The Lycan–Stich Argument and the Plasticity of “Belief”

Krzysztof Poslajko¹

Received: 12 April 2019 / Accepted: 20 March 2020 / Published online: 5 April 2020
© The Author(s) 2020

Abstract

The aim of this paper is to argue against the claim that the term “belief”, as it functions in philosophical psychology, has natural-kind term semantics; this thesis is central to the famous Lycan–Stich argument against eliminative materialism. I will argue that the current debate concerning the discrepancy between the professed opinions and actions, especially the debate concerning the idea of aliefs, shows that the concept of belief is plastic and amenable to conceptual engineering. The plasticity and amenability to conceptual engineering of the concept of belief give us, in turn, a reason to doubt that “belief” functions in a way that is presupposed in the Lycan–Stich argument. Finally, I point to an alternative to both eliminativism and the natural kind view, namely the idea that we should treat belief as a human kind.

1 Introduction

In our folk psychological practice we not only ascribe mental states to ourselves and other people, but in doing so we also implicitly categorize these mental states into several distinct kinds, like “beliefs”, “desires”, “imaginings” and the like. The philosophical question that arises with regard to this classification practice is whether the classification we ordinarily employ is the best one, or even that it is correct.

This question is of particular importance in the context of two important debates in contemporary metaphysics of mind. First, it is central to the famous Lycan–Stich argument against eliminative materialism: Lycan and Stich argue that eliminativism about propositional attitudes can be shown to be unjustified once we admit that the term “belief” might be thought of as having natural-kind term semantics.

The second issue in which the question about the relevance of our folk classifications of mental states plays an important role is the debate about the discrepancy between professed beliefs and actions; there is significant disagreement among philosophers of mind over whether situations in which a certain subject professes

✉ Krzysztof Poslajko
krzysztof.poslajko@uj.edu.pl

¹ Department of Philosophy, Jagiellonian Center for Law, Language and Philosophy, Jagiellonian University, ul. Grodzka 52, 31-044 Kraków, Poland

to have a certain belief yet their actions are inconsistent with this assertion should be classified as cases of belief. This debate raised some important methodological issues about the classifications in folk psychology that, as I will try to show, have had some far-reaching implications concerning the status of belief as a kind.

In this paper I will try to show how the debate about the phenomenon of discrepancy—and, more specifically, the dispute about the status of Gendler's (2008a, b) alief proposal—shows that we should not treat “belief” as having natural-kind term semantics. The basic idea of the argument I am going to present is that it is possible to show that “belief” is a plastic concept and that this casts doubts on the idea that this term should be treated as having natural-kind term semantics.

The plan of the paper is as follows: first, I will present the Lycan–Stich argument against eliminativism and show which assumptions need to be made for this argument to work. Then I will present the debate about the phenomenon of discrepancy and various philosophical reactions to it. I will focus on the alief proposal and the methodological debate with regard to it. I will argue that the best reading of this proposal is that the introduction of aliefs should be seen as a kind of conceptual engineering of the concept of belief. Then, I will argue that this leads to the conclusion that the concept of belief is plastic, which, in turn casts doubt on the claim that “belief” has natural-kind term semantics. Finally, I will return to the original issue of eliminativism: I will argue that although the Lycan–Stich argument fails, we are not forced to embrace eliminativism; an alternative to both eliminativism and the natural kind approach to belief is the conception according to which beliefs constitute a human kind; this conception will be briefly sketched in the final part of the paper.

2 Lycan and Stich Against Eliminative Materialism

The traditional form of eliminative materialism about propositional attitudes, endorsed by Churchland (1981), maintained that beliefs, understood as theoretical posits of folk psychology, do not exist. The reasoning of Churchland was quite elegant in its simplicity: according to him, we should treat folk psychology as a kind of proto-theory and assess it according to the rules all theories should meet. As folk psychology fails this assessment, we should deem it false. Thus, as Churchland concludes, we might claim that there are no beliefs (and desires, and so on) as they are posits of a false theory. Unsurprisingly, this conclusion has generated several controversies. This conclusion has been widely held to be flawed and many supposed that there must be some error in the argument leading to it.

One of the most popular strategies of debunking the Churchland eliminativist argument was presented by Lycan (1988) and developed by Stich (1996). As Lycan and Stich noticed, Churchland's argument relied on the assumption that the meaning of “belief” is determined by how it is understood in the context of folk-psychological theory. They also noted that this was not the only—and perhaps not the best—way of thinking about the meaning of the term “belief”.

The alternative to the descriptivist view is the claim that “belief” has natural-kind term semantics. Lycan described this hypothesis in the following way:

I think the ordinary word ‘belief’ (qua theoretical term of folk psychology) points dimly toward a natural kind that we have not fully grasped and that only mature psychology will reveal. I expect that ‘belief’ will turn out to refer to some kind of information-bearing inner state of a sentient being, (. . .) but the kind of state it refers to may have only a few of the properties usually attributed to beliefs by common sense. (Lycan 1988, p. 32).

The important feature of natural-kind term semantics is that according to this theory the speakers might be ignorant of the true nature of the kind they refer to and still be able to refer to such a kind. In Putnam’s (1975) famous example, when Archimedes used the Greek term for “gold” he referred to the same substance as we do by “gold”, even though he held some false beliefs about gold. Our situation with respect to beliefs is, as Lycan seems to suggest, similar to the situation of Archimedes with respect to gold. We might be in error about the nature of beliefs, but still we might refer to something real. This line of argument was even more explicitly restated by Perez (2004), who makes the straightforward transition from the claim that folk-psychological terms possess natural-kind term semantics to a realist claim concerning folk-psychological terms.

