- Mölder, B. (2010). *Mind Ascribed*. Johns Benjamins.
- Pautz, A. (2021a). Consciousness Meets Lewisian Interpretation Theory: A Multi-stage Account of Intentionality. In U. Kriegel (Ed.), *Oxford Studies in Philosophy of Mind* [Vol. 1, pp. 263–314]. Oxford University Press.
- Pautz, A. (2021b). Varieties of Interpretationism About Belief and Desire. *Analysis*, 81, 512–524.
- Pober, J. (2024). *Hybrid Functionalism* [Unpublished manuscript]. Centre of Philosophy, University of Lisbon.
- Posłajko, K. (2024). *Unreal Beliefs: An Anti-Realist Approach in the Metaphysics of Mind*. Bloomsbury.
- Putnam, H. (1975). The Meaning of “Meaning”. *Minnesota Studies in the Philosophy of Science*, 7, 131–193.
- Quilty-Dunn, J., Mandelbaum, E. (2018). Against Dispositionalism: Belief in Cognitive Science. *Philosophical Studies*, 175, 2353–2372.
- Robinson, W. S. (2019). *The Epiphenomenal Mind*. Routledge.
- Schwitzgebel, E. (2002). A Phenomenal, Dispositional Account of Belief. *Noûs*, 36, 249–275.
- Schwitzgebel, E. (2013). A Dispositional Approach to Attitudes: Thinking Outside of the Belief Box. In N. Nottelmann (Ed.), *New Essays on Belief* (pp. 75–99). Palgrave.
- Schwitzgebel, E. (2021). The Pragmatic Metaphysics of Belief. In C. Borgoni, D. Kindermann, A. Onofri (Eds.), *The Fragmented Mind* (pp. 350–375). Oxford University Press.
- Schwitzgebel, E. (2024). *The Weirdness of the World*. Princeton University Press.
- Schwitzgebel, E. (2026). Dispositionalism, Yay! Representationalism, Boo! In J. Jong, E. Schwitzgebel (Eds.), *The Nature of Belief* (pp. 286–308). Oxford University Press.
- Seth, A. (2024). Conscious Artificial Intelligence and Biological Naturalism. [PsyArXiv:tz6an]. *PsyArXiv*. <https://doi.org/10.31234/osf.io/tz6an>
- Strawson, G. (1994). *Mental Reality*. MIT Press.
- Toon, A. (2023). *Mind as Metaphor*. Oxford University Press.
- Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, 49, 433–460.



KRZYSZTOF POSŁAJKO \*

## DEFENDING MINIMAL NON-REALISM

**SUMMARY:** Reply to commentators in *Semiotic Studies* special issue on *Unreal Beliefs*. I defend the position developed in this book, i.e., minimal non-realism, against various charges. First, I defend the way I reconstructed folk commitments to realism about beliefs. Second, I deal with the arbitrariness charge against my definition of belief realism. Third, I critically evaluate a counterproposal to my view, namely the idea that beliefs are socially constructed natural kinds. Then I deal with the challenges levelled against my arguments against realism about belief. Finally, I reconsider some of the consequences of minimal non-realism and consider potential alternatives to it.

**KEY WORDS:** beliefs, folk psychology, anti-realism, dispositionalism, natural kinds.

### 1. Introduction

I am extremely grateful to all the commentators at the special issue on *Unreal Beliefs* for their insightful reviews, as well as to the editors of the volume, namely Agata Machcewicz-Grad and Maciej Tarnowski. Most importantly, the presented comments show that the issue of the reality of beliefs (and, more generally, of the reality of folk-psychological states) is still of significant interest to philosophers of mind. Showing that this question is vital and there are ways to make progress regarding it was the primary aim of my project.

The reviews have forced me to rethink some of the commitments I made in the book. I also had to reconsider the strength of my position, especially relative

---

\* Jagiellonian University in Kraków, Department of Philosophy. E-mail: krzysztof.poslajko@uj.edu.pl. ORCID: 0000-0002-7883-1651.

to its competitors. Perhaps more importantly, the reviews pressured me to reflect more on the motivations which underlie the positions I developed in *Unreal Beliefs* (Posłajko, 2024). All in all, I still think that minimal non-realism is a viable position regarding the metaphysical status of beliefs. However, what I have come to realize is that the attractiveness of the view I presented strongly hinges on several assumptions that might not be universally shared. Thus, its attraction might be limited to philosophers who share a certain, perhaps slightly idiosyncratic, vantage point. For this reason, the defense of the view that I offer in this paper should mostly be read as spelling out the basis on which the project is built.

Many worries regarding my project were raised by more than one commentator. For this reason, I decided to organize the response thematically. I will start with the problem that has been raised most frequently, namely the question of whether my rendering of folk commitments was adequate. Special attention will be given to the empirical versions of that worry provided by Kalis (2026) and Komorowska-Mach (2026). Second, I will deal with a worry that many commentators have raised, albeit in different forms, namely that my definition of realism was arbitrary. Then I will tackle the suggestion made primarily by Curry (2026) and Zawidzki (2026) that an attractive alternative to my position is the view that takes beliefs to be socially constructed natural kinds, and that my book does not contain a satisfactory argument against such a proposal. Then, I will deal with the charges that my arguments against realism are faulty, focusing on those provided by Kozak (2026) and Horowitz (2026). Next, I will look at the consequences of my position, and finally at some of the possible alternatives to it, especially those presented by Bantegnie (2026), Horowitz (2026), and Mölder (2026).

## 2. Folk Commitments

Several commentators (including Bantegnie, Zawidzki, Mölder, Horowitz, Kalis, and Komorowska-Mach) have complained that my rendering of folk commitments regarding the reality of beliefs is mistaken. In short, I claimed (in Chapter 7 of *Unreal Beliefs*) that the folk consider beliefs to be strongly real in the sense that folk psychology carries the commitment to treating beliefs as internal causes of behavior. This, in turn, is why I postulated a mismatch between folk-psychological commitments and the fact that beliefs are non-real.

The reviewers have nearly unanimously claimed that I overstated the strength of folk commitments, which should be seen as less realist. This sort of complaint is hardly surprising, as many of the commentators adhere to the broadly understood Rylean tradition,<sup>1</sup> according to which the actual commitments of the folk

---

<sup>1</sup> As Bantegnie (2026) notes, my reading of Ryle (2009) is relatively superficial and ignores more detailed interpretations of the actual works of Ryle. I must admit that I relied on the popular picture of Ryle: in my defence, I might only say that it is this popular reading of Ryle that influenced later discussions on dispositionalism.

are far less substantial than philosophers tend to interpret them as being. As I admitted in the book, the part about folk commitments was perhaps the most risky part of the project. The idea that the folk are committed to there being internal causally active states called beliefs seems open to empirical falsification. Still, I will try to defend this idea.

## 2.1. Folk Commitments and Anti-Realism

A common complaint (levelled by, among others, Mölder, Bantegnie, and Horowitz) was that my insistence on the substantial character of folk commitments is an unnecessary addition to my general project of minimal non-realism. The gist of the complaint is that my main thesis was metaphysical: that beliefs do not constitute natural properties in Lewis's (1983) sense. Given this conclusion, considerations about what people think about their reality should not bear too much relevance to the issue of belief-realism.

