

Mind Re-ascribed

Bruno Mölder

Department of Philosophy, University of Tartu

This paper is a reply to commentaries on *Mind Ascribed*. My response is organised into three parts. In the first part I describe the relationship between folk psychology and the scientific study of the mind. The second part replies to objections to the central tenets and presuppositions of the ascription theory. I clarify the distinction between the nature and the possession of mental states and the notion of a pleonastic entity. I explain why the ascription theory is a version of interpretivism, and not a species of instrumentalism or fictionalism. I also argue that canonical ascription should not be spelled out in terms of the ideal interpreter. The third part deals with comments on miscellaneous topics such as normativity, self-knowledge, the necessity of the brain and the proper understanding of intentional patterns.

Keywords: interpretivism, mental states, folk psychology, self-knowledge, natural kinds, normativity

Reading such sympathetic contributions has been an immensely rewarding experience. I am very grateful to Daniel Cohnitz for creating the opportunity for this book symposium and to the symposiasts for their insightful and thought-provoking comments. They have made me think hard about the foundational issues of my (somewhat unimaginatively named) ascription theory. In some cases, the objections or comments were due to misunderstandings invited by ambiguities and vagueness on my part, and those were relatively easy to rectify. In other cases, they reflected our disagreements over deeper philosophical questions. As it happens, each commentator concentrated on a different set of issues and there were relatively few recurrent concerns, if any. I can nonetheless group my replies into three parts. The first is concerned with broader questions about the relationship between the folk psychological and the scientific conceptions of the mind. The second is dedicated to the ascription theory itself, focussing on its central notions and presuppositions. The third part is more heterogeneous. It includes my

Corresponding author's address: Bruno Mölder, Department of Philosophy, Institute of Philosophy and Semiotics, University of Tartu, Ülikooli 18, 50090 Tartu, Estonia. Email: bruno.moelder@ut.ee.

responses on the remaining topics such as normativity and self-knowledge among others.

Despite the fact that quite some time has passed since the book's publication, I have not changed my fundamental outlook. However, hopefully I now have a somewhat clearer view of those issues that were still hazy for me at the time of writing the book. So in this reply I am not going to revise the basics of the ascription theory, but I offer rejoinders that restate, repair, reinforce, rehabilitate and hopefully reinstate the approach, thus making in effect the mind re-ascribed.

1. The ascribed mind and the sciences of the mind

Mind Ascribed propounded interpretivism concerning mental states as conceived through folk psychology, our common-sense framework by means of which we make sense of each other's actions. It is then only natural that one wonders how this common-sense conception of the mind is related to various models of the mind and the mind's neural basis as construed in various cognitive and neural sciences. Before we get to the ascription theory itself, I elucidate my stance on this relationship by replying to comments and criticism by Dominic Murphy and Sören Häggqvist. For the most part, I comment on the idea of natural kind-hood, examine if the danger of isolationism is engendered in such an understanding of the mind, and explain why I do not find eliminativism compelling after all.

1.1 Mental kinds and natural kinds

One way to spell out the close link between our common-sense conception of the mind and mental kinds is to say that mastering folk psychology amounts to grasping the nature of mental kinds. This means that their nature does not go further than what is contained in folk psychological concepts. We can find out about concepts by a posteriori means too (for example, by running empirical studies to find out how various communities use the terms), but there is no room for empirical discovery as there is in the natural sciences or in studies of the brain. Investigations of this latter sort may well unearth new data that shows that we need to discard our old conceptions completely. The *mind*, however, has not the hidden resources that give rise to such surprises. Or so I think. The same point can be expressed in other words as the claim that mental states are not natural kinds. In this regard, Daniel Dennett draws a vivid analogy between folk psychological mental concepts and the concept of dust:

Suppose Twin Earth is just like Earth except for having shmust where we have dust—behind the books on the bookcase, along country roads

during dry spells, etc. But surely, you protest, the concept of *dust* isn't the concept of a natural kind—shmust is dust, in spite of what anybody says! Exactly. It is a superficial concept, a nominal essence of scant interest or power. We already know enough about dust to know that science couldn't discover that dust was really something else—or that there wasn't any dust. Science could not uncover the secret nature of dust, because dust *qua dust* couldn't have a secret nature. In contrast, we already know enough about water, and gold, to know that they *are* natural kinds. ...are the concepts of folk psychology like the concept of dust, or like the concept of water? ...I think we already know enough about many of them to know that even though they may aspire to name natural kinds (unlike the concept of dust), they aren't good enough to succeed. (Dennett 1994, 535–536)

Just as there can be no discoveries about the real hidden nature of dust, there can be no scientific discoveries about the real hidden natures of mental kinds. Mental kinds are folk kinds: an outcome of the ways the folk carve the world, something that is not fixed by nature itself. This explains what is special about the mental-physical relationship, as compared with analogous relationships (e.g. biology-physics, geography-physics etc.) for kinds (e.g. biological kinds, geographical kinds etc.) besides the mental, which are studied in the special sciences. In order to locate mental entities in the natural world, an interpretivist solution becomes apt, especially if traditional reductionism is found inadequate and wholesale eliminativism is not a good option.

In the book I rejected the idea that mental states are natural kinds by drawing on Brian Ellis' (2001) conception of natural kinds. Arguably, and as alluded to by Häggqvist, given this conception, the criteria for natural kindhood are very difficult to meet. So perhaps on some more relaxed version, mental states would come out as natural kinds after all? Those more relaxed versions would require merely the groupings of similar features that are the result of common causal mechanisms and that ground success in explanation and inductive reasoning (e.g. Quine 1969; Boyd 1991). Of course, the verdict would also depend on the acceptability of these accounts of natural kinds themselves. Perhaps they are *too lenient*. But here is not the right place for a detailed study of different theories of natural kinds. I would only like to note that if folk psychological generalisations support counterfactuals, are inductively successful and are explanatory, then ultimately this success is due to the causally efficacious entities that undergird mentalistic explanations. Those could well be natural kinds, if they fulfil the right criteria, but this does not bestow natural kind status upon mental states. For the projectability of mental terms is not then due to their status as mental, but rather to the

corresponding subpersonal goings-on. Moreover, as I argued in the book¹ (230–231), the relation between the successful mentalistic explanation and the subpersonal causally efficacious basis is very loose. It could well happen that on the neural level, *several* kinds correspond to what the folk consider to be a unitary mental kind.

While we are on the topic of natural kinds, let me also clarify one minor issue about how to read a passage from Dennett (1994, 536) that I quoted in 144. Häggqvist (2017, 9) claims that I have overlooked a “natural reading” of Dennett’s quote on which he rejects the very notion of a natural kind. However, such a reading is not plausible if one looks at the wider context from which it was taken. There Dennett argues for “superficialism” about folk concepts, a view on which there are no deeper, scientifically discoverable facts of the matter on whether these concepts apply in a particular case or not. He does acknowledge that there are natural kinds (see also the quote above)—these being the cases when our terms happen to carve nature at its joints like “water” or “gold”—but warns against extending this idea to all cases (Dennett 1994, 535). His argument about nested natural kinds, which I quoted in the book, is a kind of *reductio* of the “hysterical realist’s” insistence that it *always* makes sense to ask for the exact reference of our natural kind terms. Sometimes this is a matter of negotiation in view of our interests, not of discovery of the facts. That the argument concerns the word “water”, shows that even in this case (which involves modification of Putnam’s standard Twin Earth story) there could be circumstances in which there can be no facts about which kind the term really denotes. But one should not conclude from this that water is not a natural kind or that the ‘natural kind’ idea itself is suspect.

1.2 Neither isolation nor elimination

If the nature of mental kinds does not lie deeper than what can be garnered from the folk conception, then what about all the sciences that also purport to study the mind? In particular, asks Dominic Murphy (2017a), could folk psychology be revised on empirical grounds? Or is folk psychology isolated from the scientific interventions?

I do not want to defend isolationism concerning folk psychology. I agree with Murphy that this position is unappealing, but detailing the interplay between folk psychology and the science requires some care. The commitments of the ascription theory do not preclude a mutual interplay between the folk psychological scheme and science. As noted, the influence can go in both directions.

¹ All references to *Mind Ascribed* (Mölder 2010) are by page numbers only.

On the one hand, the folk psychological concepts are often used in psychology and neuroscience to form hypotheses or to delineate the initial object for study. In the course of this, the original folk psychological notions become rather strictly specified as well as operationalised, resulting in a regimented vocabulary that can be more precisely linked to processes in the brain. This possibility was already outlined in the book and I do not see it as conflicting with interpretivism (70–72). I find myself in agreement with Francken and Slors (2014), who point out that both common-sense concepts and their scientific counterparts can be viewed as interpretations of behaviour. They argue that interpretation is required at all stages of such scientific theorising: in deciding whether those scientific concepts apply properly to the behavioural outcomes of experimental tasks as well as in deciding which brain areas are linked to the performance of those tasks.

What about influence in the other direction? Given my commitment that mental states are not natural kinds, can I allow the revision of folk psychology on scientific grounds? To a certain extent, this is a live option. Folk psychology need not be viewed as a static framework. It can be extended and modified over time, and so can our received description of it be revised. The ascription theory is not tied to any particular version of folk psychology. In (Mölder 2016) I explored the idea of an extended folk psychology that would be open to inputs from cognitive psychology and neuroscience, e.g., in the form of such terms as ‘affordance’, ‘resonance’, etc. In the case where the new terms are sufficiently strongly linked to the old folk framework they can be regarded as part of an extended folk psychology.

Francken and Slors (2018) call attention to some ways in which the folk psychological framework can be influenced by cognitive neuroscience, of which two are the most pertinent here. First, neuroscientific terminology can go to the masses, be adopted by the people and eventually transform their self-descriptions. As an example, they mention an addict who is able to give an enriched phenomenological description of one’s experiences thanks to knowing more about the underlying neural mechanisms. Second, the application criteria of some folk concepts can be changed on scientific grounds. Their example is responsibility: neuroscientific data can influence the assessments of responsibility in legal context. In addition, an important element in attributing responsibility is whether one has self-control, but this depends on various factors upon which neuroscience can throw new light. Francken and Slors’ first example can be viewed as a case of extending the folk framework, which is not just a matter of updating the folk vocabulary, but involves a change in people’s self-conception. The second case is an example of the influence of neuroscience on the application conditions of folk psychological terms. This is also in principle compatible with the ascription theory: in this

case, the evidential sources for the ascription include some neuroscientific data.

The cases in which the common-sense conception is enriched with scientific input are cases of the amelioration of our concepts. A conceptual change in the framework can be seen as a matter of the adjudication of vocabulary, not as a matter of making new discoveries about the hidden nature of mental states. To keep this consistent with the interpretivist outlook, one should not take this to imply that the terms in the mental and the scientific vocabulary are about the same natural kinds. Even if some neural terms do pick out natural kinds, mental terms pick out folk kinds. The enrichment and the extension of folk vocabulary involves thus the modification of old folk kinds or the creation of new folk kinds. Which scenario is the case—the modification of old or the creation of new—is ultimately a matter of policy, that is, how important we regard the continuity with the previous ways of speaking and how much stress we want to put on the differences between the old and the new (Godfrey-Smith 2004, 157; Dennett 1994, 534). But it would not be apt to assume that this depends on uncovering new facts about the previously hidden nature of mental kinds.

It is one thing to enlarge and improve the folk psychological vocabulary, but revising the existing folk notions is a different matter. There are limits to the range and the extent to which such revisions are possible. One example for such limits comes from Sehon (1997), who gives two examples of error, which are not intelligible for us. The first error scenario is the case in which one does not satisfy the folk conditions for being in a certain mental state (like desiring to watch a baseball game), but could still be attributed that mental state on the basis of one's brain states. The second error scenario is the case in which one satisfies the folk conditions for being in a certain mental state, but could not be attributed that mental state, since there is a certain covert property missing. The criteria for applying folk concepts cannot be overruled by factors that are not manifested in the ordinary practice. If the criteria used by neuroscientists are accessible only by special equipment and are phenomenally inaccessible to people, then it is hard to see how they could be linked to the existing folk psychological framework. But if that is the case, then it is also hard to see how this could lead to revising our attributions. The previously mentioned cases by Francken and Slors (2018) are different as there the neural factors are either phenomenally accessible or have the potential to enter into the folk conception via a change in our understanding of what counts as, say, self-control. Without links to folk psychology, scientific attempts to revise common-sense attributions would be perceived

by the folk as a change of topic: they are talking about something else, not beliefs or desires as we know them!²

By the same token, to paraphrase Dennett, there cannot be a scientific discovery that could show that the mind was really something else or that there are no minds. No matter what we could find out about the physical basis of mental processes, it cannot amount to a total rejection of the mind as conceived by our common-sense and as reflected in our everyday practice. One can get rid of mentalistic talk only by uprooting folk psychology itself, but this will not be a matter of discovery.

This should also explain my stance towards eliminativism. As I noted in 142, the dispute can only be over the meaning and prospects of conceptual elimination, since both eliminativism and interpretivism do not recognise mental entities in the ontologically strong sense. An eliminativist option in the conceptual case looks attractive if one treats folk psychology on the model of a scientific theory. Then one can find that it is not precise enough to yield testable hypotheses or that its categories are not useful for a mature science. This is an attitude that eliminativism shares with realism. Both view folk psychology as a source for scientific inquiry. They differ in that the realist expects that the scientific results will vindicate folk psychological classification, and the eliminativist denies this (cf. Sehon 1997, 342–343). In fact, I am sympathetic to the idea that if one's aim is to do good science, then one should be prepared to leave some folk categories behind. However, I think that folk psychology is more than just a source of concepts for the sciences, and it can stand on its own without requiring external vindication. It is our primary mode of making sense of ourselves and others and it is validated by daily practice. Displacing it would be a large-scale venture that involves profound change in our entire lives; it would not be just a matter of fixing one false theory (see also 146–147, Baker 1987, 130–133 or Murphy 2017b, 172). In my view, there are therefore not just two options—if vindication fails, then elimination is in order—and it is interpretivism that allows us to do justice to the centrality of folk psychology in our self-conception.