Stich (1996) and his collaborators (see Mallon et al. 2009) are more cautious in this respect. They do not claim that “belief” actually has natural-kind term semantics, rather they point out that this is an epistemically open possibility. They suggest that which semantic theory is the proper one with respect to “belief” is somehow undetermined. Still, the consequence of this more modest claim is that eliminative materialism rests on shaky ground, as one of the central assumptions of eliminativism is unfounded. Stich (1996) goes as far as to argue that if there is no correct answer to the question about the right theory of reference for the term “belief”, then there is no correct answer to the question of whether beliefs exist at all. Nonetheless, according to Stich even this weakened conclusion makes eliminativism an unwarranted position.

3 Analysing the Lycan–Stich Argument

In order to assess the validity of the Lycan–Stich argument, it is useful to make explicit its basic assumptions in order to see whether they are justified. One needs to distinguish two claims which are being made when one maintains that “belief” denotes a natural kind. First, there is a metaphysical claim that there is such a natural kind as belief, i.e. that the world is structured such that there is an appropriate mental/cognitive kind. The other claim is a semantic one that the meaning of the term “belief” is taken to be determined by the direct reference to a certain natural kind.

These two claims are logically independent. There might be a cognitive natural kind that is denoted by the term “belief” even if the term in question does not have natural-kind term semantics. As was noted by LaPorte (2000), it is possible to refer to natural kinds by description; for example, one can refer to the (arguably natural) kind *Apis mellifera* using the descriptive phrase “the insect species that is typically farmed for honey”. This shows that the fact that a term “K” denotes a natural kind

does not in itself show that the term in question works in the way the proponents of natural kind semantics envisage: the term might still be a descriptive one.

On the other hand, the semantic claim does not entail the metaphysical one either: it might be the case that a term “X” has natural-kind term semantics, but the natural kind in question does not exist, as in the cases of “phlogiston” or “unicorns” (see, e.g., Besson 2012 for a useful overview of the issue of empty natural-kind terms).

Thus, in order to claim that the term “belief” successfully refers to a natural kind, one must endorse two logically independent claims. First, that there is a cognitive natural kind such that it could serve as a referent for “belief”. Second, that the term in question has the appropriate semantic features—that it refers to the kind in question in the way that theorists of natural-kind term semantics envisaged.

The first metaphysical question has generated some attention in the recent literature: for example, Jenson (2015) argues that belief does not meet the criteria of being a natural kind, as this putative kind is fragile [i.e. “the results of multiple, independent, putatively reliable measures of that entity turn out to radically vary and this variation cannot be adequately explained away” (Jenson 2015, p. 970)] and there are no substantive inductive generalizations about this kind (however, it should be noted that Quilty-Dunn and Mandelbaum (2018) claim that there are strong generalizations concerning beliefs in cognitive science).

However, in what follows I will focus on the semantic question. This is because, even if it were possible to defend the claim that there is such a natural kind, it would not be enough to make the Lycan–Stich argument valid. One also needs to show that the term “belief” has appropriate semantic properties. Again, the question of what the right theory of reference is for this term has been subject to some scrutiny in the recent literature: Hensel (2016) critically analyses some putative psychological support for this claim, while Pinder (2017) argues that Churchland’s commitment to descriptivist theory of reference might be defended, given the premise that Churchland was involved in the project of the explication of “belief”. I will tackle this debate from a different angle and argue that the way the concept of belief functions in the discourse makes the natural-kind term semantics account of this term implausible.

In order to show this, let us first focus on the contrast between the semantics for “belief” which is assumed by the eliminativist and by the proponent of the Lycan–Stich approach. In both theories, “belief” is taken to be a theoretical term, i.e. it is treated as a term which functions to denote kinds of objects that are not directly observable.

As Stich (1996) rightly notes, there are two main competing views about the meaning and reference of theoretical terms. The first is the descriptivist view, which is most fully developed in the Ramsey–Carnap–Lewis tradition [for Stich, the best exposition of this view is to be found in Lewis (1970, 1972)], and the other is the direct reference or natural-kind term semantics view which was developed by Kripke (1980) and Putnam (1975).

According to the descriptivist view, the meaning and reference of a theoretical term are determined by the role the term in question plays in a certain theory. What is needed to establish the meaning of a theoretical term is the whole story which the theory in question tells about the kind denoted by the term. If we want to know

what, for example, the meaning of the term “species” is according to cladistics, we must assemble all the most important theoretical claims about species that this theory makes, make a long conjunction of these claims, and construe a definition out of it [this is of course a gross oversimplification of the procedure, and the details are presented by Lewis (1970, 1972)]. The most important feature of this approach is that if it turns out that nothing satisfies the conjunction, then we are entitled to say that species do not exist.

On the direct reference view, the reference of certain theoretical terms (namely natural-kind terms) is fixed via baptism and a reference–transmission mechanism. A natural-kind term refers directly to a natural kind, no matter what theory about the kind we happen to employ. As Putnam famously argued, the ancient Greeks could successfully refer to “gold” even though their theory about gold was probably mostly wrong. Natural-kind terms are, in Putnam’s view, trans-theoretical terms: this means that they continue to refer to the same kind despite our theories about them changing.

It should be obvious that in order to claim that the folk-psychological notion of “belief” can survive even if we agree that folk psychology is a seriously defective theory, one must embrace the view that “belief” is a trans-theoretical term; its reference must be fixed independently of the theory in question. The question of which theory of reference is the correct one for “belief” is, admittedly, a hard one. Stich (1996) even goes as far as to claim that there might be no correct answer to this question. This is because, as Stich contends, there is no clear answer to the question of what counts as a correct theory of reference.