I agree that one could, in principle, make one claim without the other: Mölder (2026), following Demeter (2009), distinguishes between the issues of entity realism and semantic realism. The claim of entity realism concerns the metaphysical status of beliefs, whereas semantic realism concerns the status of folk-psychological discourse. Being an entity realist (or anti-realist) concerning beliefs does not say anything about the status of folk psychology as a discourse. There is also no need to postulate any mismatch between folk psychology and reality in order to embrace entity anti-realism. Paradoxically, as Mölder points out, an entity realist might also postulate a sort of mismatch between folk psychology and reality: if folk psychology turned out to be insubstantial in its commitments, and beliefs turned out to be metaphysically real, then we would also have a mismatch between folk psychology and reality. This kind of mismatch, however, would not warrant anti-realist conclusions of any sort.

In response to this worry, I must say that my insistence on the idea of a mismatch is rooted in my motivation to preserve Churchland's central intuition of classical eliminative materialism (1991), namely that there is some important error in the way we normally conceptualize our mental lives. This motivation, as it turns out, is somehow idiosyncratic: it is rejected not only by intentional realists, but also by most Ryleans, who consider the folk to be free of error in their conceptualization of the intentional sphere. For a typical Rylean, the errors of folk psychology lie solely in the domain of realist philosophical theories.

Horowitz (2026) provides a useful terminological proposal to make matters clearer here: he proposes labelling the claim that beliefs are not metaphysically real "belief non-realism"; he reserves the name "belief anti-realism" for the further claim of a mismatch between substantial folk psychology and the non-realist metaphysics of beliefs. Framed in these terms, it would seem that most contemporary Ryleans are staunch belief non-realists, while belief anti-realism remains as unpopular as it ever was.

In my view, even showing that belief anti-realism can be coherently phrased is a theoretical gain. If we agree that eliminative materialism is a non-starter, then there is a philosophical gain in spelling out what it would mean to discover that the folk are in error about folk psychology. Even if the claim regarding the existence of this error remains tentative, the very articulation of how to make sense of the possibility of such an error widens the spectrum of possible positions.

Bantegnie (2026) points out another problem with my mismatch thesis: my reliance on the Thomassonian framework of easy ontology (Thomasson, 2015) strongly suggests that, on my view, the application criteria of the term “belief” are superficial. If this is so, then it is not clear how my view differs from dispositionalism. In response to this, I wish to point out that on my view the concept of “belief” (along with, perhaps, other folk-psychological concepts) is a complex beast as it contains both “application criteria” and “metaphysical commitments” (Section 7.3 of *Unreal Beliefs*). The application criteria of “belief” are indeed superficial; if we focused only on them, then the dispositional analysis would be mostly on track. However, if I am right, the folk have a tendency to make certain inferences about the metaphysical underpinnings of what is attributed on a superficial basis. This means that once we decide, on the basis of superficial dispositional evidence, that *S* believes that *p*, we have an inferential disposition to attribute deep, internal, causally active state of believing-that-*p* to *S*. This mechanism is not unique to folk psychology: in general, any essentializing process in human thought has this feature (e.g., Neufeld, 2022).<sup>2</sup>

This distinction between application criteria and metaphysical commitments is also important for my project as it supports the idea that the error of folk psychology is partial. It is not that we are entirely in the wrong when we attribute beliefs: there is only an aspect of our thinking that turns out to be mistaken. This diagnosis of only a partial mistake, in turn, motivates a kind of revisionary-preservationist approach, rather than wholesale elimination.

## 2.2. Functions of “Because”

Some commentators (e.g., Bantegnie, 2026; Zawidzki, 2026) have also taken truck with my principal argument for the realist reading of folk commitments, namely my thesis that the fact that the folk use “because” phrases in belief explanations is evidence for the claim that the folk take beliefs to be causes. Both Bantegnie and Zawidzki note that “because” does not necessarily mean that a causal relationship is being postulated. For Bantegnie, “because” phrases can be given a dispositionalist reading, whereas Zawidzki notes that “because” phrases can be also used in post hoc rationalizations. This is a fine point, and I agree that it is possible to propose an alternative reading of “because” in this context.

---

<sup>2</sup> Obviously, for Neufeld and other writers, essentializing is a process that regards concepts seen as internal mental states. In my approach, this essentializing is rather a feature of discourse than thought.

Still, my defeasible hypothesis is that—at least on the level of what Dennett (1989) called folk-psychological ideology—beliefs are taken to be causes; folk-psychological explanations are taken to be causal explanations; and “because” signals the commitment to such causal explanations. Admittedly, there is neither a quick and easy argument for my claim that “because” in the context of belief-explanations functions to signal causal relationships, nor for the opposite claim of Bantegnie and Zawidzki.

The only thing that can be said is that the causal interpretation seems to be the simpler one as it does not postulate a radical discontinuity between uses of “because” in psychological and non-psychological explanations. The postulate of this discontinuity, in my view, stems from the wish to fit the practice of folk psychology to the assumed theoretical framework, rather than from the observed actual difference between folk-psychological practices and other explanatory activities.

Bantegnie argues that “because” is sometimes used to provide non-causal explanations in other contexts as well. This is certainly true, but the question remains whether such uses can serve as an illuminating model for understanding uses of “because” in folk-psychological contexts. He gives two examples of such non-causal uses of “because” outside of psychology: “The temperature in the room has dropped because the kinetic energy of the air molecules has lowered” and “7 is a prime number because it is divisible only by 1 and itself” (Bantegnie, 2026, p. 155).

In my view, these examples fail to provide good analogies to the folk-psychological case as they are examples of broadly understood “a priori” explanations. Folk-psychological explanations do not strike me as being similar to a priori ones as there is a fair share of contingency involved in the relationship between actions and the reasons we postulate to explain them. All in all, I think that the onus here is on the non-causalists to provide a good model for non-causal explanations, and to show that this model is applicable to the folk-psychological case. As long as we do not have such a model, a naïve causalist interpretation seems to be the reasonable one.

### 2.3. Empirical Worries

The most substantial criticisms of my idea that the folk are committed to belief realism came from Kalis and Komorowska-Mach, who, importantly, base their criticism of my view on the empirical data. Kalis (2026, p. 89) bases her criticism of my view on folk commitments on “anthropological work in various cultural and linguistic communities”. Her main conclusion from these anthropological works is that folk psychology is not a uniform practice across cultures. More importantly, many of these practices are not “causalist” in the sense that they do not seem to rely on attributions of internal causal states. Still, the folk are doing fine in such cultures: they are able to successfully navigate interpersonal interactions and predict each other’s actions.

In a sense, the idea that causalist folk psychology is not a universal phenomenon is something which speaks in favor of the minimal non-realist view. If it is true that the folk in different cultures are able to do perfectly fine without postulating causally active beliefs, then the attraction of the traditional success argument (e.g., Fodor, 1987) for belief-realism wanes as we must seek the sources of success in interpersonal interactions in something different than the postulation of internal causal beliefs.