2. The ascription theory: clarifications and reinforcements

This part of my reply is devoted to clarifications and reinforcements to the ascription theory prompted by the objections, comments and inquiries by Sören Häggqvist, Marc Slors and Henry Jackman. I begin by clarifying the book's principal distinctions between pleonastic and natural entities and between the nature and the possession of mental states. I explain my assumptions concerning meaning and discuss whether there is a surplus meaning

² Cf. (Dennett 1994, 535) and (Sehon 1997, 337).

to mental state ascriptions that turns interpretivism into instrumentalism. I also take a stance toward pure ascriptivism as outlined by Marc Slors. Then I discuss the question of whether the notion of canonicity should be understood in terms of the ideal interpreter. Finally, I explain the role of rationality in relation to the ascription theory.

2.1 On the nature and the possession of mental states

Let me begin by recapitulating the general metaphysical framework presumed in *Mind Ascribed* (for more details, see section 2.1 of the book). One can talk about entities in the inflationary and in the deflationary sense. For example, properties in the inflationary sense are robust: they are the ways an object really is. Properties in the deflationary sense, however, include all true predications made of an object. Speaking in the deflationary sense, one can say that there is a distinct property corresponding to each predicate. This yields an abundance of properties, but this is harmless as long as we do not assume such a way of speaking to carry direct commitments to fundamental ontology. Deflationary properties are not metaphysically weighty. The trouble arises—and this is our own making—when one takes what are in effect only deflationary entities as inflationary entities. My contention is that this may have happened with mental properties, when it is assumed that they are properties in the same sense as physical properties. For given such an assumption one is bound to search for a substantial account of the relationship between inflationary mental and physical properties. However, if there are no inflationary mental properties, and some physical properties are properties also in the inflationary sense, then there is no need to look for such a substantial account.

I called entities in the deflationary sense ‘*pleonastic* entities’. The term comes from Schiffer (2003) who also outlines how one can move from using predicates to pleonastic properties. In short, whenever some predicate applies to an object, an object has a respective pleonastic property. Besides pleonastic properties and other entities, I assume that there are also some inflationary entities. Not all ontology can be easily read out from our language use. There is room and a need for a substantial ontology, which is couched in terms of inflationary or (as I called them) natural entities.

Häggqvist expresses some misgivings concerning the centrality of “Schifferian metaphysics” to the ascription theory, especially with regard to my criticism of alternative accounts of the mental state possession. In fact, my objections to alternative accounts are not and were not meant to be dependent on the distinction between pleonastic and natural entities. The purpose of my reformulations of various views on the mind-body relationship was to bring out more clearly than usual what their commitments are concerning

what exists. It turned out that in most cases, the alternative theories of the mental-physical relationship do not even require inflationary mental properties when properly reformulated. The only exception was role functionalism and there I presented separate objections to the inflation of role properties. But even these objections can be put in “neutral” terms. Apart from this, my objections to the theories stemmed from their own shortcomings, not from distinguishing inflationary and deflationary senses of ‘entity’.³

I do think, however, that it is useful for a proper presentation of interpretivism to have some conceptual resources that do not bring along commitments to inflationary mental entities. Without such resources the formulation of the interpretivist position would be very cumbersome or risks being misunderstood. One way to express such an insubstantial metaphysics is to talk in terms of pleonastic properties, but this is not the only way. See (23, fn 8) for other examples, of which it is instructive to repeat Davidson’s strategy of using the language common with his critics:

Having banished properties, states, beliefs, and so on, from my explanatory machinery, I shall now continue to allow myself to use words that superficially seem to refer to such things, on condition that I be understood to hold that such talk can be exchanged at boring length for talk that doesn’t use these words as referring to entities. (Davidson 2001, 299)

Another central distinction in *Mind Ascribed* is the distinction between the *nature* of an entity and the *possession conditions* of an entity.⁴ The nature of an entity is what determines what the entity is. The possession conditions of an entity specify what it takes to have the entity. Let us consider an entity such as a cat. To give an account of its nature is to give an account of

³ In addition, Häggqvist (2017, 11–12) is worried that the arguments against functional properties generalise also to non-mental functional properties thus leaving no room for functional and pleonastic entities posited by the special sciences. I think that for the entities invoked by the special sciences that are merely pleonastic, we can apply the same ascriptivist story, but those entities that are not merely pleonastic need a different approach. The seeds for this approach are already present in the book as I discussed a similar issue under the problem of the drainage of causal powers to microphysics (217–218, 229). The key is to construe non-mental macro-entities in such a way that they would have proprietary causal powers that do not drain to the lower layers. See (Kim 2003) for an extensive discussion of this issue.

⁴ A related distinction as applied to concepts (between their individuation conditions and possession conditions) is discussed in (Fodor 1998). He criticizes “pragmatist” views such as Peacocke’s (1992) that attempt to individuate concepts by their possession conditions and points out that traditionally it has been taken “for granted that the explanation of concept *possession* should be parasitic on the explanation of concept *individuation*” (Fodor 1998, 2). Note that I am not defending a particular direction of explanation. I am just arguing that these two kinds of conditions can be distinguished, and that one need not be parasitic on the other.

cathood: what makes something a cat rather than something else. To explain the possession conditions of a cat is to explain what it takes to have a cat. The nature and the possession condition can be connected with various entailments. One possesses a cat only if what one possesses is indeed a cat. Moreover, knowing the nature of cats helps to explain what it is to have a cat. However, it need not follow that the account of possession conditions is a trivial consequence of an account of the nature of an entity. Perhaps having a cat is such a simple matter, so that having a cat is just having the thing that cats are. But consider a more complex case such as having a house. The possession of an immovable property such as a house involves complex social and legal facts that go beyond the straightforward specification of what a house is.

But what if the entity under scrutiny is a pleonastic entity? Pleonastic entities are just shadows of concepts.⁵ Whereas the nature of a natural entity can go beyond our concept of that entity, there is nothing more to the nature of a pleonastic entity than what is contained in our concept of the entity. The nature of pleonastic entities is not hidden in the sense that uncovering it requires empirical investigation of the world. All one needs to investigate is the respective concepts. Häggqvist (2017, 10) asks why I maintain that the nature of mental entities “lies in something as intensional as (non-mental, public) concepts”. If this is a question about why I regard mental entities as pleonastic, then that is a question about the very foundations of interpretivism. What else could they be if they are not inflationary entities? I rejected the proposal that they are inflationary entities in the book. If this is a question about reducing the mental to the non-mental, then I was not attempting such a reduction. All I said is that to explain and to understand mental entities we need to study the way we use mental concepts. The reason for this is that mental states are more like pieces of text that need interpretation than they are like natural kinds such as water and gold.

Häggqvist is also puzzled about why I talk about ‘nature’ in the case of pleonastic entities at all. I think this is quite harmless, once it is made clear that their nature is not “deep”. There is also a need for such a way of speaking because the distinction between the nature and the possession conditions holds also for pleonastic entities. If we engage in semantic ascent, this translates into a distinction between the meaning of a term and its application conditions. A story about pleonastic entities can be told in terms of their nature (meaning) or possession conditions (application conditions). I am not committed to the claim that these stories must always be independent, but I am saying that there are cases where the nature and possession condi-

⁵ This metaphor is used by Schiffer (2003, 59) who got it from Armstrong (1989, 78).

tions can be given separate accounts such that, although linked, one is not parasitic on the other.

This leads us to Häggqvist's question about the connection between application conditions of mental predicates and the nature of mental states: why does the account of the nature of mental states not simply flow out from an account of what it is to have them (to apply the respective predicates)? This is definitely an option in theoretical space. For concepts, Peacocke (1992) has developed an account that individuates concepts by the conditions that a thinker has to meet in order to possess the concepts. Perhaps something similar can be developed for mental states as well. Presumably, the account of the nature of, say, belief would be along the following lines: belief is just whatever it is that one has iff one fulfils the possession condition for belief. In the terminology of the ascription theory, this would amount to: belief is whatever it is that one has iff it is canonically ascribable to one.

However, my view in the book was that an account of the nature of mental states is independent of an account of the possession of mental states and the ascription theory was explicitly presented as the latter. So if we are interested in what it takes to have mental states, it is not enough to give the meanings of mental terms. In turn, the meaning of mental terms (or the nature of pleonastic mental entities) is not to be given by explaining what it is to apply these terms. Let me rehearse the reasons why I do not think that it is a good idea in the case of mental states to explain their nature in terms of their possession conditions. To begin with, I do not think that the meanings of mental terms (or any other terms, for that matter) are very ample. They are more like dictionary entries than encyclopaedia entries, to use Ned Block's (1995, 151) suggestive metaphor. Perhaps there is no straightforward way to measure what goes into the meaning of a word and what is left out, but surely one constraint must be that those who use a word have some grasp of what it means. The grasp need not be perfect and the deference to experts in the case of some terms is possible, but for common vocabulary (and mental terms *are* common) the meanings must be accessible to their users. Now, if the meanings are lean dictionary entities, they do not incorporate much. In particular, they do not incorporate the clause that such and such a mental state is ascribable. One reason for this is the "leanness" factor, and another reason is that as far as I know, the point about interpretability does not seem to be among the things that people grasp when they grasp the meanings of mental terms. "Ascribability remains transparent when we concentrate on the nature of belief", as I said in the book (153). Knowing the meanings of mental terms does not entail knowledge of the ascription theory or any other form of interpretivism. The third reason is that the fact that mental states are interpretable does not individuate their nature (cf. Davies

1995, 295–296 for a similar objection to the intentional stance theory if construed as an account of concepts). Since every mental state is interpretable according to interpretivism, and if the possession conditions were to determine the nature of mental states, then this fact about interpretability should be part of the nature of every mental state kind (or the meaning of every mental state term). But this does not distinguish one mental kind from any other mental kind. It would be an idle part of the nature (and an idle part of the meaning) of mental states, and this should raise suspicions if it is there at all. In sum, if an interpretivist insists that interpretability enters the meaning of mental terms, she just makes the position weaker than it should be.

All these considerations could be taken as a case against interpretivism as such, but I think we need not go that far. There is room for interpretivism at the level of possession conditions, given the provision that it does not creep into an account of the nature of mental states. The latter we can glean from that part of ordinary language that is collected under the heading of ‘folk psychology’ (145–146). The meaning of mental state terms is presumed when one ascribes mental states, but it does not follow that an account of mental states cannot go beyond the account of their meanings.

However, one might wonder, why do similar considerations not also apply if we were to take possession conditions to involve interpretability; especially when I have said that, given semantic ascent, the possession conditions of mental states just are the application conditions of mental state terms. In particular, do people really grasp the fact about interpretability when they apply mental terms and are there differences in possession conditions for different mental states? My answer is “Yes” for both questions. The ascription theory aims to reflect the ordinary practice of using mental vocabulary. When we apply mental terms we do rely on the sort of evidence that I have termed “ascription sources” and we attempt to make the different pieces of information fit each other. So that much, I venture to suggest, is grasped by the folk psychologists that we are. Since to apply mental terms just is to ascribe mental states (with contents), the ascribability must be involved too: after all, when we apply the terms to one, we should be able to grasp that they are applicable to one. Of course, there is a difference between mere application of mental terms and the canonical ascription, just as there is a difference between a mere ascription and the canonical one. We can apply mental terms non-canonically. My story about canonical ascription is a non-trivial theoretical elaboration of a standard for possession. As such one should not expect it to be read off simply from what we know when we apply mental terms. As for the second question, I should note that the general possession-condition scheme in 171 is indeed the same for all mental states (with contents), but this is what one should expect from a general scheme.

However, when the condition is considered in a particular case, that is, when the details about the state, the content and the subject are filled in and complemented with the evidence for the ascription, then it becomes as specific as it can get. The possession conditions thus reflect the specifics pertaining to the distinct occasions of having the same type of mental states and to the having of distinct types of mental states, not to mention the differences brought along by the contents of mental states.

2.2 Instrumentalism, fictionalism, interpretivism

That I do not propound the ascription theory as an account of the meaning of mental terms helps me to evade some of the charges Marc Slors (2017, 19–22) has directed against intermediate interpretivism. As I construed the territory in the book, intermediate interpretivism is the middle position between *pure ascriptivism*—the view that the possession of mental states is completely a matter of interpretation—and *revelationism*—the view that the possession of mental states is independent from interpretation, but the latter can provide a good glimpse into what states one has. According to intermediate interpretivism (and I classified the ascription theory as falling into this group), having mental states is not a mere subjective projection. Although interpretability cannot be eliminated from an account of what it is to have mental states, the ascription of mental states is answerable to objective constraints.

Slors argues in the following way: intermediate interpretivism leaves the interpretivist position vulnerable to the “as if” misconstrual. The reason why this misconstrual becomes possible lies in the presumption, presumably made by intermediate interpretivism, that mental state terms have “surplus meaning”, i.e. meaning which goes beyond our ascription practices. On the “as if” construal, we pretend as if we have mental states, whereas in fact or “strictly speaking”, no one has mental states. To block the way for the “as if” construal, Slors suggests, the surplus meaning has to be rejected. There is no further meaning to mental state ascriptions beyond the one that is engendered in our practices of ascription.