Stich is definitely right in claiming that there is no easy answer to the question of whether “belief” refers directly. There is no quick route to establishing what theory of reference is correct for a given theoretical term, and testing lay-people intuitions might lead to inconclusive results (see Mallon et al. 2009).

In what follows, I argue that there is a *prima facie* argument against treating “belief” as a directly referential, trans-theoretical term, and to see the meaning of this term as being theory dependent. The argument rests on the idea that the concept of belief, at least as it is employed in philosophical psychology, is a plastic one in the sense of semantic plasticity developed by Dorr and Hawthorne (2014). If this could be shown, then we would have a reason to doubt the claim that this term’s reference is independent of theory.

In order to show that the concept of belief is plastic, I will present the phenomenon of discrepancy and one of the accounts of this phenomenon, namely the conception of aliefs. Then I will try to show that the best way to account for the status of alief proposal is to see it as an example of conceptual engineering. This in turn shows that the original concept of belief is plastic and amenable to conceptual engineering. Finally, I will try to show how this plasticity casts doubt on the claim that “belief” is directly referential.

One caveat is needed at this point. In what follows I will be primarily interested in the way the concept of “belief” functions in what might be called “philosophical psychology”, i.e. the folk-psychological discourse as it is being refined and used by philosophers in the broadly understood analytic tradition. This is an important reservation as there might be possible contexts in which someone would start to

use the term “belief” with an explicit intention to ‘baptize’ a natural kind. However, the realism versus anti-realism debate concerning beliefs targets the belief concept which is taken from the folk use and then refined within the philosophical discourse and not a homophonic counterpart of it. Should, say, computational neuropsychologists start to use the term “belief” to denote a putative cognitive natural kind, with little regard for the folk use of the term, there would be nothing “semantically wrong” in that. However, such a development would not, in my opinion, show that “belief”, as it is used by philosophers reflecting on the folk use, has a natural-kind term semantics. This is because such use would be in a way disconnected from the meaning of “belief” as it is conceptualized by philosophers of mind.

4 The Phenomenon of Discrepancy and the Concept of “alief”

Both common sense and scientific psychology acknowledge the existence of a discrepancy between apparently sincerely professed opinions and actions. In philosophical literature this discrepancy is illustrated by many familiar cases (see e.g. Gendler 2008a; Tumulby 2014). We have an overtly anti-racist professor who is verbally committed to the idea of all people being equal, and yet subtle behavioural cues suggest that she is, in fact, racially prejudiced: she looks surprised when a non-white student comes up with a well-informed answer to a question and takes it for granted when a white person does so (and so on and so forth). We also have the clever engineer who devised the glass walk above a canyon and after triple-checking all the calculations and doing the relevant stability tests is absolutely sure that the glass walk is perfectly safe. Anyhow, when she is presented with the opportunity to walk on it herself, she is very hesitant.

There are obviously many such cases of a mismatch between what is professed and actual behaviour. That such mismatches exist is obvious from the point of view of both folk psychology and contemporary empirical psychology. The philosophical question, however, is how we should conceptualize these cases: do the heroines and heroes of these short stories really believe what they are saying? In all the presented cases, their heroes and heroines are sincere in what they profess, but the cases still invite some scepticism about whether they believe in what they say, so we cannot suspect them of being hypocritical. However, the question is whether their attitude towards the professed content is that of belief, given their divergent behaviour.

There are several different possible answers to this issue: some philosophers have, for example, claimed that the discrepancy is best explained by introducing a postulate of in-between beliefs, i.e. by claiming that the ascriptions of beliefs in certain cases might lack definite truth value (see Schwitzgebel 2010; Tumulby 2014). Others have asserted that we might just as well claim that people have inconsistent beliefs (Gertler 2011). Yet another possibility would be to say that people simply have the beliefs that can be “read” from their behaviour, and their professed opinions stem from the fact that they are victims of self-deception (this view is espoused by Rey (2007), for whom the paradigmatic case of discrepancy are religious claims).

Perhaps the most interesting approach to the phenomenon of discrepancy was proposed by Gendler, who proposes distinguishing between beliefs and aliefs, where

aliefs are associative, automatic, arational, affect-laden and automatic mental states (Gendler 2008a, b). According to her, in the situations of discrepancies described above we should think of the subjects in question as having the beliefs they claim to have, but their conflicting behaviour should be accounted for by claiming that they also have aliefs conflicting with their beliefs. Thus, for example, in the glass walk case the person believes that the glass walk is safe, yet she alieves that it is not; similarly, the professor believes that people of all races have the same intellectual capacities, but she alieves that people of colour are somehow intellectually inferior. Basically, in each case of discrepancy it might be said that the subject believes what they profess, but their actions are guided by their aliefs.

5 The Status of the Alief Proposal

This solution is, as it seems, quite elegant: it explains the discrepancy without calling into question the idea that the subjects believe what they claim to. Despite the elegant simplicity of Gendler’s proposal, the claim that there are aliefs has generated several controversies, some of which concern the question of whether introducing aliefs really is the most effective way of explaining the phenomenon of discrepancy (see Albahari 2014 for an overview).

Several philosophers, however, have voiced a distinct kind of worry about the idea of aliefs; they pointed out that there are certain methodological and metaphilosophical problems with the very idea of introducing a previously unknown kind of mental state (see e.g. Bayne and Hattiangadi 2013; Matthews 2013; Muller and Bashour 2011; Thompson 2016). The question these philosophers ask is how one could justify the introduction of a new category of mental states into our conceptual repertoire, and what kind of evidence one might use in favour of the claim that there is indeed such a previously unencountered kind.