Still, the important worry that Kalis raises for my view is the issue of whether our contemporary Western culture really is committed to causalism. I agree that it might be the case that we engage in practices which are also not based on causalist assumptions: after all, if there are no beliefs understood as internal causal states, then there has to be some route by which we come to adequately predict each other, and this route must work without postulating beliefs. But the worry of Kalis is whether we “Westerners” really do postulate beliefs as causes at all. Kalis suggests that even if we sometimes do, this has more to do with the influence of philosophy and psychology on Western folk. Such a hypothesis is obviously tempting since it amounts to saying that reification of folk-psychological states is not something that comes to us naturally, but rather is an effect of theoretical impact on the folk’s linguistic practice.

But this is not the only possible way of seeing things. Speculating a bit, I would venture an opposite hypothesis: that the meteoric rise of philosophy of mind and theoretical psychology in the “West” was a result, rather than the cause, of our Western commitment to the idea of the “internal reality” of hidden causal mental states. Realist philosophy of mind (and the later theoretical psychology) might be seen as an attempt to systematize and vindicate a very parochial cultural construct of Western culture, namely the idea that there is a realm of internal, hidden intentional states. This is obviously purely speculative, and it would not be easy to verify such a hypothesis. Still, I think that the problem of the sources of our apparent commitment to causalism, especially given the assumption that beliefs are not real, is an important one.

Komorowska-Mach (2026) offers a subtle, empirically informed (by psychological literature on mindreading) proposal regarding three levels of engagement in folk-psychological practice. She distinguishes (1) mindreading as a craft, (2) folk-psychological talk, and (3) philosophical reconstruction of folk psychology. In her view, an adequate account of mindreading as a craft does not require postulating causally active internal states (with which I concur). It is also obvious that the philosophical reconstruction often takes the folk to be committed to internal states.

What is ingenious in Komorowska-Mach’s reconstruction is the fact that she postulates another layer between mindreading as a craft and philosophical reconstruction: the level of folk-psychological talk. Komorowska-Mach characterizes the roots of this level as follows: “We package a richer, holistic understanding of somebody’s perspective, their ‘informational states’, into the noun-and-clause structure our language provides” (Komorowska-Mach, 2026, p. 105).

I think that this postulate of distinguishing these three levels of engagement is on the right track, and how Komorowska-Mach depicts what happens on these levels is a much better description of the reality of mindreading than what I offered in the book. In general, most of her proposal about the way folk psychology operates can be seen as consistent with the metaphysical picture I proposed (which is what Komorowska-Mach seems ready to admit).

The point of contention is fairly subtle: Komorowska-Mach has serious reservations about the claim that the second linguistic level actually contains substantial commitments. As she writes:

At the second level, it may seem that the folk are realists about belief, but this is merely an artefact of the constraints of natural language, which is geared toward talking about macroscopic, discrete objects [...]. The burden of proof lies with anyone who claims that these linguistic chunks express genuine ontological commitments. Absent further evidence, I take it to be entirely possible that the folk are neither realists nor anti-realists about beliefs, at least until philosophers prompt them to take a stance. (Komorowska-Mach, 2026, p. 106)

In response to this, I claim that even though the folk language does not commit us to beliefs as *entities*, its superficial reading does lead to commitment to internal *states*. The idea of beliefs as entities might strike us as an unreasonable interpretation of the folk language as it suggests an unattractive picture of beliefs as sorts of solid objects. Still, we do seem to be committed to beliefs as standing states, i.e., we do seem to think that our beliefs are something that are somehow stored in our minds and wait to be activated in the right circumstances.

The sort of reading I am offering is based on the idea that the ontological commitment of folk psychology can, to an extent, be read off the superficial linguistic form of folk-psychological talk. This superficial linguistic form might be a result of the imperfect projection of our actual practice onto the linguistic structure. Still, I concur that this form serves as the basis for reasonings the folk are disposed to make: from the results of complex mindreading processes to the postulation of internal states.

To sum up, I think that the empirical arguments raised by Kalis and Komorowska-Mach force us to admit that the reality of folk psychology is far more subtle than implied by my rough-and-ready characterization of it. Still, I think that nothing they wrote disproves my tentative claim that there is some level of engagement in folk psychology that licenses a causalist interpretation of it. This sort of engagement might be parochial, and it might be only one, not very important, aspect of the richness of folk-psychological practice. Still, it seems to be there.

### 3. Arbitrary Definition of Realism

Another complaint that many of the commentators share is that the project of minimal non-realism is based on arbitrary definitions of what it means to be

a realist (this was pointed out by Bantegnie, Mölder, Zawidzki, Horowitz, Kozak, and Hensel).

In the book I was ready to admit that my definition of realism was to a great extent projective. The proposed definition was intended to do two things: first, to capture what it is that philosophers care about when they debate about the reality of beliefs; second, to be a useful guide for conducting the debate further. Although I admit that this is not the only possible way to define realism, I still think my definition can be seen as useful for the debate.

### 3.1. Bantegnie on Dispositionalism

Bantegnie's (2026) complaint is that my definition of realism makes dispositionalism an anti-realist position by definition. He thinks this is a misunderstanding, as for him there are two versions of dispositionalism that deserve to be called "realism". The first one postulates that beliefs are multi-track dispositions that cluster together, and the clustering of these sub-dispositions is guaranteed by law-like generalizations. The second version is stronger and asserts the same claims as the first one, but adds a claim that there is a categorical basis to these multi-track dispositions. Bantegnie's criticism is that I unfairly excluded these positions from being classified as realist as I equated realism about belief with representationalism, and with the postulate of internal cogs in the cognitive machine.

I admit that I made representationalism a paradigm case of realism about beliefs, but this does not mean that I intended to a priori exclude other ways of being a realist. Most importantly, I hope I have made clear that I take the issue of realism to be a matter of there being a spectrum of positions: ranging from extremely realist to extremely anti-realist, with many possible options in between. My omission of the possible realist versions of dispositionalism, although lamentable, can thus be easily fixed. A stronger version of realist dispositionalism as described by Bantegnie should be classified as an unequivocally realist position as it meets the desiderata of "objective similarities" (provided that such a version of dispositionalism could ground the idea of beliefs having objective contents) and "entering causal explanations", even if those similarities and explanations are somehow "less exact" than on the representationalist model. One might wonder, however, whether a version of dispositionalism that postulates categorical bases for belief-related dispositions still counts as dispositionalism *sensu stricto*, but this, again, would be a purely definitional debate.

A more curious case is the case of dispositionalism that does not postulate categorical bases for belief-related dispositions, but claims that a cluster of such dispositions is connected by law-like generalizations. In this case we have a theory that does not carry a commitment to anything "internal" in the explanation of beliefs, but there seems to be something realistic about it. Such a version of realist dispositionalism might then be a prime example of a position which fits perfectly in the middle of the realism vs anti-realism spectrum. Such a view would keep some of the realist commitments (beliefs are somehow explanatory if we

agree that dispositions can serve as explanations), and there is a level of objective similarity between so understood beliefs.