When Slors writes about “the meaning of mental state ascriptions” I assume that what he has in mind are locutions such as ‘John believes that snow is white’ and the mental predicates we use to make ascriptions, but not ‘I ascribe to John the belief that snow is white’. Because the latter adds the meaning of ‘I ascribe’ to the ascription made to John, this would have an extra meaning compared to the first, but this is not the issue in the present context. As I argued above, the ascription theory makes claims about the possession of mental states, not about their nature. Therefore, it does not add any surplus meaning to sentences such as ‘John believes that snow is

white'. The meaning of this sentence comes entirely from the meanings of the words that compose it. The point that the belief that snow is white is canonically ascribable to John is a point about the proper applicability of 'believes that snow is white' to John, but all of this presumes the ordinary meanings of the words.

One can distinguish the question of primacy from the question of explanatory priority. The first is a question about which comes first into being—meaning or use? Could mental terms have a meaning before we apply them to each other, before we take the intentional stance? The second is a question about what presumes what in the order of explanation. Could we rely on the meanings of mental terms to explain how the ascription of mental states is possible or do meaning and ascription need to be explained simultaneously? I do not think that meaning has primacy over use, but it does not follow that it cannot be presumed for the project of explaining the ascriptive use of mental terms. When we are in the position to ascribe mental states we are already deeply immersed in folk psychological practice and this practice could not even get off the ground if with every ascription the meanings of mental terms were to be specified from scratch.

Thus in my view, and in contrast to Davidson's, the ascription of mental states presumes that words have common meanings, and that the ascription in itself does not add to their meanings or change them. In the book I merely gestured towards the form that such a theory of meaning could take, having in mind an account of the kind that explains meaning in terms of non-mental patterns of use that are able to engender a distinction between the correct and incorrect use of the terms. Now, Slors (2017, 21) supposes that the surplus meaning might come from the presumption that the words have standards for correct use. He thinks that these standards must be normative and as such are something extra to the actual use of the terms, which is factual, and therefore non-normative. I am inclined to be suspicious about the crude oppositions of normative and factual, but even if meaning is normative, the use of the terms must be imbued with normativity as well. In any case, the normativity of the meaning, if there is such a thing, is not a surplus, something extraneous to the actual use of mental terms. The standards of correctness are public. There is no room for hidden standards of correctness in language use that would allow us to employ the terms as if we were correct superficially but state that, strictly speaking, our usage fails to meet the standards.

It is not altogether clear to me why Slors thinks that it is the surplus meaning that leads to the "as if" construal. After all, one can say that a discourse is "as if" whenever we *knowingly* treat the things in the discourse as being in a way other than they actually are. In other words, we *pretend* that

the discourse is how it actually is not. This, however, is not a point about meaning, but about how we choose to treat certain things. So I leave the meaning aside and discuss the worry about “as if” in its own terms.

Slors presents one scenario to illustrate his point about an “as if” construal of interpretivism. Imagine that folk psychology has Cartesian commitments: in ascribing beliefs and desires, people assume that these are behaviour-causing internal states of the mind. He claims that if it turns out that there are no such internal states, then “our day-to-day talk about mental states is characterized by reference to non-existing entities”. However, given that interpretivism is not eliminativism and that it attempts to save the phenomena, the upshot must be that “we merely *pretend* there to be mental states” (Slors 2017, 22). He claims that this reading is dangerous as it turns interpretivism into “an indefensibly radical position” (Slors 2017, 19), and that the ascription theory has not ruled out such a construal of interpretivism.

What is this radical position? It is instrumentalism, the view that the lore of intentional ascription is a useful instrument for explaining and predicting behaviour, but in fact the ascriptions are not true and there are no such entities to which it seems to commit us. Slors refers to (McCulloch 1990) and (Ravenscroft 2005) for such a conception of interpretivism. Note that Ravenscroft labels the position ‘fictionalism’, but given the way he construes the positions, it does not differ from instrumentalism:

Fictionalism in the philosophy of mind is the doctrine that, whilst strictly speaking there are no mental states, it’s extremely useful to pretend that there are. (Fictionalism is also known as ‘instrumentalism’ since it views the attribution of mental states as having instrumental value—and nothing more.) (Ravenscroft 2005, 72)

I think that calling instrumentalism ‘fictionalism’ is a misnomer, and I discuss fictionalism, properly understood, below.

I do not think that the ascription theory is instrumentalist in this sense. To make this clear let us consider a construal of realism by Crispin Wright and the ways it is possible to depart from realist elements. Wright (1995, 198) characterises realism about a discourse as consisting of the following three components:

- (1) *Representation*: the statements of the discourse have representational content.
- (2) *Aiming at truth*: the participants of the discourse aim at successful representation.
- (3) *Facts*: the world cooperates, that is, there are entities as represented by the statements of the discourse.

Wright notes that instrumentalists and expressivists typically deny the first, representationalist component, fictionalists reject the second component, and error-theorists deny the third, viz. that there are facts the discourse represents there to be. The ascription theory fits none of the moulds. The ascription statements are about mental states and the people who use them aim to talk truly about the mind. Although folk psychology can be put to a variety of uses, including regulative and expressive use, it does not follow that it has no descriptive use. In my view, the explanatory and predictive use of mental state ascriptions is central to folk psychology and folk psychology can be regarded as aiming at the successful representation of mental states. In a certain sense, the ascription theory offers an account of what this successful representation comes to in the mental case. The canonical ascription could be viewed as providing an objective standard for true or correct ascriptions of mental states. Thus, given that an ascription conforms to the standard of canonicity, one really has the ascribed mental states. There are corresponding pleonastic facts or the subject has corresponding pleonastic properties, so we could also say that the world cooperates. But the world does not need to contain mental properties in the inflationary sense.

Of course, I can say all this since I presumed minimalism about truth in the book, and a deflationary metaphysics for the mind: realism does not come with serious metaphysical commitments. This was also one of the reasons I preferred not to phrase my claims in terms of realism and truth in the book: given the heavy baggage these terms bring along, the potential for misreading is high. At this juncture, one could protest that truth minimalism and pleonastic ontology are too shallow to secure the realism of a robust kind.⁶ But note that many other positions, generally regarded as realist, also re-conceive the entities that the discourse purports to represent. Identity theory claims that mental states are actually brain states, and behaviourism claims that there are only dispositions to behave and behaviour. So on these views also there is some slack between what entities the world contains and what the discourse represents there to be. If we set the bar of realism so high, only dualism would count as a properly realist position since it acknowledges the existence of inherently mental robust entities. But if (say) the identity theory does not count as an error theory of folk psychology, then neither is the ascription theory an error theory. Ordinary mental state ascriptions are not erroneous, when they are canonical.

What about the previously mentioned case: what if the folk are Cartesian? I do think that the commitments of our folk psychology fall short of

⁶ Crispin Wright (1995, 207, note 15) later adds the following qualification to his threefold definition of realism: "it is *only* in a platitudinous, metaphysically non-committal sense that our anti-realist may countenance psychological states of affairs."

including a commitment to there being internal causally efficacious states of souls. Nonetheless, let us assume for the sake of argument there is a folk psychology that does include such a commitment. Let us assume also that the community of such folk is Cartesian even in a stronger sense than Slors envisions: they talk about soul states explicitly and there is a lore about souls that influences some of their actions. And this is characteristic to this community as a whole, not just some of its members. For if it were the case that just some people in the community have eccentric ideas about what mental states are, then this would not have much effect on interpretation in general. Interpretation proceeds from the communal meaning of mental terms, which is resilient to idiosyncratic deviations. Any individual differences would crop up in attributable speaker-meaning.

There are two different ways to interpret such a community. We could interpret them on the basis of our folk psychology or we could try to make sense of them in their own terms.

In the first scenario, we proceed from our own folk psychology, which presumably does not contain reference to Cartesian souls. From our point of view it then seems that the people in the Cartesian community are wrong: there are no soul states. But that need not turn all their attributions false. Some of them can be regarded as true thanks to the fact that they fulfil the canonicity requirements: they best fit with the available objective information and do not require revision. Primarily, this applies to those cases where there are overlaps between the conceptual schemes of the interpreter and the interpreted people: the ordinary-looking ascriptions of beliefs and desires and perhaps also of some other states, which are not explicitly couched as soul states. By 'those other states,' I mean states like emotions, thoughts, etc. that are designated *prima facie* by ordinary mental state terms. Although not explicitly labelled as soul states, they are still tacitly taken by the people in the community to be states of the soul.

Our story includes also cases in which it is crucial that what are ascribed are explicitly soul states. Such ascriptions occur, for example, when people make explicit remarks about mental mechanisms as they conceive them: 'The energy fluctuations in the rational part of the soul made me choose this harmful path of action.' Either the contents of attributed beliefs in such cases mention souls and soul states or soul states are among the states that they attribute to each other. Now if we let the assumption that there are no souls "contaminate" how we regard the ascriptions of Cartesian people, then we are bound to say that the ascriptions of soul states cannot be canonical. For this very assumption that there are no souls prompts the revision of any such ascription if it were actually made. At best we could convert all ascriptions of soul states into beliefs. Thus instead of the ascription 'Her soul hurts,' we

ascribe ‘She believes that her soul hurts’. Such ascriptions of beliefs, the contents of which include explicit reference to the soul and its parts, could still in principle pass the canonicity test. We could regard them simply as ascriptions of false beliefs, and false and irrational beliefs (from the interpreter’s point of view) are canonically ascribable (see 164–165).⁷

The first scenario thus results in a number of non-canonical ascriptions or canonical ascriptions of false beliefs if we reconstruct what the subjects say in terms of what they believe. However, I think that this way of approaching an alien community has some disadvantages. First, although it can throw light on how we could understand them from our vantage point it does not uncover how they make sense of themselves in their own terms. We might miss something crucial about their way of life in the first scenario (a point Peter Winch (1964) made long ago). Second, the assumption that there is an overlap between our and the Cartesian folk psychologies might be overly optimistic. Why think that we even mean the same thing when both talk about beliefs and desires? If the existence of souls is indeed the centrepiece of Cartesian folk psychology and the soul states are strongly linked to other mental states and observable action, then this would presumably have an effect on their use of mental state terms and ultimately on what they mean by ‘beliefs’, ‘desires’ or ‘mental states’. If this is true, we would be dealing with a different folk psychology, and its similarities to ours would be merely superficial.

Let us then look at the second way of approaching the Cartesian culture. This scenario holds that the people of the Cartesian community understand themselves via a folk psychology that is distinct from the one that we are accustomed to in our daily activities. Let us also make the admittedly un-Davidsonian assumption that the Cartesian folk psychology is still accessible to us or whoever are the interpreters of the Cartesian community. In this scenario, the ascription of soul states and other mental phenomena is based on a folk psychology that differs from ours, but there is nothing in principle that stops such ascriptions from coming out as canonical. The Cartesian folk psychology does not yield defeating assumptions that would require the ascriptions to be revised (as happened in the first scenario). Given their folk psychology, there can be no difference in the status of soul states as compared to that of beliefs. If the predicates that they use to attribute soul states have

⁷ Here I sense some parallels with Dennett’s heterophenomenology, which accepts as data what the subjects say with “a deliberate bracketing of the issue of whether what they are saying is literally true, metaphorically true, true under-an-imposed interpretation, or systematically-false-in-a-way-we-must-explain” (Dennett 2007, 252). But based on this initial data, the theorist could adopt an eliminativist position towards some things the subjects have said.

proper application conditions, one could even get pleonastic properties from these predicates by Schifferian something-from-nothing transformations.

The point is general. This in effect is what it is for the possession of mental states to be interpretation-dependent. When we change the folk psychology upon which interpretation relies, we also change the outcome. If an interpreter with a radically different folk psychology were to assess our ascriptions on the basis of her folk psychology, she would probably dismiss them as non-canonical. The interpretation is relative to the underlying folk psychology. But one need not succumb to a total relativism here. Even if there is no privileged viewpoint from which one could tell which mental states one “in fact” has, we should prefer the folk psychology that enables us to make more canonical ascriptions. One could then still say that the interpreter who dismisses the majority of our ascriptions as non-canonical has not configured the lens of the intentional stance properly. She is like Dennett’s Martian who misses a real pattern in the data (Dennett 1987, 25–27). To make the pattern in the data visible, one has to use the proper scheme and the proper scheme is the one that maximises canonical ascriptions.

Note that neither of these scenarios involves pretence. The ascription theory is not committed to the claim that ordinary people or Cartesian folk psychologists merely pretend that they have the states that are ascribable to them. They purport to make claims about how the mind is (according to their conception of it). And as I said earlier, these claims can be given a representationalist reading. The ascriptions are more than merely useful instruments. If the interpreter shares the folk psychology with the people, there can be no inclination to consider the mentalistic discourse as a mere make-belief, for then our own mental states would also become a matter of pretence. If the interpreter does not share the folk psychology with the people who make attributions, but proceeds from the alien scheme nevertheless, then the ascriptions can come out as canonical (given the scheme). If the interpreter does not share the folk psychology with the people and proceeds from her own scheme, the inclination is rather to reject the misguided attributions than to pretend that they are true.