One of the clearest, although slightly dramatic, expressions of the metaphilosophical problems with the notion of alief is Sylvan Bromberger’s question (revealed to Gendler in personal communication, see Gendler 2008b). The question is why Gendler “should have been so fortunate to have discovered a category of thought that has evaded the eyes of philosophers for two millennia” (Gendler 2008b, p. 557).

The Bromberger question might be treated as an expressive device which draws our attention to a central issue with regard to the methodological status of the alief theory (this point is stressed by Matthews 2013). The intuitive worry seems to be how a philosopher of mind is able to make a true claim about there being a new previously unknown kind of mental state. Does it make sense to speak about a genuine discovery in this context? The case of Gendler introducing the notion of alief is, at least on the level of primitive intuitions, very much unlike the case of Māori people discovering New Zealand or Marie Skłodowska–Curie discovering radium.

The model of discovery seems obviously out of place in the case of aliefs: there are no interesting empirical generalizations about aliefs, and the term in question is not a part of scientific psychology. Many (see Bayne and Hattiangadi 2013; Thompson 2016) have noted that in order to claim that aliefs have been discovered, one would have to show that there are inductive generalizations about aliefs, as in order

to claim that one has discovered a new natural kind in psychology, one must show the aptitude of this putative kind to function in inductive generalizations. These generalizations about aliefs, however, are notably absent.

Partly in response to these worries, Gendler underscored the importance of pragmatic considerations in the introduction of this term. As she wrote:

The alief papers make the simple claim that it is conceptually useful to have a single category for the family of mental states that accompany these propensities (...) The main point is simply that in developing a relatively coarse-grained folk psychology – one that makes use of the notions of belief and desire to explain intentional behaviours, and of the notion of reflex to explain behaviours that consistently occur in response to stimuli in the absence of representational mediation – it is crucial to have an intermediate category (Gendler 2012, p. 799–800).

Other participants of the alief debate have also seen the pragmatic aspect of this proposal as central; for example, Tumulty (2014), who rejects the alief proposal and argues for the in-between beliefs option, justifies this solution by appeal to normative and pragmatic considerations (specifically, it is important for her that folk psychological attributions retain their role in shaping behaviours).

Given the fact that there is little reason to suppose that aliefs have been discovered (in any meaningful sense of ‘discovery’) and that an important motivation for the introduction of this category is pragmatic and normative, it is best to see the alief proposal as an example of conceptual engineering. “Conceptual engineering” is a fashionable term which nonetheless lacks a universally accepted definition. It might best be regarded as a blanket concept that is used to denote a family of views that is associated with theories of philosophers as diverse as Haslanger (2000), Plunkett (2015), Eklund (2017), Cappelen (2018), Machery (2017) and Chalmers (2011) (many of whom do not embrace the label). In the present paper I do not wish to subscribe to any of the existing conceptions of conceptual engineering, nor to propose a new theory of conceptual engineering, but rather to adopt a rough and ready concept of conceptual engineering and to apply it to the case at hand.

According to this simplified vision, conceptual engineering is a philosophical activity that aims to improve our defective representational tools. A philosopher involved in conceptual engineering recognizes that the ways in which we think about a given phenomenon are somehow defective and proposes change in our concepts/ways of representing the world. The question that a philosopher involved in the practice of conceptual engineering asks is “What concepts should we use?”, rather than “What concepts do we use?”, which is often associated with traditionally understood analytic philosophy. An important aspect of the idea of conceptual engineering is that often such changes to our concept are motivated by pragmatic and normative considerations.

I think that the alief case nicely fits this general schema of conceptual engineering. Gendler identified an important deficiency in our conceptual repertoire that we use to think and talk about mental states, namely that it does not provide a good explanation for the discrepancy between professed opinions and actions. This deficiency shows up in the fact that competent speakers of English are often unable,

when confronted with typical cases of discrepancy, to decide whether to ascribe beliefs to the subjects in question. The fact that we are not certain what to say when confronted with such cases might be seen not as a proof of our cognitive shortcomings, but rather as the result of the fact that our representational devices that are designed to track mental states are not fine-grained enough to capture certain phenomena. Thus, the idea of aliefs should be seen as a proposal of a re-conceptualization of the ‘belief discourse’. The proposal is that we restrict the concept of belief to certain cases and introduce a new concept to include some aspects of the problematic cases.

6 Status of Aliefs and Plasticity of “Belief”

The conclusion that “alief” should be seen as an engineered concept and not as a discovered kind might not seem very surprising. However, as I will presently argue, it has significant ramifications for the status of the concept of belief as it shows that we might treat the concept of belief as a plastic one.

The idea of concept plasticity has been developed by Dorr and Hawthorne (2014) and expanded upon by Cappelen (2018). According to Dorr and Hawthorne’s definition, a given concept is plastic if we might use the term related to the concept in question with a changed extension (and probably intension), yet it seems we are talking about the same thing. For example, the concept of salad is plastic as, arguably, both the extension and the intension associated with it have significantly varied through history. This variation, however, does not lead us to the claim that we have somehow changed the subject in our salad-talk. Rather, the idea is that a plastic concept such as “salad” can survive changes to its extension and intension.