In this way, my definition of realism about beliefs does not a priori exclude the kind of view Bantegnie proposes, but rather shows how they are located in relation to other more and less realist proposals. In general, I think that Bantegnie's insistence that I conflated the reality of beliefs with their internal character is a bit of a mischaracterization of my view. A representationalist view that postulates internal beliefs is just a paradigm of the realist position because it clearly meets the criteria of objective similarity, causation, and relation to fundamental properties. But this does not mean that positions other than the standard representationalist one cannot meet these criteria. Thus, I feel guilty of a failure of imagination as I have not predicted the possibility of the realist version of dispositionalism. However, I still feel that my definitions stand. I will revisit the idea of realist dispositionalism in Section 7.1.

### 3.2. Zawidzki on Imprecision

Zawidzki (2026) complains that I remain elusive about the exact location of beliefs, as I see them, on the spectrum of realness. He contrasts beliefs with arbitrary modal collections on the one hand, and with social kinds such as money on the other; he himself likens beliefs to "natural social kinds", such as money. I will discuss Zawidzki's positive proposal in the next section, but here I want to focus on the question of where I postulate locating beliefs. On my take, beliefs are less real than money (assuming that there indeed are non-trivial generalizations about money), but they are not fully arbitrary collections of modal objects.

If I were to postulate an analogy from social ontology, then I would suggest that beliefs as a category are somewhat similar to the category of "heads of states".<sup>3</sup> Heads of state clearly exist: there are numerous people, past, present, future, actual, and non-actual, to whom we might predicate the property of being a head of state. There are some features that all heads of state seem to share, including certain social statuses and expectations. There are good pragmatic reasons for singling out this category, e.g., reasons concerning how heads of state should be treated according to diplomatic protocol. Yet, heads of state do not constitute any "natural kind", even if we allow social kinds to be natural. Heads of states vary immensely, both in their internal properties and in the social statuses that they are given (presidents can be either mostly ceremonial positions or hold near-dictatorial powers, etc.). Most importantly, there seem to be no important, non-trivial generalizations about heads of state, and the prospects for the discovery of such generalizations remain dim.

---

<sup>3</sup> This analogy is inspired by Schwitzgebel's analogy between consciousness and "kings" (Schwitzgebel, 2025). However, I do not feel committed to the whole package of Schwitzgebel's views about consciousness.

Beliefs, on my view, are similar to such “thin” social categories. There are good reasons to single out beliefs as a category, and there is certainly a certain status that can be attached to being a believer that  $p$ ; however, if I am right, there are no interesting generalizations about this kind. More on the distinction between beliefs and natural social kinds will be said in Section 4.

### 3.3 Horowitz on Functional Properties

Horowitz (2026) claims that my definition of realism excludes functional properties from being natural. According to him, my definition of realism would require beliefs to have internal essences to be considered natural kinds. Thus, as beliefs are almost universally accepted to be functional kinds, then my conclusion that they are not natural kinds follows trivially. I think this is a misunderstanding. My definition of realness/naturalness does not exclude functional properties in general. Still, it puts some demands on those functional properties which we would like to classify as natural properties. Such properties would have to be projectible: there have to be some non-trivial generalizations regarding such properties. This is because we can claim that we have genuine explanations only if there are such generalizations.

It would be spurious to claim that all functional properties are natural: there are some functional properties which constitute classes whose members do not share anything more than the functional property used to define the kind in question. Take the example of “valve”: being a valve is a functional property, but there are lots of types of valves, and nothing of substance can be said about this property other than what is packed in its definition. On the other hand, there are more interesting functional properties. Being a predator is also an admittedly functional property. However, objects belonging to the class of predators apparently share lots of different properties, not only that of being an animal that hunts other animals. If indeed predators constitute such a class that shares numerous surprising properties, then “predator” counts as a natural, albeit functional, kind.

The question is then whether the class of “believers that  $p$ ” is more akin to the class of valves or to the class of predators. This translates to the question of whether this kind is such that its members share a number of different surprising properties. My answer to this question is negative, but it is not a matter of pure stipulation: it could turn out to be the other way round, and the claim of there being no natural category of believers is based on the available evidence, not on definitional fiat.

### 3.4. Hensel on Realism and Natural Kinds

Hensel (2026) is skeptical about my proposal to define reality in terms of natural kinds. For him, the idea that we can achieve progress in adjudicating the question of realism about beliefs in this way is mistaken, and we should revert to the original formulation of Churchland (1981), for whom the question of

realism about beliefs was to be seen as the question of the truth of folk psychology as a theory.

Hensel identifies two main problems with the natural kinds approach. The first is what he calls “the problem of disappearing substance”; the second is “the problem of disappearing reality”. The first problem goes as follows: natural kind theories together with natural kind-term semantics allow us to drop any commitment that we seem to associate with a given kind *K*. So, if we discover that there are no things that have feature *F* and we previously thought that *F* is essential to *K*, we are in no position to declare that this kind does not exist. A natural kind-style realist about *K* might postulate that we keep the commitment to *K* but drop the commitment to the idea that *K* must be characterized by *F*. In this way, a realist about *K* can survive (almost) any refutations of there being *K*.

In the case of beliefs, we have a clear example of this strategy. Once upon a time a realist about beliefs used to be committed to the idea of beliefs being “sentences in mentalese”. But what happens once the idea of mentalese is rejected? Realists about beliefs (as a natural kind) simply drop their commitment to mentalese and say that natural kind “beliefs” exist. We were just wrong about beliefs’ essential properties: we mistakenly assumed that they must be sentences in mentalese. In such a way, it might turn out that any negative result regarding beliefs might be brushed off by belief-realists.

The second problem identified by Hensel, “the problem of disappearing reality”, is that once we drop the classic micro-essentialist view of natural kinds, then kinds—understood in a less robust manner—become conventional. Moreover, it is very hard to ascertain which kinds are projectible as we still do not know which generalizations are true: it always might turn out that a kind we deem not to be projectible will become the basis of some future inductive generalizations.

I acknowledge that the problems identified by Hensel are very much real, but I do not find them unsurmountable. First, to avoid the problem of realists dodging their commitments, we should focus on what we care about when we debate the reality of beliefs. As Hensel notes, I claim that there are two essential commitments of belief realism that are worthy of its name: that beliefs have propositional semantic properties, and that they are causal explanations of actions. Of course, a realist could come and say that they would drop this commitment and still defend the reality of what they would call “beliefs”. But the reality of such understood “beliefs” would be of nearly no interest to philosophers of mind. We care about the reality of beliefs because we care about there being states that can combine being vehicles of propositional content with being causes of behavior. If there are some other states that are somehow related to beliefs so understood, then fine—there are no semantic police to stop anyone from calling them “beliefs”. But if there are no states that are at the same time propositional and causal, then we can claim victory for belief anti-realism because there is nothing that has the properties we care about.

This also resolves the worry about conventional kinds. As I argued in my earlier work (Poslajko, 2022), there are many possible precisifications of the term

“belief”, but there does not seem to be one that is objectively best. Does this bar us from making progress in the realism debate? I do not think so. If there is evidence that at least one kind that embraces our core postulates is projectible, then realism is “winning”. But as long as there is no acceptable—from the point of view of what we care about—precisification of belief that meets the criterion of projectibility, the anti-realist views retain the upper hand.