Finally, I would like to distinguish the ascription theory from fictionalism. As properly understood, fictionalism about the mental is the view that the statements in the mental discourse are representational, but do not aim at truth. They are just like statements in a fictional story, which are truth-apt and have meaning, but do not purport to give a true description of the world. Instead, they have some other, non-epistemic benefits. For instance, according to Demeter (2013, 500), who presents a version of mental fictionalism, the value of folk psychology can be seen in its impact on “social navigation, the formation of alliances, and our psychological sensibility.” In

contrast, I think that even if there are all these further things that we can do with mental state ascriptions, they actually presume the aiming at successful representation. For how could we navigate successfully in the social world if we do not believe that our ascriptions aim to represent our minds and those of our partners? The ascription theory preserves this aim. It says that we really have those mental states that are canonically ascribable to us. It is not that we have mental states only according to some mentalistic fiction. Albeit pleonastic, there truly are mental facts. The mentalistic talk is indeed a text that needs interpretation, but it is not a fictional story. It is not a chapter from the book of the world, for it does not carve nature at its joints. It is a description of our conception of the mind.

2.3 Pure ascriptivism and counterfactuals

Slors (2017, 17) advocates pure ascriptivism, which he takes to be a “more stringent form of interpretivism”, for it leaves no room for the “surplus meaning” of the ‘mental’. I have already dealt with the putative “surplus meaning” and with the “as if” reading of interpretivism. In this section, I focus on pure ascriptivism as Slors presents it.

As argued in section 2.2, the ascription theory does not rely on any surplus meaning over and above our ascription practices, and neither is the “as if” reading apt for it. If that is true, then this alone takes the force out of Slors’ argument that pure ascriptivism is the best form of interpretivism. But let us consider the view on its own merits.

Slors claims that all mental state ascriptions are linked with counterfactuals about what one would do in various situations:

Pure ascriptivism, then, amounts to the idea that the meaning of mental state ascriptions hinges on the counterfactuals associated with these ascriptions and the practices of mental state ascription determine the full range of counterfactuals associated with a given mental state term. (Slors 2017, 24)

Although it is not fully clear what kind of dependence-relationship is intended by “hinging on”, we can see that this view postulates a tight fit indeed between the ascription practice and the meaning of the mentalistic ascriptions. But note that Slors presents pure ascriptivism as a claim about what the mental state ascriptions mean. In the first place, it is a view about the meaning, not an account of what it takes to have mental states. However, in *Mind Ascribed*, I was concerned with the latter account only.

As I already noted, in principle I can agree with Slors that the mental state ascriptions do not have any further meaning over and above our folk psychological practices. But let me voice a suspicion that counterfactuals as such are not a sufficient basis for capturing everything that is contained in

this practice. Must all implications of the prospective and retrospective mental state ascription take the counterfactual form? And counterfactuals about the behaviour even more? For one, I am not convinced that all of the various commitments and entitlements incurred by the ascription of propositional attitudes can be recast as counterfactuals or if they can, that it is illuminating to do so.

Be that as it may, as an account of the meaning of mentalistic ascriptions, Slors' pure ascriptivism seems to be prone to a damaging kind of holism. The counterfactuals linked with each token ascription from the same type could be different, for the predictions about what one would do depend upon one's individual features. But then, on this account, there is no single fixed meaning to 'John believes that it is raining'. One could argue that this is fine when we apply this sentence to different Johns, for as different persons, they might all behave a bit differently. But I think that the same worry holds when we apply this sentence to the same John at different times. Depending on John's other current mental states and dispositions, this ascription could entail different predictions, and hence the sentence must mean something different at each occasion of its use. This problem with holism does not come up in the ascription theory, as the ascription theory concerns the possession of mental states, not the meanings of mental state terms or ascription sentences. Although every ascription is individual—they involve different source information and can lead to different predictions—the meanings of the sentences used for making the ascriptions are stable and do not vary from case to case.

Slors' pure ascriptivism—a view that brings counterfactuals to the centre of things—seems similar to Lynne Rudder Baker's practical realism (Baker 1995; 2003). On this view, believing something "is a property that someone has in virtue of there being a relevant set of true counterfactuals about the person" (Baker 2003, 185). For Baker those counterfactuals also include conditionals with actually true antecedents. According to practical realism then, what beliefs a person has is fixed by what that person does, says and thinks as well as what the person would and would not do, say and think in different situations (Baker 1995, 154–155; 2003, 185–186). However, there is an important difference between pure ascriptivism and practical realism. The latter is not intended as a reductive account of the meaning of believing and other mental terms. Baker (2003, 185) presents it as a non-reductive account that lists the necessary and sufficient conditions for the possession of mental properties.

One reason why Baker does not analyse believing that p in terms of sets of counterfactuals is that ascriptions of the same belief on different occasions can involve quite different counterfactuals. Hence believing that p cannot be reduced to a single set of counterfactuals (Baker 1995, 156). It is basically the

same worry as the problem with holism mentioned earlier and she avoids this worry by not giving the counterfactuals a meaning-determining or reductive role in her account.⁸

Is pure ascriptivism, as Slors construes it, a species of interpretivism at all? I cannot recognise it as the position I described as “pure ascriptivism” in the book (81). On that view, it is purely a matter of interpretation what mental states one has. Interpretability is both a necessary and a sufficient condition for the possession of mental states according to pure ascriptivism. I regarded it as an extreme position because it leaves no room for objective constraints on interpretation: if it is necessary and sufficient for having a belief that *p* that one is interpreted accordingly, then it seems that it is only the very fact of being interpreted that makes it the case that one has the belief that *p*. This worry does not go away if one talks about interpretability instead of actual interpretation, for the potential to be interpreted in itself does not suffice to turn interpretation into an objective matter. My guess is that Slors calls his version of ascriptivism “pure” because he wants to keep the point that something (interpretability in my version or the ascription practices in Slors’ case) exhausts something else (the possession of mental states, or in his case, the meaning of mentalistic ascriptions). However, being an account of the meaning, it is distinct from pure ascriptivism as I construed it.

Furthermore, interpretability is not constitutive in Slors’ version at all. For instance, he rejects my suggestion that Bakerian counterfactuals could be regarded as covertly interpretive and thus that the counterfactual account could be reconciled with interpretivism.⁹ For Slors, the holding of the relevant counterfactuals is an interpreter-independent matter. The interpretability comes into play insofar as “highlighting those counterfactuals serves practical purposes, such as predicting further behaviour” (Slors 2017, 25). Here, the interpretability plays a useful, but an ultimately dispensable role. The main job is done by the sets of counterfactuals. If one wishes to find a place for this view in the interpretivist spectrum, and I do not think that this is

⁸ In my view Christopher S. Hill’s objection to Baker is misplaced. Hill (1997, 138–139) argues that it follows from this holism that belief states lack a uniform nature. Given that the relevant sets include different counterfactuals, two tokens of the same belief could have different natures and it becomes unclear what makes them tokens of the same type. However, Baker does not define the *nature* of belief via sets of counterfactuals. In her view, the set of true counterfactuals are necessary and sufficient conditions for *having* beliefs. But—as noted above—a similar problem plagues Slors’ pure ascriptivism, which fixes the meanings of mentalistic ascriptions by the sets of counterfactuals.

⁹ The suggestion was that a full explication of what it is for someone to have a mental state (e.g. for *S* to believe that *p*) even if couched in terms of counterfactuals involves interpretive considerations, although ‘*S* believes that *p*’ can be true independently from any actual interpretation (154–157).

obligatory, then at most, it would lie at the other extreme from pure ascriptionism (in my sense). It would be a version of revelationism.

2.4 Canonical ascriptions and ideal ascriptions

Henry Jackman concentrates on canonical ascriptions, that is, ascriptions that fix the possession conditions for mental states. Although he acknowledges in the end that my “relaxed” notion of canonicity is superior to the conception of canonicity in terms of the ideal interpreter, he is not happy with my reasons for rejecting the ideal conception. In this section I clarify my position with respect to the criterion of canonicity and give my answer to Jackman’s critical remarks.

If an interpretivist explicates the possession of mental states through their ascription, a certain standard is required for distinguishing between such ascriptions that count for fixing the mental states and ascriptions that are not comprehensive enough for the task. To recap, in the ascription theory, this role is played by the canonical ascriptions, which approximate the maximum coherence of the information from various sources of evidence for the ascription and are such that they do not require revision (see 175 for more exact definition).

Now, Jackman asks why I do not fasten the standard of canonicity to the ideal ascription. On this view, an ideal ascription would be “*maximally* coherent with complete knowledge of all of the ascription sources” (Jackman 2017, 29). In my view, which I also outlined in the book, this notion of canonicity would make the standard inaccessible to us, the ordinary ascribers of mental states.¹⁰ We are not ideal ascribers, since we are not omniscient. We are not in possession of a complete knowledge of all the sources of evidence for the ascription. We cannot compute the perfect coherence, we can only approximate to maximum coherence. If the ideal ascription were canonical, our limitations would make our ascriptions substandard, not good enough for fixing the correct interpretation and consequently, the possession condition.

Jackman points out that in the case where our ascriptions agree with the ideal ones, we could make canonical ascriptions even if we are not ideal ascribers ourselves. He adds that “in the absence of any concrete reason why we should expect such a divergence from the ideal interpreter, it seems

¹⁰ We are discussing here the typical ascriptions that involve assessing the coherence of several sources of evidence. I am thus not considering the admittedly rare cases where one could be an “ideal interpreter” on the cheap, cases where there is just a single relevant piece of evidence for the ascription or very few evidence, so that computing the coherence is undemanding. Such cases, even if they occur, cannot set the standard for the rest.

doubtful that we should need anything like an *a priori* guarantee that we will match such ascriptions.” (Jackman 2017, 33)

Even if our ordinary ascriptions sometimes happen to match those that the ideal interpreter would make, this would be nothing more than a lucky coincidence. There does not seem to be any principled ground as to why our ascriptions must match with the ideal ones. It would not be mere luck if the omniscience and absolute computational resources were in principle within our reach, but they are not. The standard of canonicity is the standard for mental state possession according to interpretivism. What mental states one has should be in principle accessible to ordinary interpreters, otherwise the very sense of attributing minds and mental states is lost. Linking canonical ascriptions to ideal ascriptions would make the standard accessible for us merely by a lucky accident.

Jackman claims that there is no reason to expect a difference between our ascriptions and those of the ideal interpreter, but it is our imperfection that provides such a reason. Take an ordinary ascription of a mental state. An ideal ascription can differ from it in at least two dimensions: first, it can take into account a wider range of evidence, and second, even if the considered information is the same, it can measure the coherence in a better way. It is thus very likely that the ideal ascription will come out as somewhat different. My conjecture in the book was that our ordinary ascriptions are usually good enough so that they could count as canonical (and my relaxation clause allows this). But if the ideal ascription is canonical by definition, then even in cases where the ordinary ascriptions do not differ much from the ideal ones, they cannot count as canonical. Only in those lucky and presumably rare occasions where they match the ideal ascriptions, they would have canonical status. But of course, we would not really know if we were lucky or not, and given our imperfections, there is a *prima facie* reason to expect that any particular ascription is of the unlucky sort. This is a more serious condition than a common fallibilism.

But how could ascriptions that fall short of the ideal ever be deemed canonical? Jackman (2017, 33–34) points out that even on my relaxed conception, a different ascription made by an ideal interpreter would create the need to revise the original ascription. However, this overlooks one crucial difference between the ascriptions we could make and the ascriptions of the ideal interpreter. All canonical ascriptions according to my theory are such that they could be actual even if they happen not to be made actually. The second clause of the definition of the canonical ascription states that if the ascription were made in fact, it would not require revision (175). This presumes that it can in fact be made: there is nothing that in principle precludes such an ascription. The same cannot be said about the ascriptions

of the ideal interpreter. Ideal interpreters do not dwell in the actual world and in the worlds relevantly similar to the actual world (called ‘ w_c worlds’ in the book). When one talks about ascriptions rather than interpreters, then a similar point applies: in typical cases, in the actual world and in the w_c worlds, no ascriptions are made that are based on the full knowledge of all the ascription sources and on a complete computation of maximal coherence. If so, then there can be *no* actual (and w_c -world) scenario in which an ordinary ascriber would be required to revise one’s ascription in light of a different ideal ascription. Note that the possibility that some ordinary ascriptions made in a non-ideal way match in content the ascriptions of the ideal interpreter is not relevant to the present point. For that would be a case of congruence, not a case of revision-requiring divergence.

Eventually, however, Jackman comes to reject the proposal of defining canonicity in terms of the ideal interpreter. This has to do with the uncodifiability of interpretation: there is no full codification of the process of interpretation—not even of the process of finding out which interpretation fits best with the evidence. Jackman (2017, 36) admits that “there may be no determinate answer to the question of what set of ascriptions would be maximally coherent with the ascription sources, and thus no determinate answer to the question of what mental states an ideal ascriber would ascribe to us.” Indeed, if ideal ascriptions were to set the canonicity standard, then given the aforementioned uncodifiability, the upshot is that there would be no determinate canonical ascriptions and therefore, given the link between canonicity and possession, nothing would determine what mental states we have. Such an upshot presents a good reason to look for an alternative account of canonicity.

I believe that my more relaxed conception fits the bill nicely. There is an explicit discussion of uncodifiability in *Mind Ascribed* (179–181) as well as an awareness of the danger that it could lead to the disappearance of the mind:

The stipulation about revision gives canonical status to a variety of ascriptions which pass the relaxed criteria. Hence, there is no need to worry that uncodifiability *cum* interpretivism leads to the conjecture that, since there is no codifiable way to fix our mental profiles and since our mental profiles depend on such fixing, they are either completely undetermined or we do not have mental states at all. (181)

Nevertheless, it did not occur to me that one could make use of the uncodifiability considerations as an argument against ideal interpretation. I am thus grateful to Jackman for pointing this out. However, I do not think that this argument makes my reasons against ideal interpretation as canonical ascription superfluous. The argument from uncodifiability can serve as an-

other nail in the coffin of the ideal conception of canonicity, but it does not have to be the only nail!