When an expression is semantically plastic, then, according to Dorr and Hawthorne and Cappelen, there is a whole range of possible meanings of this expression; as Dorr and Hawthorne put it, “an expression’s actual meaning is surrounded by a vast cloud of slight variants that seem just as well qualified to be possible meanings” (Dorr and Hawthorne 2014, p. 282). Plastic expressions are those which can take any of these distinct possible meanings and still count as being the same expression. There is no uniquely ‘best’ meaning for a plastic expression, and we are able to switch between those possible candidates without changing the subject.

The thought that certain concepts might be plastic is essential to the project of conceptual engineering. If we want to say that a concept of e.g. “woman” should be amended, then we must ensure that the concept that will be the effect of our conceptual tinkering would be still the same one, otherwise the proponent of conceptual engineering of the given concept would be easily accused of changing the subject (this issue is discussed at length by Cappelen 2018). Thus, the plausibility of the idea of conceptual engineering depends on the plausibility of the idea that some concepts are indeed plastic.

It is important to note that the claim about the plasticity of a concept is not a simple observation that a given concept might have fuzzy/vague boundaries. The basic notion behind the idea of vagueness is that there are certain objects that we are unable to classify as being or not being part of the extension of the concept in question.

The idea of plasticity runs much deeper: the idea is that a concept might change in time and yet remain the same concept. The ordinary concept of “green” is most probably vague (there are things which we cannot classify as being neither green nor not-green), but it seems to be a relatively stable one: both the extension and the intension of this concept have probably not changed much in recent centuries. What is important to the idea of plasticity is that a concept might undergo changes, both in extension and intension.

The conclusion that the concept of “alief” is an engineered one gives us a strong reason to suppose that the concept of belief is a plastic one. The argument goes as follows: the possibility of introducing a new concept to our classification of mental states shows that our conceptual repertoire that is used to ascribe mental states is amenable to change. This in turn shows that the concept of belief is plastic as it can survive said change.

For all the opposition that the idea of aliefs has faced, the baseline of the discussion seems to be that Gendler’s proposal is worthy of consideration, even though it might not be, all in all, feasible. The idea of reorganizing our conceptual apparatus regarding our mental states in order to make room for a new category of previously unacknowledged states might be thought not to be particularly useful, but there is no suggestion that the whole enterprise is misguided.

If we were to accept the alief category in our conceptual repertoire, this would inevitably change the concept of belief because the concept of aliefs is engineered via, as it were, carving out conceptual space from the existing concept of belief. It is important to note that aliefs are not devised as a sub-category of beliefs: Gendler does not claim that aliefs are simply subspecies of an existing category of belief. They are a wholly new category that is distinct from beliefs. They differ from beliefs in many respects, and alieving that *p*, in most situations, means not believing that *p*. If we consider the example of a “verbally non-racist” professor, then, according to Gendler’s analysis, she believes that all people are equal, yet she alieves that non-white students are not as academically apt as white ones. In the case of the skywalk, the person believes that it is safe to walk on, but alieves that it is not, etc. In all cases, the idea is that the mental reality which we tried to conceptually cover with just one category (that of beliefs) would be better understood if we used two distinct concepts. In this way the possible introduction of the concept of aliefs would also change the existing concept of belief.

If such a change is possible, then it shows that the concept of belief is plastic enough to survive the change that consists of excluding the alief cases from the extension of the concept in question; it also seems plastic enough to survive the accompanied change in intension. After the possible introduction of the concept of alief, the concept of belief would change its content: it would become not only more precise, but the engineered version of belief would be more focused on professed and conscious opinions than the pre-engineered version of “belief”. When we use a pre-engineered version of the belief concept, we are often uncertain whether to apply the concept to certain cases (such as the unconsciously racist professor case), because our concept of belief requires some sort of conformity of actions and declarations. However, if we accept Gendler’s proposal, then we should have no problem in attributing the belief that all people are equal to the professor in question, as in

the new version of the belief concept what counts for belief attributions are, primarily, the professed opinions.

There are, as it has been noted, other ways in which the concept of belief might be altered in order to accommodate the phenomenon of discrepancy: one might opt for Schwitzgebel (2010) option and postulate states such as in-between beliefs; alternatively, one might claim that, despite our initial reservations, we should be ready to attribute inconsistent beliefs to subjects in such situations (Gertler 2011). Even though these changes are not quite as dramatic as the one proposed by Gendler, they also point to the possibility of rearranging our concepts which serve to describe our mental life. Thus, it seems right to claim that our belief-related conceptual scheme is plastic and amenable to conceptual engineering.

Given this conclusion, one might ask whether the differences between the different solutions to the issue of discrepancy are not merely verbal: once we admit that the concept of belief is plastic, are we not forced to conclude that the parties to this debate just talk past each other, as they are using different precisifications of the concept of “belief”? Admittedly, the debate is not substantial in the sense that there are no facts about the objective boundaries of belief that would settle it. However, it does not follow that the debate is “merely verbal” in the disparaging sense; rather it should be seen as a case of a pragmatically and normatively motivated dispute about the proper way of using the word in question; this debate can be seen, for example in the model of meta-linguistic negotiations (see Plunkett 2015).

7 Plasticity of “Belief” and Natural-Kind Term Semantics

The idea that the concept of belief is amenable to conceptual engineering has important implications for the debate on the status of “belief” as a possible natural-kind term: the plasticity and amenability to engineering of the concept of belief shows that this concept cannot be treated as having natural-kind term semantics.