This brings us to the last issue, namely that we do not know which kinds will turn out to be projectible. This is uncontroversial, but this does not change the fact that we can make claims given the available evidence. I admitted in the book (e.g., Poślajko, 2024, p. 112) that the issue of the reality of beliefs is an empirically open one. Still, my anti-realist preference is based on what I take to be the current state of knowledge: currently, we do not have positive proof of there being genuinely surprising generalizations about the category of beliefs; therefore, we can, tentatively, accept non-realism.

All in all, there are genuine problems that arise when we tie realism to natural kindness, but these problems are not insurmountable. Moreover, the alternative that Hensel proposes—to go back to folk psychology as a theory—seems even more problematic. Gone are the days when seeing folk-psychological explanations as species of theory-based explanations was universally accepted. Today, those theorists who are still likely to compare folk-psychological explanations to scientific ones are more likely to see folk psychology as a model (e.g., Godfrey-Smith, 2005), but even this kind of assimilation of psychological to scientific vocabulary is controversial.

#### 4. Socially Constructed Natural Kinds

Both Curry (2026) and Zawidzki (2026) locate a lacuna in my arguments. They note that I envisage the possibility of beliefs being seen as socially constructed natural kinds, and I reject this possibility (Poślajko, 2024, Section 8.6), but according to them my arguments against this view are too weak.

On their approach, socially constructed kinds can be natural because members of such kinds might share important properties, and this projectibility can be supported by social mechanisms. For Curry and Zawidzki, beliefs should be seen exactly as this kind of social natural kind. As Curry characterizes the issue: “The question, then, is whether folk-psychological processes do construct beliefs such that those beliefs go on to play a distinctive role in the causal nexus of the world (rendering them projectible)?” (Curry, 2026, p. 119).

Curry starts from the assumption that beliefs are for attributors, and that once an attributor attributes beliefs veridically, then they are able to make many successful predictions about the person to whom they attributed the belief. He contrasts this case with the case of race (an analogy I relied on heavily in my book; Poślajko, 2024, Section 7.4), because for Curry if all there is to race is the racialization process, and there are no true generalizations about racial groups, then race attributions are never veridical.

In my view, Curry's argument relies on the erroneous conflation of veridicality and projectibility. Curry seems to claim that if a given social predicate is not projectible, then it cannot be veridically attributed. In my opinion, this view is patently false. There are many social categories that are non-projectible but veridically applicable. Take the example I used above, namely that of the category "heads of state". There seem to be no interesting, surprising generalizations about heads of state. Still, it seems to me that obviously it is true of many people that they are heads of state.

Still, the question of whether beliefs-as-socially-constructed are projectible is a vexed one. It hinges on the question of whether beliefs-as-constructed-by-attribution really provide us with a basis for predicting behavior. Curry and Zawidzki certainly think that this is the case. As Zawidzki claims: "to believe that  $p$  is to adopt the very fine-grained social role of a believer-that- $p$ , maintained in the same way as more canonical social roles, like motherhood: in virtue of deontic attitudes prevailing in communities" (Zawidzki, 2026, p. 24). The fact that believers occupy such roles, and these roles are stabilized by social mechanisms, is the basis on which we might make substantial predictions about the behaviours of believers.

I agree with Zawidzki that being a believer-that- $p$  is a kind of social role, but I think that this role is relatively thin and not much can be inferred from it. This is because beliefs are hopelessly holistic. Let us take a look at an example of holism from Crispin Wright: "I may manifest my belief that the sea is cold by refusing to bathe (if I do not like cold water bathing) or plunging in joyously (if I do, or want to present an example of British fortitude, etc.)" (Wright, 1993, p. 22).

What such examples show is that we hold some normative expectations for believers-that- $p$ , but I contend that we are surprisingly tolerant about deviations from "normally expected behaviour". We are well aware that people suffer from weakness of will, that they might have eccentric desires, that they might have non-standard background assumptions, etc. Thus, even in the simplest cases of beliefs, such as the belief that the sea is cold, there is a surprisingly rich body of behaviours that count as being coherent with holding this belief. This means that the role of a believer-that-the-sea-is-cold is very thin: it probably involves very little besides saying so or nodding approvingly in the relevant contexts.

Zawidzki seems to be aware of these holistic considerations, and he explicitly contrasts belief-explanations with causal-explanations (Zawidzki, 2026, p. 19). However, it is unclear to me how this non-causal-explanatory role that he envisages for belief-explanations is to be squared with the idea of there being a rich social role of a believer-that- $p$ .

This skepticism about the richness of the social role of a believer-that- $p$  corresponds nicely with the considerations presented in Section 2.3 above. If Komorowska-Mach and Kalis are right that the actual mechanisms of mindreading are independent of the (parochial) linguistic practice of belief-attributions, then it is no wonder that, appearances to the contrary, we gain little predictive power by making explicit belief-attributions.

All in all, Curry and Zawidzki are right that I have not provided conclusive proof that beliefs cannot be a natural social kind. On the other hand, they have not provided a decisive reason to think that beliefs are such a natural social kind. Rather, they have merely expressed their optimism about the to-be-discovered importance of beliefs-qua-social-roles for the process of social predictions. In my view, however, such optimism is not enough to warrant the claim that beliefs are natural social kinds.

## 5. Faulty Arguments Against Realism

Some commentators have expressed misgivings about my arguments against realism. Zawidzki (2026) claims that it would be very hard to explain the entrenchment of folk psychology if there were no genuine generalizations about belief. This seems to me to be a variation on the theme of success argument: why would we use belief-talk if the attribution of beliefs did not lead to epistemic gains? But, to repeat myself, if the underlying sub-personal processes of mindreading and mindshaping are distinct from the outward processes of linguistic belief attributions, then this argument does not lead us to any form of realism. Moreover, the very idea of entrenchment might be simply false: if Kalis is right about the scope of variance in cross-cultural folk psychology, then beliefs are not as entrenched as is assumed.

### 5.1. Kozak on Rationality

Kozak (2026) offers an interesting argument for belief realism, which he calls “the rationality argument”, which is different from the standard success argument. This argument is based on the general idea that norms of reasoning can alter our behavior, especially how we reason. The very aim of there being norms of reasoning, according to Kozak, is that they can guide our reasoning. So, norms of rationality can alter our behavior (broadly conceived). If this is so, there has to be some intermediary between norms of rationality and behavior: beliefs are precisely what can be said to be this bridge. How can a norm of reasoning, say *modus tollens*, influence my reasoning practice? According to Kozak, it is only because I believe that said norm is true and correct. However, if this picture is on the right track, then beliefs must be treated as causes. Only if my belief that a given norm of reasoning is treated as a cause can we then treat it as something that changes my reasoning activities. This argument is not meant to show us what kind of theory of beliefs is correct: it could turn out to be either representationalist or dispositionalist. What needs to be accepted, as Kozak argues, is the basic realist commitment that beliefs are causally efficacious.

