

WITOLD M. HENSEL *

TROUBLES WITH THE NATURAL-KIND APPROACH TO THE DEBATE OVER THE REALITY OF BELIEF¹

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.

¹ 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 H₂O.

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
- 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>