To see how the question of plasticity weighs on the issue of “belief” having natural-kind term semantics, it is useful to employ the concept of candidate meanings, i.e. the possible extensions and intensions a given term might have. There appear to be several candidate meanings for the term “belief”. There is B_0 , i.e. the concept that we presently use (arguably an open-ended and vague one), but also there are several possible alternative concepts we might decide to employ. One of them is the concept as amended by Gendler (we might call it B_1), which differentiates between aliefs and beliefs; the other alternative is the concept amended according to Schwitzgebel’s suggestion to include the idea of in-between beliefs (B_2); yet another (B_3) is a concept of belief that includes the possibility of inconsistent beliefs; there are also possibly many more. The idea that the concept of belief is plastic in the relevant sense boils down to the suggestion that the choice of either of these candidate meanings (B_1 , B_2 ,... etc.) does not constitute changing the subject. In other words, we might decide on one of these candidate meanings and still talk about the same thing, namely beliefs.

The observation that the concept of belief is plastic leads to the conclusion that the boundaries of the kind in question might be changed via normatively and

pragmatically guided conceptual decisions. This conclusion, however, directly contradicts the idea that beliefs have natural boundaries which are determined by reality itself, independently of our cognitive activities. If “belief” were a natural-kind term, then this kind of conceptual tinkering would be unlicensed. What is characteristic of natural-kind terms is that they are very non-plastic. As Dorr and Hawthorne put it, “When a content corresponds to some ‘joint in nature’, there is little pressure to think of it as belonging to a large set of variants that could just as easily be expressed” (Dorr and Hawthorne 2014, p. 283). When a given term is a natural-kind term, there is a best candidate meaning for that term; “gold”, for example, has a uniquely best candidate meaning – a chemical element with atomic number 79.

The claim that there is no uniquely best candidate meaning for the concept of belief directly undercuts the assumptions of Lycan–Stich’s hypothesis that “belief” has natural kind semantics (although it does not necessarily lead us back to eliminativism—see next section). The semantic plasticity of the concept of “belief” shows that we, the competent users of the term, have no intuition of there being a unique best candidate meaning for this term. There are several distinct options which we might rationally consider, and there seems to be no recourse to the idea that there is one best option which is delivered by reality itself.

In the case of natural-kind terms, if the Putnam–Kripke model is to be believed, the extension is, as it were, provided by reality itself: what needs to be done in order to establish extension of a given natural-kind term is to epistemically engage with the substance/species that is denoted by the term. It is only by determining the “nature” of the given thing that we are able to establish the correct semantics of the term in question.

Such a procedure is notably absent from the debate about the way we should conceptualize the phenomenon of discrepancy: little credence is given to the idea that we should first establish ‘the essence of belief’ in order to decide which mental states we should include in the extension of the term “belief”, and which should be classified as not being beliefs. Instead, the debate on how to handle discrepancy and whether to, for example, include the category of aliefs in our classification of mental states focuses on pragmatic and normative considerations.

A similar observation was presented by Zimmerman (2018) in the context of the issue of whether subjects who have implicit biases should be classified as harbouring racist/sexist beliefs. According to Zimmerman, it would be a mistake to think that we could expect to solve this issue by just deferring to science and adopting the definition of belief provided by psychology. In such cases, as Zimmerman rightly claims, we must make a pragmatic and moral decision concerning whether to classify such subjects as harbouring these objectionable beliefs. This seems about right to me, and although Zimmerman does not take the hypothesis that “belief” has natural-kind term semantics into account, it seems that his way of looking at the matter would also conflict with this idea (on a side note, this does not mean that I wish to fully endorse Zimmerman’s first-level pragmatic definition of belief here).

The argument presented in this paper does not constitute a direct proof that “belief” does not have natural-kind term semantics: as was noted in Sect. 2, such an argument is notoriously hard to produce. The idea is rather to show that there is little reason to suppose that the concept of belief—at least as it functions in philosophical

psychology—derives its meaning from direct contact with a natural cognitive kind. It is important to note that one cannot just declare by theoretical fiat that a term possesses natural kind semantics; there must be some reason to suppose that a given term has meaning fixed by a causal chain to a natural kind, and there is little reason to think that this is the case with belief, as our conceptual practice seems to support the alternative hypothesis.

The line of argumentation presented above might be challenged in the following way: there might other terms which we would be ready to classify as having natural-kind term semantics that have apparently undergone changes in their meaning. For example, the concept of “atom”, which might be considered to have natural kind semantics, has been differently conceptualized throughout the ages. Thus, it would seem that the observation that the meaning of an expression might be modified is not enough to justify the claim that this expression does not have a natural kind semantics.

Two things might be said in response to this worry. First, that it is not that certain that expressions commonly used in the literature as examples of natural kind semantics really have such semantic properties. As has been noted, it is extremely hard to establish whether a given theoretical term has descriptive or natural kind semantics, and the mere assurance from a philosopher of language that “atom”, “water” or “gold” have this kind of semantic properties is not enough to prove that this is the case. It might be turn out that even in such cases a detailed analysis would show that these terms had meanings determined by theory.

Second, it might be said that even if we agree that certain terms which have natural kind semantics have undergone meaning changes, there is a significant difference in the way the natural-kind terms and the plastic terms change their meaning. If the natural kind semantic theory is to be believed, when people change their opinions about “gold” or “atoms”, their activity should be reconstructed as being guided by their aim of discovering the hidden essence of what they are inquiring into. There is no room for conventional redefinitions of such terms: the changes of opinions should be seen as hypotheses about the nature of the kind denoted.