This is an intriguing line of thought, and it brings attention to certain, admittedly controversial, features of my position. If we treat denial of causal efficacy as a central commitment of non-realism about belief (as I am inclined to do), then a non-realist such as myself must reject Kozak’s conclusion. And, as his

arguments seem sound, the only way to do it is by denying the premise that norms of rationality can causally influence behavior (this sort of response was suggested to Kozak by Maja Kittel).

Contrary to what Kozak suggests, this does not mean that norms of rationality are not *applicable* to behavior. It is vital in this context to distinguish between norms of rationality being *applicable* to humans, and norms of rationality being able to *causally influence* behavior. That norms of rationality apply to human behavior means that we can judge ourselves and others by whether our reasonings adhere to norms of rationality. As I see it, we can do this without postulating that people can intentionally change their behavior by adopting norms of rationality.

The subtle and crucial difference here is between norms of rationality being “in force” and norms of rationality being prescriptive. This distinction was crucial in the debate about normativity of meaning back in the day (e.g., Hattiangadi, 2006). Semantic norms are, uncontroversially, in force, in the sense that people can and do judge others by the standards of linguistic correctness. But to say that semantic norms are prescriptive is to say something more controversial—namely, that people ought to follow them. But if it were true that people *ought to* follow semantic norms, it would mean that they should *be able to* decide to follow them. Still, it is far from obvious that many ordinary people have ever made any conscious decision of the form “from now on I will follow rules of standard English”.

In the case of logical norms, the very question of whether it is at all possible to decide to follow the rules of standard or non-standard logic is a subject of heated debate under the banner of the “adoption problem” (e.g., Birman, 2024). The worry is, in short, that it is unclear whether one can really *adopt* a logic, because doing so seems to presuppose already using a certain logic. So, one would need logic to adopt logic.

The issue of adopting logic is a vexed one, and I do not aim to claim that it shows the impossibility of norms of rationality influencing behavior. I only mean to show that the answer to the question of whether norms of rationality actually do influence behavior is far less obvious than Kozak suggests. It is true that an anti-realist about belief must deny that norms of rationality can be adopted by an intentional act. Indeed the central tenet of belief anti-realism is that the link between the “realm of reasons” and the “realm of causes” must be severed. However, my contention is that such an anti-realist picture is not as hopeless as is suggested.

The view of how a belief non-realist views norms of rationality goes roughly like this: We pick up, by means of sub-personal, non-intentional processes, certain ways of reasoning. These ways of reasoning are then subject to social criticism, as others might be prone to treating our ways of reasoning as deficient. Then, another subtle sub-personal, non-intentional process of mindshaping might get us to reason better. In this way, the norms of rationality are still applicable to us, and we still might get to be better reasoners, but the process does not involve

intentional acts of changing norms of reasoning.<sup>4</sup> I admit that denying that norms of rationality can be intentionally adopted might be a high price for belief non-realism. However, a non-realist about beliefs has a story about the role of norms of rationality, and this story might be closer to the truth than the one espoused by rationalistically inclined philosophers.

## 5.2. Horowitz on Fine-Grained Contents

Horowitz (2026) concurs that if we individuate beliefs by their folk contents, then beliefs should turn out to be non-natural and not causally active. However, he points out that one could also individuate the contents of beliefs in a more fine-grained psychofunctional manner. Horowitz gives Loar (1998) as an example of a theorist who distinguishes between contents of beliefs as we ordinarily attribute them and “psychological contents”, which are determined by the conceptual roles of the subject. Horowitz does not follow Loar, but he notices the possibility that if we individuated beliefs in a more-fine grained manner than we do in common practice, then such more finely individuated beliefs could play the explanatory role. So, if two subjects’ behavioral dispositions differ, then more finely individuated beliefs also differ. In this way, we could, potentially, save the idea of beliefs being explanatory because the differences in fine-grained beliefs are mirrored by differences in behavior.

My worry with this suggestion is that there is a whiff of circularity and arbitrariness about such a way of defining contents of beliefs. If we decide to reject folk attributions as a way to determine contents of beliefs, then there are very many ways in which we might determine contents in a more-fine grained manner. In fact, one could devise such a radically fine-grained way of content individuation such that no two people—nor even two distinct time-slices of the same person—could harbor beliefs with the same content. For example, the belief that Birmingham is north of Oxford that I held before I visited either of these cities might be said to have different psychofunctional content than when I had visited them because my, say, evidential situation became different. But now, as my memories of England fade, perhaps the fine-grained content starts to differ again. Or perhaps it has always been the same? How are we to tell?

Horowitz toys with the suggestion that we should differentiate the contents of beliefs on the basis of subjects’ behavioral dispositions, but this seems to be on the wrong track. If we define contents of beliefs by dispositions and then use this definition to show that beliefs are causally active, then this reasoning seems hopelessly circular. This is because we first use the causal profile to define beliefs, then we show that beliefs defined by reference to their causal profile indeed have a causal profile.

---

<sup>4</sup> This picture is obviously inspired by the Kripkensteinan skeptical solution to the problem of rule-following (Kripke, 1982).

All in all, Horowitz does not back the standard version of realism about beliefs, and I find his remarks about the limits of the success argument most helpful—specifically, his remark that “the explanatory and predictive success of a theory only indicates the truth of those parts of it that are explanatory or active in the theory’s explanations and predictions” (Horowitz, 2026, p. 62) seems to be right on track. Where I disagree with Horowitz is the extent to which the success of folk psychology commits us to anything resembling beliefs as ordinarily understood: I will contrast my views with the positive proposal of Horowitz in Section 6.

## 6. Consequences of Minimal Non-Realism

### 6.1. Schwitzgebel on Robots

In his intriguing review, Schwitzgebel (2026), who expresses sympathy with many aspects of my view, focuses on an entirely different issue: the problem of the consequences of the general superficialist approach (to which camp my proposal clearly belongs) to the issue of the attribution of mental states to robots. His contention is that the superficialist should welcome such a possible extension of the intentional idiom: given that some future robots would become behaviorally complex enough, then our linguistic practice would be such that we would find it apt to describe them using the intentional idiom.

I think that this is an extremely important aspect of my view that was somehow underdeveloped in the book: the question of the consequences of adopting minimal non-realism about beliefs for the issue of the correct scope of belief-attributions. *Prima facie*, as Schwitzgebel suggests, the consequence is clear: if we consider beliefs to be not strongly real, then we should allow for a significant degree of liberalism in the scope of beliefs’ attribution. If there is no hidden essence to beliefs, then there seems to be no grounds to deny that artificial systems are believers. Or so it seems.

On my view, this conclusion seems too hasty, and the link between superficialism and extreme liberalism about the scope of belief attributions—although in a way natural—could and perhaps even should be severed. As I see it, there should be a way for a superficialist to say that although a certain system seems, at first glance, to be a belief-holding system, in fact it is not really one. It would be, as I reckon, an extremely uncomfortable position to hold if one could not ever claim that what we are dealing with is a system that is just a seeming believer. For example, Dennett, who for a long time was a champion of artificial systems being true believers, ultimately rejected the idea that contemporary LLMs are true believers, calling them “counterfeit people” (Dennett, 2023). So, even a staunch superficialist might get fed up and decide that what we have in front of us is not the real McCoy.

