STUDIA SEMIOTYCZNE (SEMIOTIC STUDIES), 39(1), 11–23
ISSN 0137-6608, e-ISSN 2544-073X
DOI: 10.26333/sts.xxxix1.02

Article

TADEUSZ ZAWIDZKI *

LOCATING BELIEF ON THE SPECTRUM OF EXISTENCE

SUMMARY: Krzysztof Posłajko'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" (Posłajko 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 Posłajko's favored "minimal non-realism". However, as Posłajko realizes, the distinction between natural and non-natural properties is "graded" (Ibid), e.g., categories of the special sciences, like money, are not *as natural* as the categories of our best physics, but they're 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 Posłajko 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. Posłajko'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. 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 don't 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" (Ibid) 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*" (Ibid). 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 (Ibid, p. 64). Despite Posłajko's laudable metaphysical nuance concerning this general point, he is surprisingly inexplicit about *precisely where* on this ontological "spectrum" (Ibid, 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 (Ibid, pp. 98-112), it's not clear how far towards the "natural" end

of the spectrum belief belongs. Posłajko 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 (Ibid, pp. 106-112) and classic puzzles about mental causation (Ibid, 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 Posłajko 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's 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.

Posłajko 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" (1991; Posłajko 2024, p. 55). He argues that, unlike categories posited by special sciences, belief does not qualify as a natural kind according to this criterion (Ibid, 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 Posłajko 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 Posłajko's concerns regarding revisionism about the folk concept of belief: although it's 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 Posłajko 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" (Ibid, 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" (Ibid, 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 et al. 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,

> [t]here 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" (Ibid, 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 (Ibid, 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 (Ibid, 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 (Ibid, 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. (Ibid, 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. (Ibid, 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 (Ibid, 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" (Ibid, 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" (Posłajko, p. 68) mechanism which explains this: natural selection. For Dennett, it is not *just* that, as Posłajko notes, "intentional interpretation reveals important behavioural patterns to us … attributions of beliefs seem to capture some important objective similarities" (Ibid, p. 81). There is *also* a fundamental physical mechanism which explains this: natural selection. For this reason, Posłajko 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" (Ibid). 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 Posłajko notes, the folk seem committed to the view that beliefs are concrete causes of behavior (Ibid, 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" (Ibid, 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. Posłajko explicitly problematizes such revisionist proposals (Posłajko 2024, pp. 153-154); however, as I argue in the concluding section, it is unlikely that *any* theory of belief, including the representationalist theory Posłajko 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 isn't 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 Xs 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. (Ibid, 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). 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 (Ibid). This interpretive practice is endlessly dynamic and recursive: discursive statuses are constantly being revised and negotiated, partly because their attributions also constitute contestable moves in the game. However, such a game could never get off the ground if there weren't

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. Posłajko himself acknowledges that "it seems perfectly reasonable to engage in inquiry into the patterns of language use" (Posłajko 2024, p. 143) in support of inferentialist theories of meaning like Brandom's (Ibid, p. 143). And deontic attitudes might reasonably be made sense of in terms of Strawsonian "reactive attitudes" (1963), and "normative cognition" (Kelly et al. 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. Posłajko 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; Posłajko 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 Posłajko'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, n. 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. Posłajko 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[1], 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

[1] 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 et al. 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 needn't 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 Posłajko 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 Posłajko'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 Posłajko'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 Posłajko'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). 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.