On the other hand, highly plastic terms can be redefined without regard to any putative essences: when feminist philosophers propose reconceptualizing our existing gender categories, they do so without any intention of claiming that they have discovered the essence of, say, “womanhood”; rather, they propose a conceptual change based on normative and pragmatic considerations. There is no claim that such a change better tracks a hidden joint in nature; rather, the idea is that choosing a certain conceptual option might be useful in attaining some aim.

The argument provided in this paper aims to show that the reconceptualizations of the notion of “belief” that are proposed within philosophical psychology are akin to the way feminist theorists try to change the notion of gender. The proposed changes are motivated by pragmatic and normative considerations, not by putative discoveries about the essences of mental kinds. If this is true, then it might be argued that the concept of “belief” is plastic and does not have natural-kind term semantics. One instructive example of such purposive re-conceptualization of the notion of “belief” in philosophical psychology might be Van Leeuwen’s (2014, 2017) idea of distinguishing between factual beliefs and religious credences. It is important to note that

Van Leeuwen is perfectly aware of the fact that in order to justify this distinction he has to redefine the concept of belief: as he writes, “‘factual belief’ for me is a term of art that refers to mental states that have the features my theory describes” (Van Leeuwen 2017, p. 208). This methodological approach, namely that we first stipulate the proposed re-conceptualization of the notion of belief and then see if such a redefinition yields important theoretical results, strikes me as the correct one (obviously, for some readers Van Leeuwen’s views on the philosophy of religion would not be terribly convincing, but the point discussed here is methodological, not substantive).

8 Consequences for the Metaphysics of Belief

If the reasoning presented in the previous section is correct, then the idea of “belief” having natural-kind term semantics is not a silver-bullet response to the charge of eliminativism. Once we see that the claim that “belief” has natural-kind term semantics lacks support, then the Lycan–Stich style of argumentation loses much of its theoretical bite. There seems little reason to suppose that straightforward realism about beliefs can be easily rescued by pointing to the possibility that “belief” directly refers to some cognitive natural kind if there is little reason to suppose that the term in question actually possesses natural-kind term semantics. The mere supposition that “belief” (or indeed any other term) might (just might) have natural kind semantics cannot be seen as an ontological silver bullet if it is not backed by some positive reason to think that the term in question indeed functions as a name for a natural kind.

However, rejecting the idea that “belief” has natural-kind term semantics does not necessarily mean that we should go back to Churchland-style eliminative materialism. Disproving the idea that the meaning of “belief” is determined by direct reference to some cognitive kind does not force us back to the idea that beliefs are to be defined by their theoretical role in a (flawed) proto-theory of folk psychology.

A promising alternative to both natural-kind-style realism and eliminative materialism is the idea that we should retain beliefs in our ontology, even if we accept that they do not form a natural kind. In order to do so, we should claim that folk psychological kinds, such as beliefs, are in fact human kinds, i.e. kinds shaped by our culturally determined ways of looking at the world. This line of reasoning is proposed explicitly by Dewhurst, according to whom “the category ‘belief’, for instance, might not exist as a fundamental cognitive scientific kind (whatever that might mean), but could still refer to a non-arbitrary set of traits and behaviours, and could figure in folk psychological predictions and explanations (Dewhurst 2017, p. 183)”. A broadly similar pragmatic approach to the question of kinds in folk psychology, although without much emphasis on social construction, has been also promoted by Matthews (2013). The advantage of such a pragmatic, human kind treatment of “belief” is that it might be easily squared with the observation that the concept of belief is amenable to conceptual engineering. If when using the term “belief” we are talking about a human kind, then

it seems natural to suppose that the boundaries of this kind are something which might be changed by the process of our conceptual decisions.

This position would obviously require much more detailed elaboration, as there are many competing accounts of what human kinds are on a general level and the application of this general concept to the area of folk psychology might be not a straightforward matter. For the purposes of the present essay, however, it is important to note that there is a possible alternative to the Churchlandian account of folk psychological terms, once we reject the natural-kind term semantics in this case.

9 Conclusion

The main conclusion of this paper is that once we accept the claim that the concept of belief is plastic and amenable to conceptual engineering (as reflection on the phenomenon of discrepancy and the idea of “alief” shows), then there is little reason to suppose that this concept has natural kind semantics. This conclusion is important with regard to the current debates in metaphysics of belief: it shows that there is no easy way to rescue realism about propositional attitudes by way of claiming that “beliefs” denote some kind of cognitive natural kind, as there is no reason to suppose that the concept of belief has the appropriate semantics. However, the failure of the natural kind view does not mean that we should revert to eliminativism; the best alternative seems to be the view that belief is a human kind. This view makes it possible to accept both semantic intuitions about the plasticity of the concept of “belief” and moderate metaphysical realism about beliefs. Most importantly, this view does not carry the commitment to the idea that the reality of beliefs is to be vindicated by discoveries of appropriate kinds in cognitive science.

One question that arises here is to what extent does the conclusion generalize to other folk-psychological concepts like thoughts or imaginings? This is not an easy one to answer: on the one hand, it would be tempting to claim that all folk-psychological concepts are plastic as they belong to the same family of concepts as the concept of “belief”. On the other hand, the case for the plasticity of “belief” is based on specific considerations regarding the discrepancy cases and their conceptualization, and there seems to be no evident reasons why such considerations should be mechanically transferred to other cases. The issue of the proper semantics for other folk psychological concepts definitely requires more attention.