The central question that Schwitzgebel’s commentary forces us to consider is the following: On what basis can a superficialist deny that a system has genuine

beliefs and relegate this system to the sphere of mere copycats? This is admittedly a vexed question that would require much more consideration that can be given here, but let me offer some preliminary thoughts.

The opponents of superficialism have it easier when it comes to denying the status of a true believer: a proponent of identity theory can appeal to similarities (or lack thereof) in brain structure; a dualist can appeal to non-physical phenomenology; and a representationalist can appeal to considerations about the cogs in cognitive machinery. A superficialist must, however, restrict themselves to purely dispositional considerations to achieve this result. So, my preliminary proposal is that in denying that a system is a true believer, we must resort to considerations about the richness of its dispositional profile. There can be a system that has *some* behavioral dispositions that are more or less the same as those of a typical believer; but this, after reflection, could turn out to be not enough to treat this system as a true believer. To classify a genuine believer, it is necessary to have these dispositions embedded in a thick web of many other dispositions that are characteristic of other mental states that typically interact with beliefs.

Consider a robot that has some belief-related dispositions: it can answer questions and even sometimes argue in favor of a given answer. It can also guide its behaviours using its “beliefs” to achieve some pre-determined aims. But this robot has no dispositions that would justify ascribing any independent motivations or plans to it (so, it is basically a mix of a chatbot and a Roomba). So, on the basis of its web of dispositions being too thin, a superficialist could tentatively justify their doubt that this robot is a true believer.

This is what differentiates my view from Schwitzgebel’s proposal. He also treats the thickness of the web of dispositions as a criterion for belief ascriptions, but his understanding of this web is far less substantial than mine. For Schwitzgebel, for a system to meet the superficial criteria for a believer, it is enough to have a coherent set of verbal dispositions and the ability to guide its own behavior (so, for him a, LLM-powered Roomba might “deserve” to be called a believer if there is enough consistency in its verbal outputs). On my view, this might not be enough. For a system to qualify as a believer, verbal dispositions and behavior guidance should be connected to a richer set of motivational and plan-related dispositions. The robot must be interpretable as being an independent planner, capable of, as it were, being its own center of actions and having independent motivations. In this way, as I see it, a dispositionalist might set a higher bar for believers than Schwitzgebel envisaged.

Additionally, as I indicated in Section 9.3 of my book, there is also an important pragmatic aspect to the question of the scope of belief attributions. This means that we should ask not only the question of whether we *can* extend the use of belief-attributions to robots, but also whether we *should*. Given that beliefs do not have hidden essence, there is a substantial wiggle room to make the concept of belief answer our pragmatic and moral needs. So, the question we might wish to consider before we agree to call robots believers is whether this kind of change in use would actually help us with our social and moral aims.

As I said, these are just preliminary thoughts, and I hope to be able to say more about restrictions of the scope of belief attributions in the future.

## **6.2. Mölder on the Aim of Attributions**

Mölder (2026, p. 10) raises an interesting and vexed question: What would be the aim of belief attributions if we revised the concept of beliefs along the lines that I proposed in the book? On my proposal, in order to retain the concept of belief, it would have to undergo a significant revision: we would have to drop the realist commitment to the idea of beliefs being explanatory of actions in a causal sense. Mölder then asks: What use would belief-attributions be in such a scenario if we are now using them to describe causal relationships?

This is indeed a very good question, and I can offer only a gesture at an answer here, with full knowledge that it demands a more thoughtful elaboration. The proposal would go along these lines: once we get rid of the causal-explanatory function of belief attributions, we would start to use these attributions as parts of certain stories that could render people's behaviours understandable. In this way, belief attribution practices would take an explicitly justificatory aim. Instead of asking the "Why did they do it?" question with an intent to look for causes, we would start to ask questions such as "In light of what reasons are their actions justified?"

Obviously, such a justificatory aim is already present in our current belief-attribution practice, but it is intermixed with the causal aim. So, the endgame of the proposed revision would not be to propose an entirely new function of folk psychology, but rather to purify it from the unwanted causal element. This might be seen as a highly speculative proposal, but let us remember that—according to non-causal theories of folk psychology—this is what our practice already looks like. So, my proposal is that even if folk psychology does not look like the non-causalists envisage it, we should try to make it look like that. Obviously, more should be said about the specifics of such a proposal, but this must be left for a later occasion.

## **7. Alternatives to Minimal Non-Realism**

My proposal of minimal non-realism was intended to be an alternative to both the strong realist view on the one hand, and to error theories on the other. As it turns out, there are many other theories of belief that try to achieve the same aim, and several commentators pointed to them in their reviews. Here I will discuss some of them.

### **7.1. Mölder on Interpretivism**

Interpretivism, fathered by Dennett (1989), and more currently championed by Mölder (2010) and Curry (2020), is a theory that places itself in roughly the

same place on the spectrum of realism about belief as my proposal. My earlier discussion of interpretivism (Posłajko, 2020) focused on the issue of judgement-dependence and metaphysical constitution, and Mölder in his commentary tries to answer precisely these challenges.

It must be admitted that Mölder's version of interpretivism, namely the canonical ascription approach, is a theory with which my view shares a great deal of affinity, and it served as a vital point of inspiration to my project. However, I still think Mölder fails to answer the most important challenge I raised to his view: to specify, in a non-trivial manner, in what sense are attitudes constituted. Mölder agrees with me that interpretivism should not overgeneralize so as to include cases such as boulders. The existence of boulders, even though they are pleonastic entities, does not involve any process of interpretation. However, on Mölder's version of the interpretivist view, the existence of beliefs is closely tied to interpretation. As Mölder puts it: "the relevant distinction lies between cases where meeting the application conditions of a term is straightforwardly empirical (as in the case of boulders) and cases where meeting them is sensitive to interpretation, as with mental states" (Mölder, 2026, p. 142–143). Furthermore,

mental state ascription is not simply a matter of applying fixed criteria in a straightforward empirical manner. Making an ascription requires constructing a coherent arrangement of information from various ascription sources. This arrangement is unique in each case of applying a mental term, such as "belief". (Mölder, 2026, p. 143)

This proposal, in my view, fails to answer the challenge fully. I am still not sure whether the difference between boulders and beliefs is, on Mölder's account, epistemological or metaphysical. Does the difference lie solely in the fact that it is epistemically harder to ascribe beliefs than to ascertain that a given rock is a boulder? If so, then this is something that nearly any middle-ground position in the metaphysics of beliefs would agree on. Thus, the epistemological reading does not distinguish between interpretivism and dispositionalism.

The metaphysical reading would provide the relevant distinction, but I have an impression that Mölder shies away from it. He invokes a "lightweight" version of constitution, but I fail to see how this version would provide a relevant difference. All in all, I think that much more would need to be done to provide a satisfying notion of constitution of beliefs, and this is, in my view, necessary to render interpretivism a satisfying alternative.

## 7.2. Bantegnie on Realist Dispositionalism

As was noted in Section 3.1 above, Bantegnie toys with the idea of what he calls the "realist dispositionalist" view (although he falls short of endorsing it). He considers two versions of it. The first, weaker one, postulates that beliefs are identical with clusters of dispositions, and that elements of these clusters are tied

by law-like generalizations. The stronger version postulates that these clusters share some categorical basis.