I developed two points: the ascriptions made in the ideal manner, even if possible, would still be beyond our reach, but *qua* ideal they would trump our ascriptions; there are no ideal ascriptions or ideal interpreter, at least not in the actual world or its vicinity. Jackman's point is a stronger one: even an ideal interpreter might not come up with a single maximally coherent ascription. They all render the notion of an ideal interpreter or an ideal interpretation useless for the purpose of defining canonicity.

Compare this with possible considerations about whether we should take into account what God does. The analogue for the first point would be that we could not enact God's perfect deeds: by definition they would be better than ours and therefore would constitute an unreachable standard for us. The analogue for the second point would be to note that there is no God anyway (at least not around here). Jackman's point would be then that there may be no determinate answer to the question of what God does. They all render God's actions useless for our deliberations.

2.5 Why isn't it interpretivism?

In this section I answer the questions Häggqvist (2017, 13–15) posed about central features of the ascription theory. They pertain to the evidence for an ascription and the role of rationality in interpretation.

Let us discuss the issue of evidence first. Häggqvist notes that I did not touch upon the question of how much evidence is sufficient for an ascription. This leads to the worry that if very little evidence is acceptable, then it becomes too easy to find an ascription that coheres with all the evidence. In reply I would note that the question of the admissible volume of evidence can only crop up in the case of an *actual* ascription (and I did discuss singularly compelling ascriptions in connection with the actual ascription of perceptual states, cf. 236–237). There we can ask how much the interpreter should know in order to make an acceptable ascription. However, what matters for the ascription theory is the canonical *ascribability* and this is a hypothetical situation in which all relevant information is available. In view of this,—and this is a point that Häggqvist also noted—one can expect that the actual ascriptions made on the basis of scant evidence are not up to the standard of canonicity.

Häggqvist also asks if other ascribable mental states can be viewed as evidence source equal to the other kinds of evidence such as behaviour, stimuli and personal background. As these are also ascribable, should they not be the product of interpretation, rather than its source? My answer is that their status as evidence is relative to each ascription. The fact that these other

mental states are ascribable in a given situation constrains the ascription of a particular mental state, since mental states are interlinked in a holistic network. Of course, evidence is needed for the ascription of these other mental states too, but that evidence should in turn include information about yet further mental states. I do not think that the interpretation could proceed solely on the basis of a subject's behaviour and environment, without taking into account its mental profile.

The holism of mental states brings us to the role of rationality in interpretation. I did indeed reject the common notion that interpretation must presume the rationality of its objects. However, I do not see this as so big a break with tradition to warrant the suspicion that thereby the ascription theory ceases to be a form of interpretivism—*contra* what Häggqvist suggests in his title and in the last passage of his contribution. My point was simply that the stress on rationality generated lots of unnecessary disputes that overshadowed what is really crucial for interpretation. And this is that interpretation requires only links between mental states, behaviour, environment and background, such that it would be possible to move from the sources to the ascription. If those links are rational and can be subsumed under some theory of rationality, it is all well and good, but they do not have to be rational or conform to some generally accepted norms. One could deviate from rational norms in all sorts of ways and one's mental states could be related in ways we normally deem to be irrational. However, this does not make the interpretation of such a subject impossible, since we can become aware of such deviations and irrational relations. It just makes the interpretation more difficult in actual cases.

Similarly, the internal consistency of a subject's beliefs is not necessary for an ascription, but it does not follow that we cannot sometimes assume such consistency. When the interpretation should aim to maximise the coherence of an ascription with the sources of evidence and that evidence also includes other ascribable mental states, then this would also maximise the consistency of a subject's beliefs. But this does not preclude the possibility that in some cases the best interpretation is the one in which a subject's beliefs are not all internally consistent or rationally linked. I do not see the ascription of inconsistent bundles of beliefs as presenting a special kind of worry. This can be done on the basis of information about a subject's idiosyncrasies, which could range from severe delusions to merely misaligned sets of attitudes. Some of this evidence can come from what the subject talks about and the way he or she talks. The so-called verbal behaviour can constitute an important piece of evidence for the interpretation. To be sure, in some cases one might mean something uncommon by one's words, but this

would merely be a matter of ascribing a different speaker meaning, and this can be accommodated by the ascription theory (cf. 125).

What to make of Dennett's rabbit-bird example that Häggqvist (2017, 14–15) presents as a case showing the need for internal consistency? As it stands, it is just an excerpt, which cannot be made intelligible without further information. I thus agree with Dennett and Häggqvist that a fuller story is needed for its interpretation. I suggest that this story would bring out those presumably idiosyncratic connections between the man's beliefs or his deviant meanings. Perhaps this makes him internally fully consistent. Perhaps not. In any case, given such further information we would also be able to interpret him within the framework of the ascription theory. I thus do not see a reason to worry that my view is not a species of interpretivism.

3. Normativity, patterns and self-knowledge

For the last part of this paper, I have collected those comments on topics that do not fit under the headings of the first two parts. This part is thus inevitably more diverse—ranging from normativity to self-knowledge. I also comment on the role of the brain in having mental states and on the correct understanding of Dennett's notion of real patterns. The contributors whose comments I address in this section are Dave Beisecker, Sören Häggqvist and Dominic Murphy.

3.1 The curious case of normativity

In this section I address two significant and related omissions in the book highlighted by Dave Beisecker: Sellars and normativity. I comment first on Sellars rather briefly, but the bulk of this section is dedicated to normativity, a tangled topic of its own.

Beisecker (2017, 47–48) draws attention to the lack of engagement with Wilfrid Sellars' ideas in my book. In this respect I can echo Dennett (1987, 349), who quipped, when talking about himself, Davidson and Bennett: “Almost no one cites Sellars, while reinventing his wheels with gratifying regularity”. He explains this by pointing to the fact that philosophers tend to be slow to even recognize approaches that are similar to their own views, while being very fast to acknowledge approaches that are the outright opposites of their own views. In my case this is more due either to me not having been sufficiently familiar with all the potentially relevant Sellarsian ideas that Beisecker mentions or to me not seeing their relevance, and I am grateful for his suggestions. However, I am less inclined to tie thought so

tightly to speech, and thus I am not sure if I could adopt Sellars' ideas on self-knowledge in a straightforward manner.¹¹

With respect to direct influences of Sellars on me, I am aware of just two: first, the force of the general contention that our mental concepts are in some respects theoretical and second, I made a conscious effort not to fall into the trap posed by the myth of the given when building my account of perception. If there are any other influences, then these must be second-hand, through Sellars' influence on Dennett, Davidson and other writers.

Beisecker (2017, 52) actually takes Sellars to be a kind of interpretivist, which I find a bit of a stretch. Commonly, Sellars is seen as an influence on the development of functionalism and theory theory, both of which stand apart from interpretivism. As Beisecker notes, Sellars can indeed be regarded as the father of normative functionalism, but he considers this view to be a version of interpretivism. Normative functionalism conceives its targets (e.g. assertions, concepts or meanings) in terms of their inputs and outputs, where these are both understood normatively: i.e. with reference to the difference that their use makes on the normative status of the participants of the interaction. This in itself need not amount to interpretivism. It will not if the interpretive considerations are not constitutive of those inputs and outputs. I imagine that normative functionalism can be developed in ways compatible with interpretivism, but I am not convinced that Sellars himself gave interpretation a constitutive role in his account. I will come back to the normative functionalist approach later in this section, but before that, some general comments on normativity are in order.

The attention to and even the mention of normativity is curiously absent from my book in Beisecker's (2017, 49) opinion. He suspects that this is not accidental and he is right. At the time of writing the book I had not really made up my mind concerning the relationship between normativity and the mental and at the same time I felt the force of the anti-normativist arguments in the literature to the effect that meaning and content are not normative. I thus thought it is wise not to make my arguments for interpretivism depend upon an unexamined reliance on normativity.¹² For if it turns out that mental states are not normative in the required ways, this damages my case for interpretivism. And if they are normative in the right kind of way, then perhaps this could be used to make a further case for the ascrip-

¹¹ For crucial differences between thinking and speaking from the pragmatist perspective, see (Kukla and Lance 2016).

¹² Thus I do not agree with Beisecker (2017, 51) that the normativity of belief ascription explains why I regard the mental as pleonastic and with no direct empirical commitments. My main reason for treating mental entities as pleonastic is that there simply are no mental entities in the inflationary sense. If so, one should not expect to locate such entities in the brain or to think that they line up with other inflationary entities of the natural basis.

tion theory, but this may be left as a matter for future work. At present, I am still not entirely settled on the issue of normativity. However, I have become even more inclined to think that truth, meaning and intentionality are not constitutively normative (see Mölder 2008; Mölder 2011).

What then are the reasons for my scepticism concerning normativity? Let me say at the outset that I do not deny that our relationships with each other are sometimes normative: we obey norms, hold each other accountable for what we say and try to honour our commitments. Our actions are guided by how things ought to be, and sometimes we do and say things because we should do and say them. I also think that “mindshaping” (Mameli 2001; McGeer 2007; Zawidzki 2013) is a neat idea, which—when its normativity is properly “tamed”—goes a long way toward explaining why folk psychology holds us sway: since folk psychology regulates how we make sense of each other, it shapes our minds into such forms that can be understood in terms of that very framework itself.

However, I advise extreme caution against jumping to philosophically substantial conclusions from all this. It seems to me that sometimes people assume that since something is related to norms, correctness conditions, rules, rationality, sociality or values, it necessarily makes that thing normative. But this is not the case. Separate arguments are needed to move from, say, rationality or sociality to normativity proper (cf. Gibbard 2003, 85).¹³

Moreover, I take it that the normativity of some phenomenon becomes philosophically puzzling if and only if that phenomenon is *essentially* normative. This is to say that the nature of that phenomenon is determined by normative facts or, epistemically, that a complete account of the phenomenon must inevitably be couched in normative terms. In only that case is there a substantial problem about the place of that phenomenon in the natural (non-normative) world and its reduction has slim prospects: we are unlikely to be able to account for the phenomenon in non-normative terms. After all, a phenomenon can be related to normative affairs (e.g. having normative consequences), and yet not itself be essentially normative (Horwich 1998, 188).

The kind of normativity at issue in the case of meaning and intentionality is deontic normativity. It is common to explicate this in terms of prescriptivity (Glüer and Wikforss 2018). Thus if meaning or intentionality are truly normative, they are inextricably bound with prescriptions, which guide our

¹³ Timothy Schroeder (2003) argues in similar vein that although Davidson’s account relies on rationality and consistency that by itself does not make his account normative. In Schroeder’s reading, Davidson’s account includes norms, understood as categorization schemes, but the “force-makers” of these norms, which are required for genuine normativity, play no role in his theory.

actions and tells us what we *ought* to do. I already mentioned the compelling anti-normativist trend in the literature.¹⁴ One strategy the anti-normativists employ is to distinguish kinds of ‘ought’, which fall short of essential normativity, and argue that the alleged normativity that is linked with correctness conditions or rules is in fact not the essential normativity of the required kind, but instead upon closer inspection turns out to be descriptive. I cannot rehearse these discussions here, but I bring out some of their pertinent points concerning the normativity of meaning and content.

One such relevant distinction is a division between categorical and hypothetical prescriptions (Wikforss 2001; Hattiangadi 2006; Hattiangadi 2007). A categorical prescription holds irrespective of one’s desires and it cannot be overridden by them. It is not instrumental on one’s wanting to achieve some goal. Hypothetical prescriptions, by contrast, command one only if further pragmatic conditions hold: one ought to do something if one wants to get a certain result. For example, ‘If you want to finish your draft during the week, you ought to start writing now’ is a hypothetical prescription, which holds on the condition that you want to get your draft done during the week. The anti-normativists’ claim is that the essential normativity of meaning must be of the categorical kind, and it has not been established that it amounts to more than hypothetical prescriptivity. The latter is not enough to secure the essential normativity of a phenomenon, because hypothetical prescriptions can be analysed as being essentially descriptive (Hattiangadi 2006, 228). The statement in the above example can be viewed as describing a way of finishing a paper on time, not as instituting a genuinely normative commitment. Similarly, if one wants to be understood by others, one ought to speak in a certain way (cf. Bilgrami 1993, 135). But this is conditional on one’s desires and intentions.

Another distinction is that between normativity and norm-relativity (Hattiangadi 2007, 7; Finlay 2010, 332; Hattiangadi 2017, 650–651). Normativity involves categorical prescriptions as outlined above. Norm-relativity pertains to norms that set a standard. For instance, norms that fix the use of words as correct and incorrect are normative in the sense of being norm-relative. What makes this distinction important is that meeting a standard itself is not normative (Glüer 1999). It does not yield prescriptions. On the contrary, the statements of standard-setting norms are descriptive. They describe the standard that something either meets or not. This applies to correctness as well, as Hattiangadi (2007, 60) argues: “To say that some use of an expression is correct is ... to say that it *refers to* or is *true of* the thing to

¹⁴ The representative publications of this trend include (Glüer 1999), (Glüer and Pagin 1999), (Wikforss 2001), (Hattiangadi 2006), (Hattiangadi 2007) and (Glüer and Wikforss 2009). For recent overviews, see (Hattiangadi 2017) and (Glüer and Wikforss 2018).

which it has been applied. If we keep this firmly in view, it no longer makes sense to gloss ‘correct’ as ‘ought to be pursued’ but rather as ‘meets a standard’”. On this view, distinguishing between the correct and incorrect use of a word by some norm of use is to supply a standard that sorts the uses of words into those that refer to what they have been applied to and those that do not. This is not normative; it just describes a way of sorting.