Acknowledgements The work on this paper was funded by National Science Centre, Poland, Grant under Award Number UMO-2014/15/D/HS1/02722. The first version of this paper was presented at the Polish Semiotic Society meeting in Warsaw. I am grateful to the participants for a heated and fruitful discussion; especially to Joanna Komorowska-Mach, Witold Hensel, Tadeusz Ciecierski, and Maciej Sendlak. I am also grateful to Szymon Bogacz, Pawel Banas, Bartosz Janik, Katarzyna Kijania-Placek, Zuzanna Krzykalska and anonymous referees at Erkenntnis. The initial research for this paper was conducted during my EPSA Fellowship at the The Centre for Philosophy of Natural and Social Science of the London School of Economics.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Albahari, M. (2014). Alief or belief? A contextual approach to belief ascription. *Philosophical Studies*, 167(3), 701–720.
- Bayne, T., & Hattiangadi, A. (2013). Belief and its bedfellows. In N. Notelmann (Ed.), *New essays on belief* (pp. 124–144). London: Palgrave Macmillan.
- Besson, C. (2012). Empty natural kind terms and dry-earth. *Erkenntnis*, 76(3), 403–425.
- Cappelen, H. (2018). *Fixing language: An essay on conceptual engineering*. Oxford: Oxford University Press.
- Chalmers, D. J. (2011). Verbal disputes. *Philosophical Review*, 120(4), 515–566.
- Churchland, P. M. (1981). Eliminative materialism and propositional attitudes. *The Journal of Philosophy*, 78(2), 67–90.
- Dewhurst, J. (2017). *From folk psychology to cognitive ontology*. Dissertation, University of Edinburgh.
- Dorr, C., & Hawthorne, J. (2014). Semantic plasticity and speech reports. *Philosophical Review*, 123(3), 281–338.
- Eklund, M. (2017). *Choosing normative concepts*. Oxford: Oxford University Press.
- Gendler, T. (2008a). Alief and belief. *The Journal of Philosophy*, 105(10), 634–663.
- Gendler, T. (2008b). Alief in action (and reaction). *Mind and Language*, 23, 552–585.
- Gendler, T. (2012). Between reason and reflex: Response to commentators. *Analysis*, 72(4), 799–811.
- Gertler, B. (2011). Self-knowledge and the transparency of belief. In A. Hatzimoysis (Ed.), *Self-knowledge* (pp. 125–145). Oxford: Oxford University Press.
- Haslanger, S. (2000). Gender and race: (What) are they? (what) do we want them to be? *Noûs*, 34(1), 31–55.
- Hensel, W. M. (2016). Why apply causal reference to intentional concepts? A polemic with Michael and MacLeod. *New Ideas in Psychology*, 40, 65–76.
- Jenson, J. C. (2015). The belief illusion. *The British Journal for the Philosophy of Science*, 67(4), 965–995.
- Kripke, S. (1980). *Naming and necessity*. Cambridge: Harvard University Press.
- LaPorte, J. (2000). Rigidity and kind. *Philosophical Studies*, 97(3), 293–316.
- Lewis, D. (1970). How to define theoretical terms. *The Journal of Philosophy*, 67(13), 427–446.
- Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50(3), 249–258.
- Lycan, W. (1988). *Judgement and justification*. Cambridge: Cambridge University Press.
- Machery, E. (2017). *Philosophy within its proper bounds*. Oxford: Oxford University Press.
- Mallon, R., Machery, E., Nichols, S., & Stich, S. (2009). Against arguments from reference. *Philosophy and Phenomenological Research*, 79(2), 332–356.
- Matthews, R. J. (2013). Belief and belief's penumbra. In N. Notelmann (Ed.), *New essays on belief* (pp. 100–123). London: Palgrave Macmillan.
- Muller, H., & Bashour, B. (2011). Why alief is not a legitimate psychological category. *Journal of Philosophical Research*, 36, 371–389.
- Perez, D. (2004). Mental concepts as natural kind concepts. *Canadian Journal of Philosophy*, 34(sup1), 201–225.
- Pinder, M. (2017). The explication defence of arguments from reference. *Erkenntnis*, 82(6), 1253–1276.
- Plunkett, D. (2015). Which concepts should we use? Metalinguistic negotiations and the methodology of philosophy. *Inquiry: An Interdisciplinary Journal of Philosophy*, 58(7–8), 828–874.

- Putnam, H. (1975). The meaning of ‘meaning’. *Minnesota Studies in the Philosophy of Science*, 7, 131–193.
- Quilty-Dunn, J., & Mandelbaum, E. (2018). Against dispositionalism: belief in cognitive science. *Philosophical Studies*, 175(9), 2353–2372.
- Rey, G. (2007). Meta-atheism: Religious avowal as self-deception. In L. Antony (Ed.), *Philosophers without gods: Meditations on atheism and the secular life* (pp. 243–265). Oxford: Oxford University Press.
- Schwitzgebel, E. (2010). Acting contrary to our professed beliefs or the gulf between occurrent judgment and dispositional belief. *Pacific Philosophical Quarterly*, 91(4), 531–553.
- Stich, S. P. (1996). *Deconstructing the mind*. Oxford: Oxford University Press.
- Thompson, J. (2016). Believe it or not: On multiplying classes of belief-like states. *Studia Philosophica Estonica*, 9(1), 79–110.
- Tumulty, M. (2014). Managing mismatch between belief and behavior. *Pacific Philosophical Quarterly*, 95, 261–292.
- Van Leeuwen, N. (2014). Religious credence is not factual belief. *Cognition*, 133(3), 698–715.
- Van Leeuwen, N. (2017). Two paradigms for religious representation: The physicist and the playground (a reply to Levy). *Cognition*, 164, 206–211.
- Zimmerman, A. Z. (2018). *Belief: A pragmatic picture*. Oxford: Oxford University Press.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.