This is a *prima facie* attractive proposal and worthy of at least brief consideration, even if Bantegnie only presents it as a possible view. It is, to an extent, a simpler view than mine, as in my view belief-dispositions are somehow haphazard. So why should we reject realist dispositionalism? The answer is pretty simple—because such a view would require positive proof of there being generalizations that are responsible for clustering of belief-dispositions. My insistence on the idea that our belief-attribution practices pick up sets of dispositions that are not strongly related to each other is based on the epistemic principle that we need to have positive evidence to the effect that a given category is projectible in order to postulate it being so. Such available evidence is scant in the case of beliefs.

### 7.3. Horowitz on Syntax Without Content

Horowitz's view, which he also develops in his recent book (2024), is that in order to account for the success of folk psychology we should postulate internal psychofunctional states, but only postulate that these states have syntactic structures without propositional content. This is also an interesting position: it rejects the standard realist view because standard realism is tied to the idea of psychological reality of content. On the other hand, Horowitz's view is more realist than mine as it postulates that there is at least some correspondence between what we talk about in folk discourse and the internal reality.

This proposal would require much more engagement than can be provided here, but let me offer some preliminary remarks. What needs to be admitted is that a position like mine does carry a commitment to there being internal psychofunctional states which are the actual causes of human behavior. There is also no easy way to a priori exclude the possibility that such internal psychofunctional states would have some sort of syntactic structure. The question of there being such an internal syntax is an empirical one, and philosophers cannot answer it from an armchair.

What I take to be highly doubtful, however, is the idea that these possibly discovered psychofunctional states with syntax would have syntax that could reasonably mirror the syntax of beliefs as we ordinarily attribute them.<sup>5</sup> In short, postulating such syntactic mirroring would require believing in some sort of pre-established harmony, for which an extraordinary amount of evidence would be needed. And, as I have argued above, what we care about in debating the reality of beliefs is whether there is really something that is similar to beliefs as we ordinarily attribute them. One cannot exclude that there is some sort of internal

---

<sup>5</sup> I am unclear, however, to what extent Horowitz actually endorses the idea of such a mirroring. Section 5.3 of (Horowitz, 2024) does seem to imply certain skepticism about it, but more careful analysis would be needed.

syntax, but this hypothesis of internal syntax in itself does not save belief-realism, even in its weaker form.

Again, my argument relies on the idea that the lack of positive evidence for the realist hypothesis lends credence to anti-realism. However, I think that is exactly right: it is the belief-realist who incurs substantial metaphysical debts, and they need to provide us with positive, even if tentative, reasons to accept this view. And as the evidence is lacking, anti-realism should be the default position.

## REFERENCES

- Bantegnie, B. (2026). Dispositionalism and Realism. *Studia Semiotyczne*, 39(1), 147–157.
- Birman, R. (2024). The Adoption Problem and the Epistemology of Logic. *Mind*, 133(529), 37–60.
- Churchland, P. M. (1981). Eliminative Materialism and Propositional Attitudes. *The Journal of Philosophy*, 78(2), 67–90.
- Curry, D. S. (2020). Interpretivism and Norms. *Philosophical Studies*, 177(4), 905–930.
- Curry, D. S. (2026) Beliefs Are Real(ly for Social Animals). *Studia Semiotyczne*, 39(1), 113–130.
- Demeter, T. (2009). Two Kinds of Mental Realism. *Journal for General Philosophy of Science*, 40(1), 59–71.
- Dennett, D. C. (1989). *The Intentional Stance*. Cambridge, MA: MIT Press.
- Dennett, D. C. (1998). Two Contrasts: Folk Craft versus Folk Science and Belief versus Opinion. In D. Dennett (Ed.), *Brainchildren* (pp. 135–148). Cambridge, MA: MIT Press.
- Dennett, D. C. (2023). The Problem With Counterfeit People. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2023/05/problem-counterfeit-people/674075/>
- Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Vol. 2). Cambridge, MA: MIT press.
- Godfrey-Smith, P. (2005). Folk Psychology as a Model. *Philosopher's Imprint*, 5(6).
- Hattiangadi, A. (2006). Is Meaning Normative? *Mind & Language*, 21(2), 220–240.
- Hensel, W. (2026). Troubles With the Natural-Kind Approach to the Debate Over the Reality of Belief. *Studia Semiotyczne*, 39(1), 29–47.
- Horowitz, A. (2024). *Intentionality Deconstructed: An Anti-Realist Theory*. Oxford University Press.
- Horowitz, A. (2026) Poślajko on Belief Similarities and Belief Causation. *Studia Semiotyczne*, 39(1), 49–66.
- Kalis, A. (2026). The Folk Are Doing Fine (But the Philosophers Are in Trouble). *Studia Semiotyczne*, 39(1), 89–100.

- Komorowska-Mach, J. (2026). Rethinking Folk Realism About Beliefs. *Studia Semiotyczne*, 39(1), 101–111.
- Kozak, P. (2026). Why Should We Be Realists About Beliefs? The Rational Indispensability Argument for Realism. *Studia Semiotyczne*, 39(1), 67–88.
- Kripke, S. A. (1982). *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Harvard University Press.
- Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61(4), 343–377.
- Loar, B. (1988). Social Content and Psychological Content. In R. H. Grimm, D. D. Merrill (Eds.), *Contents of Thoughts* (pp. 99–110). Tucson: University of Arizona Press.
- Mölder, B. (2010). *Mind Ascribed. An Elaboration and Defence of Interpretivism*. Amsterdam: John Benjamins.
- Mölder, B. (2026). Belief: Minimal Non-Realism and the Constitutive Role of Interpretation. *Studia Semiotyczne*, 39(1), 131–146.
- Neufeld, E. (2022). Psychological Essentialism and the Structure of Concepts. *Philosophy Compass*, 17(5). <https://doi.org/10.1111/phc3.12823>
- Posłajko, K. (2020). Can Deflationism Save Interpretivism? *Philosophia*, 48(2), 709–725.
- Posłajko, K. (2022). The Lycan-Stich Argument and the Plasticity of “Belief”. *Erkenntnis*, 87(3), 1257–1273.
- Posłajko, K. (2024). *Unreal Beliefs: An Anti-Realist Approach in the Metaphysics of Mind*. Bloomsbury Publishing.
- Ryle, G. (2009). *The Concept of Mind*. New York: Routledge.
- Schwitzgebel, E. (2025, March 06). Kings, Wizards, and Illusionism about Consciousness. *The Splintered Mind*. <https://eschwitz.substack.com/p/kings-wizards-and-illusionism-about>
- Schwitzgebel, E. (2026). Superficialism About Belief, and How We Will Decide that Robots “Believe”. *Studia Semiotyczne*, 39(1), 159–171.
- Thomasson, A. L. (2015). *Ontology Made Easy*. Oxford: OUP.
- Wright, C. (1993). *Realism, Meaning, and Truth*. Blackwell.
- Zawadzki, T. (2026). Locating Belief on the Spectrum of Existence. *Studia Semiotyczne*, 39(1), 15–27.