What to make of Beisecker’s (2017, 49) charge that I have “downplayed” “the essentially *normative* dimension of mental states and their ascriptions”? Armed with the outlined distinctions, it is possible to argue that intentional states are not essentially normative, although they can be related to norms and be linked to hypothetical prescriptions. Of course, this point depends on how the wider battle between the normativists and anti-normativists pans out, but I think that the anti-normativists have made a good case for the claim that the mind is not essentially normative. If that is true, then an account of the mind or intentionality worthy of its name need not bear the burden of explaining their essential semantic normativity, although it needs to acknowledge the role of norm-relativity and hypothetical prescriptions. And of course, if the aim is to account for the mind in a wider cultural and social context, then the other kinds of norms, such as prudential, moral, evaluative etc. should be heeded too, but I suppose they are less resistant to naturalization than the alleged semantic normativity.

However, what about the fact that in the ascription theory the possession of mental states is constituted by their canonical interpretability, while the interpretation relies on various presumably normative principles? Does this not all by itself make the possession of mental states an essentially normative affair? In reply I would like to point out first that canonical ascribability condition is best viewed as providing a standard for sorting the ascriptions into those that are constitutive for the mental state possession and those that are not. As noted before, such standards are descriptive, rather than normative. Second, the principles of ascription (161–163) can indeed be formulated as rules for the interpreters. However, the normativity of those principles thus articulated is then hypothetical in the sense above, and not categorical. They take the conditional form: if you want to ascribe mental states, here are certain things you should take into account. As such the principles merely describe a way of coming to a definitive interpretation of the subject’s mind. Neither the principles nor the canonical ascribability standard have “normative oomph”.¹⁵ That is, the failure to follow the principles and to meet the canonicity condition is not normative failure. It only results in non-canonical ascriptions. These principles and the canonicity standard do not bind us: we do not have to follow them no matter what. One has

¹⁵ See (Copp 2015, 142) for this notion.

to follow them if one wants to ascribe mental states, but this is so on the condition that one has such an aim.

Beisecker (2017, 50) stresses also the normativity of folk psychological platitudes. He compares them to norms of conduct and conceives them as “normative principles guiding our enforcement of rational norms”. He suggests that conceiving folk platitudes as normative is more illuminating than treating them as empirical generalizations. I agree that we can do many things with folk psychology besides explaining and predicting others’ behaviour. We can use it to regulate the actions of others. This shows that folk psychology can be used in normative ways, but I do not think that this normativity is categorical. If folk platitudes can be laid out (and I harboured some doubts about this in Chapter 4), then even if they are not empirical generalisations, they would be storable in a straightforwardly descriptive form. I think this holds especially if folk platitudes are likened to codes of conduct, since the latter are never categorical. They are hypothetical prescriptions, which are at bottom descriptive.

At this point, a defender of the essential normativity of the mental might object that my rejoinders have missed what is crucial on this issue: namely, that the communal background of norms is a condition for the very possibility of thought and language. It is not something that one could simply adopt if one wants. Neither can one opt out of following these norms without ceasing to be treated as a rational thinker by other members of the community.¹⁶ Put in the terminology of the ascription theory, the possession of mental states is constituted by canonical ascribability, and the latter relies upon the principles of interpretation and upon folk psychology. The normativist could argue that those make up the normative and inescapable background for the possession of mental states. If it were not for these normative relations, one would not have mental states. For this reason, it cannot be claimed that the normativity at issue is merely hypothetical, for one cannot merely choose to be bound by these norms.

I am not swayed by this line of reasoning. For a start, I would like to point out that not following these norms does not preclude the ascription of mental states. Only those ascriptions would not count as canonical. This reply, however, does not touch the point that the canonical ascription constitutes the possession of mental states. But note that the claim that canonical ascription is imbued by norms requires independent backing. For it might turn out that it is not normative after all. To show this I will make use of

¹⁶ Similar considerations are presented against anti-normativists by e.g. Cash (2008, 104) and Buleandra (2008, 185).

yet another argument from the anti-normativist toolbox.¹⁷ The basic idea is that constitutive conditions cannot fulfil a normative role, since they cannot generate prescriptions. Prescriptions imply that it is possible to violate them. But if the principles of ascription and folk platitudes are constitutive of mental state possession, then they determine (or play a role in determining) what it is to have mental states. They cannot prescribe what a canonical ascriber ought to do, since she could not violate such a prescription. The canonical ascriber has no other option than to make the ascription in the canonical manner, for otherwise she would, by definition, not be a canonical ascriber. Thus these principles and platitudes should not be construed as norms.

In sum, I proposed two separate escape routes from the conclusion that the ascription theory implies that the mental is essentially normative. These routes depend on the way in which the consideration in favour of normativity is spelled out. On the first route, the defender of normativity proposes to recast the principles of ascription as rules for interpreters. However, in that case I would argue that if construed in this way, those rules are hypothetical in nature and this is not enough for genuine normativity. On the second route, the normativity is alleged to stem from the constitutive role that the principles play in the possession of mental states. But then the escape lies in the very constitutivity: for constitutive conditions determine but do not prescribe.

Let us return now to normative functionalism, in the direction of which Beisecker tries to nudge me. Could the ascription theory be complemented with an approach along the lines of left-wing Sellarsians? I take it that it is characteristic of these philosophers, of which Robert Brandom (1994) is the most prominent example, that they take Sellars' (1997, 76) metaphor of "the logical space of reasons" very seriously and conceive meaning and intentionality as normative phenomena residing in a different "space" from the one in which phenomena are described by the empirical sciences. This approach understands norms in relation to social practices, for the norms govern common social practices and can be manifested in them. Without

¹⁷ This tool is used by Glüer (1999; 2001), Wikforss (2001, 215), Glüer and Wikforss (2009, 48–52) and Glüer and Wikforss (2018). It is mostly applied on Davidson's principle of charity. Here is one characteristic statement of this point:

the very idea that content attributions have to respect the basic standards of rationality, implies that these standards cannot be taken to be prescriptive. That intentional creatures by and large do accord with the basic standards of rationality is, on this view, a condition of their being intentional creatures, a condition of being able to follow and accept norms, not something prescribed to already intentional creatures. (Glüer and Wikforss 2009, 52)

going into full details of Brandom's account, I close this section with some more or less impressionistic remarks about the prospects of such complementation.

In certain respects, this has the potential to strengthen the case for the ascription theory. Remember that the ascription theory presupposed the feasibility of a general theory of meaning, but did not provide one itself. An account along Brandomian lines could perhaps provide such an account of meaning. A wholehearted embrace of normativity could also provide further support for interpretivism if one could show that only an interpretivist approach could do justice to the inherent normativity of intentional states. Finally, such a partnership could bring the nature and the possession of mental states together again, provided that the latter is understood in ascriptive terms. For Brandom's account can be construed as giving the nature of attitudes in terms of how we acknowledge and attribute them. In his terms, this amounts to explaining entitlements and commitments or "deontic statuses" through "deontic attitude", which is the attitude of taking one as thus entitled or committed (see Brandom 1994, 165–166).

However, even if all this would seem to make the theory stronger I am not sure it could really work out. First, there is the danger that fusing nature and possession in such a way would obliterate the theoretical role of ascription entirely. There is no point in adding my notion of canonical ascription to Brandomian deontic practice. The interactive "practice of giving and asking for reasons" does not require that each mental state be canonically ascribable. Furthermore, one could also dispense with mental states as *states* given the role deontic attitudes and statuses play in Brandom's account, and along with mental states the pleonastic entities should go as well.¹⁸

Second, adding a theory of meaning to the ascription theory should avoid circularity. There is a danger of circularity if the ascription theory presumes that words have meanings, but an account of meaning in turn crucially relies on the possession of mental states, where this possession is conceived in the interpretivist way. And of course if mental states in such a theory are understood along some non-interpretivist lines, that would be even worse, for then the theory of meaning would posit mental states that are not captured by the ascription theory.

At first glance, Brandom's approach seems quite promising in this regard as he elucidates meaning through our communal practices. However,

¹⁸ Gauker's (2003) communicative conception would be another example of an account that rejects beliefs as internal states. He understands the nature of beliefs in terms of their attribution and conceives the latter as assertions performed on someone else's behalf. I, on the other hand, think that there is no need to eschew mental states as long as one conceives them in an ontologically deflationary way. See also the discussion on 148–150.

it seems that this elucidation still invokes the ascription of mental states or intentionality. Brandom (1994) sets out to explain how the correctness—that some applications of concepts are correct and others incorrect—, which he regards as a normative matter arises from social practices, which are also essentially normative. In his account, deontic statuses are explained in terms of deontic attitudes. That is, commitments and entailments are explained in terms of the attributions of commitments and entailments within the deontic practice. At this point one could start to worry about the status of the deontic attitudes themselves (cf. Zawidzki 2015, 600). These consist in “taking or treating someone as committed or entitled” (Brandom 1994, 166), and look very much like intentional states. To be sure, Brandom does not construe them as explicit attitudes. These “takings” can be implicit in a community practice in the pattern of sanctioning certain applications of concepts and reinforcing others. Sanctions and reinforcements in turn can in the most primitive case be conceived as behavioural dispositions.¹⁹ Thus the practice implicitly comes to embody norms for the correct application of concepts.

It seems that if such a picture is joined to the ascription theory, then such a picture—one which finds correctness conditions on behavioural roots—would not create the circularity problem, although it may pose difficulties on a different front; namely in adjudicating the relationship between interpretivism and dispositionalism. However, Brandom cannot rest content with a purely dispositional picture (one probably accepted by Sellars), since that seems not to have enough resources to dispel the familiar rule-following concerns (see Millikan 2005, 62–63; Hattiangadi 2007, 164–165). For one cannot read off which rule has been followed (which use of a concept is correct) by attending to mere dispositions as manifested in the practice. Brandom alleviates such concerns by reference to the external interpretation of the practice that discerns which norms the community in fact has:

The general point is that while normative interpretation of a community as engaged in one set of practices rather than another is underdetermined by nonnormatively specified actual behavior, regularities of behavior, and behavioral dispositions, *relative to such an interpretation*, concepts nevertheless are objective, shared, and unambiguously projectable. (Brandom 1994, 633; see also 648)

Thus beside the non-normative behavioural dispositions, there is a need for an interpreter who interprets the whole community as taking part in practices that are linguistic and implicitly normative. But if this reading of Brandom’s project is correct, then his approach invokes interpretation at the cru-

¹⁹ Thus writes Brandom (1994, 166): “Adopting this practical attitude can be explained, to begin with, as consisting in the disposition or willingness to impose *sanctions*.”

cial juncture, and that makes it incompatible with an account like the ascription theory, which presupposes that meaning is determined independently from interpretation.

Third, I have already expressed my reservations about the essential normativity of the mental. I do not deny that the application of mental terms may have normative consequences, but I do not think that these consequences define those terms or are somehow essential to them. If this is on the right track, then the consideration from normativity cannot be marshalled into an argument for interpretivism. For if the mind is not essentially normative in the relevant sense, then it cannot be argued that the interpretivist approach is preferable, since it accommodates the normativity better than the other theories.

I also share the naturalist's scruples about primitive and ungrounded normativity. To echo Mackie (1977), the categorically prescriptive things in the world are spooky if they are not grounded on or explained by the non-normative facts. Normative functionalists like Sellars and Brandom think that such worries can be alleviated by presenting genetic stories or myths about how the problematic terms became adopted in communities that originally had no use for them (cf. Beisecker 2012, 114). I can see some appeal in such an approach but to my mind the utility of genetic myths is still restricted given that such stories depict idealised situations, which were never actually the case.²⁰ And even if one could concoct an illuminating genetic story about how the norms evolved in a pre-normative community, we still need to explain what, metaphysically speaking, norms are. Even if normative vocabulary is irreducible, then what about the norms themselves? What grounds them? Do they have causal powers? For such questions, and I think they are legitimate questions, genetic stories give no answers.

To conclude, let me emphasize that I do not want to leave the impression that in my view norms and normativity play no role in understanding the mind. My contention is rather that the normativity at issue is of a rather innocuous kind which neither creates grand philosophical puzzles nor gives a reason for regarding the mental as uniquely prone to an interpretivist treatment.

²⁰ Their utility is also limited if such stories smuggle in the very properties they were devised to illuminate. For instance, Haugeland's (1990, 404–405) story about the origin of norms involved creatures with "wired in" conformism, a disposition to imitate and censor others. But this sidesteps the clearly relevant questions of where the disposition to conform comes from and why they evolved to censor other members of the community?

3.2 Brain, beliefs and the folk

Beisecker (2017) takes me to task for the (presumably) thoughtless endorsement of Lynne Rudder Baker's "incautious statement" that the brain is necessary for having beliefs.²¹ In fact, no such endorsement was present in my text. The context of that discussion was an objection to the idea that mental properties are composed by or constituted by physical properties; for this leads to an unreasonable reification of mental entities. My whole point of quoting Baker in 66 was to illustrate one crucial difference between mental states like beliefs and brain states: beliefs are not spatiotemporal entities like brain states. If they are entities in some other sense or of some other kind, how could they be composed of brain states? Now, Baker's claim was that there is this difference between beliefs and brain states even if having the latter is a necessary condition for having the former. I think it is charitable to read her (at least in that quote) as supposing this for the sake of argument, but not necessarily as an endorsement.

Notwithstanding this particular quote by Baker, is having a brain necessary for having beliefs and other mental states? What is the stance of the ascription theory on that very question? As also hinted on 193, the ascription theory leaves this question for folk psychology to settle. The ascription theory presumes that mental states are pleonastic entities and, accordingly, their nature is fixed by the respective mental terms. What it takes to have mental states—their possession conditions—is fixed by what it takes to apply mental terms. In other words, we need to find out whether the ascription of beliefs and other mental states relies upon the fact that the subject of ascription has a brain. Note that such a connection between the ascription and the necessary condition is enabled by interpretivism. Without assuming interpretivism, the ascription practices of people could not be helpful in this regard. The issue of whether brain states are necessary for mental states would have to be settled independently in such a case.

It is possible to object that the ascriptions that lay people make are not canonical, and therefore have no bearing on the metaphysical issue even when we presume interpretivism. According to the ascription theory, how-

²¹ The whole quote is the following:

Having certain kinds of brain states may be necessary for having beliefs; but it does not follow that particular brain states constitute particular beliefs. Brain states are ordinary spatiotemporal entities. Spatiotemporal entities are not widely scattered objects, but are compact objects that have more or less definite boundaries in space in time. A belief is no more constituted by an ordinary spatiotemporal entities than is the British Constitution. (Baker 1995, 184–185)

ever, these ascriptions are canonical if no revisions are required or if they are *stable* (to use Häggqvist's helpful term). We should thus consider only stable ascriptions when looking at lay people's ascription practices.

Recently, experimental philosophers have conducted several studies investigating how lay people attribute mental states. The results seem to suggest that people do not hesitate to ascribe intentions, beliefs and in some cases also phenomenal states to groups, robots, ghosts and other entities that do not have a biological brain, and hence do not have any brain states. It could turn out that the answer to the question of whether having a brain is necessary for having mental states will not be a uniform one: there are some experimental studies showing that people are reluctant to attribute certain mental states like feeling pain to groups or robots, while routinely ascribing beliefs and other attitudes to them (Knobe and Prinz 2008). The jury is still out on how the data resulting from these studies is to be interpreted (see (Sytsma 2016) for a recent overview). Philosophers also debate whether the ascription of phenomenal states presumes the existence of a biological body or not, and there is some evidence that even phenomenal states are ascribed on the basis of functional cues such as behaviour, other mental states and stimuli with no regard to embodiment (Buckwalter and Phelan 2014). However, one can glean the following minimal message from these studies: if lay people attribute at least some kinds of mental states to disembodied and other entities that have no brains, then this suggests that having a brain is not regarded as a necessary condition for the ascription of every kind of mental state. Given interpretivism, this allows us to move from ascription to possession and to deny that the brain plays a necessary role in having mental states.

This brings me to a related question about the relationship between folk psychology and the ascription theory. Häggqvist (2017, 13) records his disappointment that I do not present a separate account about what it is to be a believer. Instead I rely on folk psychology for this. Indeed I do not think that such a separate story is needed, but I do not see a reason for disappointment over this. This is a feature of the ascription theory, not a bug. Given that the theory does not set substantial conditions on believers it can be complemented with different folk psychologies or even with a scientific psychology. For instance, certain subpersonal constructs of a psychological theory can also be canonically ascribable in accordance with the ascription theory and thus be placed within the interpretivist framework (cf. 71–72). By deferring to the folk, the ascription theory secures a link with the empirical reality. What we regard as a thinker is an empirical question about folk practice, not something that should be fixed from the armchair. Häggqvist writes as if the fact that the ascription theory is a metaphysical account of the posses-

sion conditions of mental states requires such an armchair account. But I do not think that there is such a requirement. Even if the ascription theory is not an empirical account of our psychology, it is a theory about the mental states as we, the folk, understand them. This is why I stressed the importance of not losing contact with ordinary practice in the book.

3.3 On intentional patterns

What is the proper understanding of Dennett's "real patterns"? Häggqvist (2017, 16) suggests that, for Dennett, patterns are just "patterns in behaviour" and thus "they are not themselves "intentional"". He notes correctly that when discussing Dennett's position in *Mind Ascribed* I sometimes wrote about *intentional* patterns and understood them in the sense of patterns of intentional states. Was it then a slip-up?

Let me point out first that the term "intentional pattern" is not my invention; it is widely used in the literature (e.g. Bennett 1991, 25; Baker 1994; Slors 2001, 212–213; Brook and Ross 2002, 21). Even one subsection in Dennett's "Real Patterns" is titled "The Reality of Intentional Patterns" (Dennett 1991, 42). The term also occurs in his earlier writings, collected in *The Intentional Stance*.²²

The idea behind Dennett's talk about patterns is to highlight that a pattern is real if it becomes discernible as a result of the compression of data. If no efficient compression is possible, and the data can be reproduced only through an exact copy, then it lacks a real pattern in Dennett's sense (Dennett 1991, 34; Zawidzki 2007, 145). Dennett applies this idea to the explanation of behaviour in intentional terms. If approaching a certain process from the intentional stance can yield a pattern that would be invisible amidst the myriad of data, as conceived in physical terms, then we have detected a real intentional pattern.

But what is this pattern, really? Is it just a complex behavioural trajectory? Note that this would make Dennett's position almost indistinguishable from behaviourism. If the patterns were merely behavioural, then it would be unclear what the fuss is about their being real. Nowadays it is hard to find someone who denies the reality of behaviour, and it would be uncharitable

²² Here are two examples from this book: "It is important to recognize the objective reality of *the intentional patterns* discernible in the activities of intelligent creatures, but also important to recognize the incompleteness and imperfections in the patterns." (Dennett 1987, 28; my italics)

"It is as hard to say where the intentional order is as it is to say where *the intentional patterns* are in the Life world. If you "look at" the world in the right way, the patterns are obvious. If you look at (or describe) the world in any other way, they are, in general, invisible." (Dennett 1987, 39; my italics)

to hold that Dennett was defending only a “mild” realism of behavioural patterns. By invoking the notion of real patterns, Dennett sought to secure a status for intentional states that would be more realist than mere instrumentalism. But if the real patterns were only behavioural patterns, then that by itself would do nothing to secure a realist standing for intentional states. Furthermore, when the patterns in behaviour were understood in non-intentional terms, then those would also be easily discernible from the physical stance, as they are just complex physical movements. The intentional stance would be redundant on such a reading.

Let us search for another answer. Consider the following quotation from Dennett:

Are there beliefs? Of course. We already know that there are robust, reliable patterns in which all behaviorally normal people participate—the patterns we traditionally describe in terms of belief and desire and the other terms of folk psychology. (Dennett 1988, 501)

As Dennett writes, the patterns under scrutiny are patterns described in folk psychological terms. What emerges from this is a reading of Dennett’s patterns as structures that are given descriptions in intentional vocabulary. These descriptions involve belief and desire attributions, a conception of actions as driven by intentions and so on. In short, to discern intentional patterns is to make sense of something that from the physical stance would look like incomprehensible strings of physical movements. To discern intentional patterns is to make sense of intentional *action* in terms of intentional states (cf. also Nelkin 1994, 61).

It might seem, however, that “the patterns that we characterize in terms of the beliefs, desires, and intentions of rational agents” (Dennett 1987, 27) still blurs some important distinctions. For instance, Brian McLaughlin and John O’Leary-Hawthorne (1994, 209) claim that it runs together “the patterns of dispositions to peripheral behavior that get interpreted from the intentional stance” and “the belief-desire patterns that patterns of dispositions get interpreted as being”.

I think that there are actually two related distinctions here: one is worth making, but the other cannot be applied to Dennett’s patterns. Let us start with the latter. This is a distinction between the ontological nature of patterns (“What patterns really are?”) and our epistemic take on them (“How we make sense of the patterns?”). But no such distinction can be had on a reading of Dennett’s patterns as recognition-dependent entities (which I also defended in the book; see the quotes on page 116). If patterns are such that they can be discerned only by taking the intentional stance and by applying intentional vocabulary, then they do not have non-epistemic identity condi-

tions. What could be discerned and individuated from the physical stance is an unintelligible complexity, not an intentional pattern.

The distinction that is at least sometimes important to keep is the one between patterns of action and dispositions to act, on the one hand, and the patterns of beliefs, desires, thoughts and other mental states, on the other hand. But in some cases it is not clear if that distinction can be drawn: after all, discerning an act of trying, a coming to a decision or a bout of sadness are not relevantly different from ascribing beliefs and desires as they all require adopting an intentional stance.²³

3.4 Self-knowledge

Dominic Murphy suggests that a transparency account along the lines of Gareth Evans (1982) and Richard Moran (2001) would fit better with my interpretivist allegiance. I will first briefly outline the account presented in the book and then I will comment on the approach suggested by Murphy.

In the book I was looking for an account of self-knowledge that would be consistent with the general form of the ascription theory, and which would do justice to the intuitive difference between knowing one's own mind and the minds of others. This may seem challenging, for the canonical ascribability conditions are in their nature third-personal and give no special role to considerations from the first-person point of view. I began by pointing out that it is part of common-sense psychology that people have authority over their minds. More precisely, when one says that one has a certain mental state, this is generally regarded as true. However, this folk psychological presumption cannot stand on its own. One has to explain why it holds and I reached out to subpersonal mechanisms for the required explanation. On this view, first-person authority is not a conceptual truth, but a contingent principle that holds due to the way we are built. This is how we normally function, where 'normally' is here understood in a statistical, not a strong normative, sense.

In line with the ascription theory, the possession conditions for mental states are constituted by their canonical ascriptions. Actual self-ascriptions (or avowals or reports) of mental states are not automatically canonical sim-

²³ McLaughlin and O'Leary-Hawthorne (1994, 209) regard the following difference as crucial: "It may be that types of belief-desire patterns are expressible only in an intentional vocabulary. However, each of the types of patterns of dispositions to peripheral behavior that get interpreted as belief-desire patterns can, *in principle*, be expressed in a nonintentional vocabulary." I think that this holds only when 'dispositions to behaviour' are understood in the sense of dispositions to bare physical movements. However, this is not the sense of the patterns on which the adoption of the intentional stance is needed to discern them. Thus, in the context of the intentional stance theory, this particular difference does not come up.

ply on account of their being self-ascriptions, though normally they pass the standard for canonicity. This is because of the subpersonal machinery that maintains the congruence of one's behaviour in a given environment, generally speaking. A part of that congruence is the harmony between one's actions and one's self-reports. Remember that the environment and the behaviour are also among the sources of the canonical ascription. It is then no wonder that whenever a state is canonically ascribable, one's subpersonal machinery can produce the self-ascription of the same state, provided one is appropriately prompted.

Although this view has some affinity with reliabilism, the relevant subpersonal machinery is understood quite broadly. There is no commitment to a single mechanism for self-knowledge, like on some inner sense views that postulate a dedicated system for introspection. For this reason, my account is not inconsistent with interpretivism, which it would have been—as Murphy rightly suggests—if it were committed to a mechanism that detects or observes mental states as they pass through one's mind.

Murphy does not really object to my account apart from when he notes that the idea of a subpersonal mechanism *triggering* self-ascriptions remains obscure. Perhaps the choice of that term was not entirely befitting, indeed. I meant to speak of the subpersonal processes that lead to a self-ascription. Since the self-ascription under consideration is an actual act, not just the possibility of the act, this act must be underpinned in each of its instances by certain processes, and those in turn will be caused by certain efficacious subpersonal processes (in line with the view on mental causation developed in Chapter 7). Thus, whenever one does actually self-ascribe a mental state to one, this act (or more exactly, the process underpinning this act) is caused by certain mechanisms in one's brain. Being an interpretivist, I took some pains to avoid the traditional view that it is either one's unspecified experience, or the mental state itself, that triggers the self-ascription (271–272). Since that would invoke interpretation-independent mental entities, this option is not available to an interpretivist. Another aspect that I wanted to convey with the notion of triggering is that sometimes self-ascriptions can be caused by internal subpersonal processes without any external prompts. Why this happens is something for psychology to explain.

Let us now move on to the Evansian transparency account of self-knowledge. Here the basic idea is that self-knowledge is a kind of deliberation: to find out if I believe that p I rely on the same procedures I usually employ in order to find out if p is true. These procedures are directed outwards to the world, not to the inner stage of one's mind, as the traditional picture has it. There is no dedicated mechanism for knowing one's own mind. I just make up my mind by finding out truths about the world. This account ap-

plies in the first instance to one's knowledge of one's current attitudes, and Murphy (2017a, 44) suggests extending it to dispositional states with the help of memory. In that case, one just recalls if one thinks that p is true: "I take a content from memory storage—say in episodic or semantic memory—and put it in working memory".

The leanness of this approach is appealing indeed. However, I have some scruples about it. They are reservations about the account for the most part from an interpretivist point of view. The issues I mention are however not decisive objections.

First, I do not see it as "a better bet" for interpretivism like Murphy suggests. It might not even be consistent with the view. The view relies crucially on the cognitive processes for fixing one's world-directed attitudes: Murphy (2017a, 44) writes about "the procedure for forming that belief or desire". This and his remarks on retrieving "content from memory storage" sound very much in the spirit of a standard functionalist cognitive psychology. There seem to be ascription-independent cognitive processes operating on stored memories, as well as on beliefs and desires as internal states. Beliefs and memories participate in those processes *qua* beliefs and memories, not as subpersonal items, and operations on them are sensitive to their contents. Both these cognitive processes as well as the memories and beliefs with contents that these processes implicate seem to be independent from the interpretation in this view. I am not saying that it is impossible to give an ascriptivist gloss on such cognitive processes, but that does not appear to be a part of Murphy's model. On the face of it, the transparency account has traded the inner mechanism for internal states.

Second, as presented, the account does not seem to yield a straightforward explanation for the authority of self-knowledge. Why assume that the first-person access to one's mental states has authority over third-person access? Is it because the others do not have access to my cognitive procedures? But since the procedure is a world-directed means of deciding whether p is the case, then they do not seem to be special in the way self-knowledge seems special. There is more promise for explaining authority in that manner in the case of memory, for others normally cannot get into my memory storage, but this does not help with the basic case.

Third, the range of application of the transparency account is *prima facie* restricted to one's occurrent and remembered attitudes. This does not include knowledge of one's own emotions, sensations and bodily feelings, to name just a few kinds of mental states apparently left out by the account (cf. Coliva 2016, 123). What would be the outwards-looking or memory-involving cognitive procedure for deliberating about whether one is in pain, for instance? There do not seem to be any simple tweaks in the offing to

accommodate such states if one wants to keep the idea that self-knowledge is an outwards-looking matter. To propose to follow whatever procedure one follows in order to find out if one is in pain would merely be to beg the question. For that would presume an access to pain, but the nature of this access is precisely what needs to be accounted for. Apart from this, if the proposal is that in the case of pain I am somehow reaching inwards to the pain itself to form the respective belief that I am in pain, then this solution is not available for an interpretivist, who cannot help herself with such internal and ascription-independent mental states.

However, I do not want to sound uncharitable here. In fact, I am inclined towards pluralism about self-knowledge; that is, I could admit that since different kinds of mental states have different properties in our folk psychology, they are also known in different ways.²⁴ Thus an account of the self-knowledge of beliefs need not be the same as the account of the self-knowledge of bodily feelings. An interpretivist can adopt such different accounts, provided that their commitments are consistent with interpretivism. Accordingly, I do not want to preclude the possibility that, say, the deliberative kind of self-knowledge can be explained along the lines of the transparency account or better (as already hinted in 273–274), along the lines of some amalgam of the ascriptionist and the transparency models.

Acknowledgements

The research for this paper was supported by the Estonian Science Foundation grant ETF9117 (“Mind without mental states?”), the Centre of Excellence in Estonian Studies (European Regional Development Fund) and is related to research project IUT20-5 (Estonian Ministry of Education and Research). I am grateful to Uku Tooming for comments and to Alex Davies for checking my English.

Bibliography

- Armstrong, D. (1989). *Universals: An Opinionated Introduction*, Westview Press, Boulder.
- Baker, L. R. (1987). *Saving Belief: A Critique of Physicalism*, Princeton University Press, Princeton.
- Baker, L. R. (1994). Content meets consciousness, *Philosophical Topics* 22: 1–22.

²⁴ For one possible pluralist treatment of self-knowledge, see (Coliva 2016).

- Baker, L. R. (1995). *Explaining Attitudes: A Practical Approach to the Mind*, Cambridge University Press, Cambridge.
- Baker, L. R. (2003). Belief ascription and the illusion of depth, *Facta Philosophica* 5: 183–201.
- Beisecker, D. (2012). Normative functionalism and its pragmatist roots, *Social Epistemology Review and Reply Collective* 2: 109–116.
- Beisecker, D. (2017). Whither normativity?, *Studia Philosophica Estonica* 10.2: 46–54.
- Bennett, J. (1991). Analysis without noise, in R. J. Bogdan (ed.), *Mind and Common Sense: Philosophical Essays on Commonsense Psychology*, Cambridge University Press, Cambridge, pp. 15–36.
- Bilgrami, A. (1993). Norms and meaning, in R. Stoecker (ed.), *Reflecting Davidson*, Walter de Gruyter, Berlin, pp. 121–144.
- Block, N. (1995). An argument for holism, *Proceedings of the Aristotelian Society* 95: 151–169.
- Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds, *Philosophical Studies* 61: 127–148.
- Brandom, R. (1994). *Making it Explicit: Reasoning, Representing, and Discursive Commitment*, Harvard University Press, Cambridge, MA.
- Brook, A. and Ross, D. (2002). Dennett's position in the intellectual world, in A. Brook and D. Ross (eds), *Daniel Dennett*, Cambridge University Press, Cambridge, pp. 3–37.
- Buckwalter, W. and Phelan, M. (2014). Phenomenal consciousness disembodied, in J. Sytma (ed.), *Advances in Experimental Philosophy of Mind*, Bloomsbury, London, pp. 45–73.
- Buleandra, A. (2008). Normativity and correctness: A reply to Hattiangadi, *Acta Analytica* 23: 177–186.
- Cash, M. (2008). Thoughts and oughts, *Philosophical Explorations* 11: 93–119.
- Coliva, A. (2016). *The Varieties of Self-Knowledge*, Palgrave Macmillan, London.
- Copp, D. (2015). Rationality and moral authority, in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics, Vol. 10*, Oxford University Press, Oxford, pp. 134–159.
- Davidson, D. (2001). Comments on Karlovy Vary papers, in P. Kotatko, P. Pagin and G. Segal (eds), *Interpreting Davidson*, CSLI Publications, Stanford, pp. 285–307.

- Davies, M. (1995). The philosophy of mind, in A. C. Grayling (ed.), *Philosophy: A Guide Through the Subject*, Oxford University Press, Oxford, pp. 250–335.
- Demeter, T. (2013). Mental fictionalism: The very idea, *The Monist* **96**: 483–504.
- Dennett, D. C. (1987). *The Intentional Stance*, The MIT Press, Cambridge, MA.
- Dennett, D. C. (1988). Précis of *The Intentional Stance, Behavioral and Brain Sciences* **11**: 495–546.
- Dennett, D. C. (1991). Real patterns, *Journal of Philosophy* **87**: 27–51.
- Dennett, D. C. (1994). Get real, *Philosophical Topics* **22**: 505–568.
- Dennett, D. C. (2007). Heterophenomenology reconsidered, *Phenomenology and the Cognitive Sciences* **6**: 247–270.
- Ellis, B. (2001). *Scientific Essentialism*, Cambridge University Press, Cambridge.
- Evans, G. (1982). *The Varieties of Reference*, Clarendon Press, Oxford.
- Finlay, S. (2010). Recent work on normativity, *Analysis* **70**: 331–346.
- Fodor, J. A. (1998). *Concepts: Where Cognitive Science Went Wrong*, Clarendon Press, Oxford.
- Francken, J. C. and Slors, M. (2014). From commonsense to science, and back: The use of cognitive concepts in neuroscience, *Consciousness and Cognition* **29**: 248–258.
- Francken, J. C. and Slors, M. (2018). Neuroscience and everyday life: Facing the translation problem, *Brain and Cognition* **120**: 67–74.
- Gauker, C. (2003). *Words without Meaning*, The MIT Press, Cambridge, MA.
- Gibbard, A. (2003). Thoughts and norms, *Philosophical Issues* **13**: 83–98.
- Glüer, K. (1999). Sense and prescriptivity, *Acta Analytica* **14**: 111–128.
- Glüer, K. (2001). Dreams and nightmares: Conventions, norms and meaning in Davidson's philosophy of language, in P. Kotatko, P. Pagin and G. Segal (eds), *Interpreting Davidson*, CSLI Publications, Stanford, pp. 53–74.
- Glüer, K. and Pagin, P. (1999). Rules of meaning and practical reasoning, *Synthese* **117**: 207–227.
- Glüer, K. and Wikforss, Å. (2009). Against content normativity, *Mind* **118**: 31–70.
- Glüer, K. and Wikforss, Å. (2018). The normativity of meaning and content, in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.

URL: <https://plato.stanford.edu/archives/spr2018/entries/meaning-normativity>

- Godfrey-Smith, P. (2004). On folk psychology and mental representation, in H. Clapin, P. Staines and P. Slezak (eds), *Representation in Mind: New Approaches to Mental Representation*, Elsevier, Oxford, pp. 147–162.
- Hattiangadi, A. (2006). Is meaning normative?, *Mind and Language* **21**: 220–240.
- Hattiangadi, A. (2007). *Oughts and Thoughts: Rule-Following and the Normativity of Content*, Clarendon Press, Oxford.
- Hattiangadi, A. (2017). The normativity of meaning, in B. Hale, C. Wright and A. Miller (eds), *A Companion to the Philosophy of Language: Second Edition*, Wiley-Blackwell, Oxford, pp. 649–669.
- Haugeland, J. (1990). The intentionality all-stars, *Philosophical Perspectives* **4**: 383–427.
- Häggqvist, S. (2017). But is it interpretivism?, *Studia Philosophica Estonica* **10.2**: 8–17.
- Hill, C. S. (1997). Critical study of Lynne Rudder Baker, *Explaining Attitudes: A Practical Approach to the Mind*, *Noûs* **31**: 132–142.
- Horwich, P. (1998). *Meaning*, Clarendon Press, Oxford.
- Jackman, H. (2017). Interpretivism and “canonical” ascriptions, *Studia Philosophica Estonica* **10.2**: 28–37.
- Kim, J. (2003). Blocking causal drainage and other maintenance chores with mental causation, *Philosophy and Phenomenological Research* **67**: 151–176.
- Knobe, J. and Prinz, J. (2008). Intuitions about consciousness: Experimental studies, *Phenomenology and the Cognitive Sciences* **7**: 67–85.
- Kukla, R. and Lance, M. (2016). Speaking and thinking, in J. R. O’Shea (ed.), *Sellars and His Legacy*, Oxford University Press, Oxford, pp. 81–99.
- Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*, Penguin, New York.
- Mameli, M. (2001). Mindreading, mindshaping, and evolution, *Biology and Philosophy* **16**: 597–628.
- McCulloch, G. (1990). Dennett’s little grains of salt, *The Philosophical Quarterly* **40**: 1–12.
- McGeer, V. (2007). The regulative dimension of folk psychology, in D. D. Hutto and M. Ratcliffe (eds), *Folk Psychology Re-Assessed*, Springer, Dordrecht, pp. 137–156.

- McLaughlin, B. and O'Leary-Hawthorne, J. (1994). Dennett's logical behaviourism, *Philosophical Topics* **22**: 189–258.
- Millikan, R. G. (2005). The father, the son, and the daughter: Sellars, Brandom, and Millikan, *Pragmatics and Cognition* **13**: 59–71.
- Mölder, B. (2008). Normativity and deflationary theories of truth, *Studia Philosophica Estonica* **1.2**: 179–193.
- Mölder, B. (2010). *Mind Ascribed: An Elaboration and Defence of Interpretivism*, John Benjamins, Amsterdam.
- Mölder, B. (2011). Normatiivsusse mitu tahku, *Studia Philosophica Estonica* **4.1**: 52–82.
- Mölder, B. (2016). Mind and folk psychology: A partial introduction, *Studia Philosophica Estonica* **9.1**: 1–21.
- Moran, R. (2001). *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton University Press, Princeton.
- Murphy, D. (2017a). Ascribing minds and knowing what you think, *Studia Philosophica Estonica* **10.2**: 38–45.
- Murphy, D. (2017b). Can psychiatry refurbish the mind?, *Philosophical Explorations* **20**: 160–174.
- Nelkin, N. (1994). Patterns, *Mind & Language* **9**: 56–87.
- Peacocke, C. (1992). *A Study of Concepts*, MIT Press, Cambridge, MA.
- Quine, W. V. O. (1969). Natural kinds, *Ontological Relativity and Other Essays*, Columbia University Press, New York, pp. 114–138.
- Ravenscroft, I. (2005). *Philosophy of Mind: A Beginners Guide*, Oxford University Press.
- Schiffer, S. (2003). *The Things We Mean*, Clarendon Press, Oxford.
- Schroeder, T. (2003). Donald Davidson's theory of mind is non-normative, *Philosophers' Imprint* **3**: 1–14.
- Sehon, S. R. (1997). Natural-kind terms and the status of folk psychology, *American Philosophical Quarterly* **34**: 333–344.
- Sellars, W. (1997). *Empiricism and the Philosophy of Mind*, Harvard University Press, Cambridge, MA.
- Slors, M. V. P. (2001). *The Diachronic Mind: An Essay on Personal Identity, Psychological Continuity and the Mind-Body Problem*, Springer, Dordrecht.
- Slors, M. V. P. (2017). Interpretivism and the meaning of mental state ascriptions, *Studia Philosophica Estonica* **10.2**: 18–27.

- Sytsma, J. (2016). Attributions of consciousness, in J. Sytsma and W. Buckwalter (eds), *A Companion to Experimental Philosophy*, Wiley-Blackwell, Oxford, pp. 259–278.
- Wikforss, A. M. (2001). Semantic normativity, *Philosophical Studies* **102**: 203–226.
- Winch, P. (1964). Understanding a primitive society, *American Philosophical Quarterly* **1**: 307–324.
- Wright, C. (1995). Can there be a rationally compelling argument for anti-realism about ordinary (“folk”) psychology?, *Philosophical Issues* **6**: 197–221.
- Zawidzki, T. W. (2007). *Dennett*, Oneworld Publications, Oxford.
- Zawidzki, T. W. (2013). *Mindshaping: A New Framework for Understanding Human Social Cognition*, The MIT Press, Cambridge, MA.
- Zawidzki, T. W. (2015). Dennett’s strategy for naturalizing intentionality: An innovative play at second base, *Philosophia* **43**: 593–609.