

# Interpretivism and the Meaning of Mental State Ascriptions

Marc Slors

Faculty of Philosophy, Theology and Religious Studies, Radboud University

---

Interpretivism is often seen as the theory according to which mental state ascription is useful, even though mental states do not really exist. This “as if” theory is widely held to be untenable. In this paper I argue that in order to avoid an “as if” reading of interpretivism, we should embrace the strongest version of this theory.

*Keywords:* mental states, mind-body problem, interpretivism

---

## 1. Introduction

To lay my cards on the table straightaway: I was already a convert of interpretivism. The nature of a person’s mental states depends, as far as I am concerned, on the interpretability of her behaviour. Thus, reading Bruno Mölder’s wonderfully thorough and detailed elaboration and defense of that position did not change my mind—I found myself in agreement with the general thrust of *Mind Ascribed*<sup>1</sup> and with very many of its details. The book is an important contribution to the literature on philosophy of mind, precisely because it presents the undervalued and underrepresented position of interpretivism in minute detail. It positions interpretivism vis-à-vis the more current positions in philosophy of mind—something that has not been done in this way since Dennett’s “Midterm Examination” (Dennett 1987, 339–350), which may be one of the reasons why interpretivism is usually treated as the “odd one out” in the philosophy of mind arena. Mölder’s grounding of interpretivism in a clear metaphysics of pleonastic and natural properties, his connecting interpretivism with functionalism, practical

*Corresponding author’s address:* Marc Slors, Faculty of Philosophy, Theology and Religious Studies, Radboud University, Erasmusplein 1, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands. Email: m.slors@ftr.ru.nl.

<sup>1</sup> Unless indicated otherwise all references with just a page number will be to (Mölder 2010).

realism or dispositional theories, and the (long awaited) description of an interpretivist take on issues such as mental causation (where I was very glad to see a variation on the program model being put to use) and self-knowledge make for the most complete picture of the interpretivist position I have encountered so far.

Despite my general agreement, in this comment I shall concentrate on where Mölder's version of interpretivism differs from the version I would prefer. I shall argue in what follows that Mölder's position is not radical enough to dispel what I take to be a misrepresentation of interpretivism as an "as if" theory (McCulloch 1990) or a theory that may just as well have been called 'fictionalism' (Ravenscroft 2005). On this misconception, interpretivism amounts to the idea that we *pretend* that humans and other systems have minds for practical purposes, even though, for all we know, there may not be such things as minds. This misconception of interpretivism hinges, as I will explain below, on the idea that there is a meaning of 'mental' that is not exhausted by our practices of mental state ascription. It is this "surplus meaning" that gives sense to the idea that there may be an element of pretense in practices of mental state ascription; we may be taken to refer to something that does not exist. In the next section I shall describe why Mölder is inclined to accept this "surplus meaning" of 'mental' as a means of avoiding an overly radical and hence possibly indefensible reading of interpretivism. I shall also argue, however, that by accepting this idea and hence allowing—inadvertently—for an "as if" reading of interpretivism, interpretivism will precisely be perceived as being an indefensibly radical position.

In Section three, however, I will argue that, paradoxically, accepting the more stringent form of interpretivism that Mölder labels "pure ascriptivism" is the best way of avoiding an "as if" reading of interpretivism. For on this more radical position there is no meaning of 'mental' beyond our ascription practices, which implies that there really is no sense to be made of an "as if" claim. I will also argue that pure ascriptivism is a much more defensible position than Mölder makes it out to be. In the final section of this comment, I will argue further that pure ascriptivism allows for an easier and more straightforward connection between interpretivism and mainstream positions such as functionalism or dispositionalism.

## 2. Intermediate ascriptivism and the "as if" problem

Mölder describes his position as "intermediate ascriptivism." This position covers the middle ground between "revelationism" and "pure ascriptivism." According to revelationism, intentional patterns in the behaviour of, say, humans exist independently of our intentional interpretations. Pure ascriptivism, on the other hand, holds that the existence of mental states is ex-

hausted by the interpretability of behaviour in intentional terms. Intermediate ascriptivism is the “idea is that the ascription of mental states uncovers an intentional pattern, which is not merely a matter of projection of interpreters” while it is nevertheless the case that “the interpretability is ineliminable from the full account of the pattern” (82). Mölder’s motivation for opting for intermediate rather than pure ascriptivism is the fact that he portrays pure ascriptivism as a radical theory. The theory appears radical because it seems to make *my* having a mental state dependent on interpretation *by others*.

But the price that is to be paid for this strategy of avoiding radicality is to accept that the meaning of ‘mental’ is not exhausted by our practices of ascription of mental states. Intermediate ascriptivism rejects, in Mölder’s words, the idea that “one’s having mental states is exhausted by one’s being accordingly interpreted or interpretable.” It rejects the notion that “[i]nterpretability is ... a necessary and a sufficient condition for having mental states” and hence that ascriptivism “is a constitutive account of mental-state possession.” (81) The upshot of this is that the meaning of ascribing a mental state to someone involves more than “merely” the appropriateness or fruitfulness of that ascription in the context of our interpretive practices. There is a *surplus meaning* of mental state ascriptions that is not captured by an (imaginary) complete description of our ascription practices.

Mölder makes it clear that this surplus meaning is not to be understood in terms of reference to internal states. Mental terms are not applicable, according to him, “in virtue of the terms corresponding to internal mental entities. Rather, they are applicable in virtue of us sharing a folk-psychological conception that determines which sort of mental states are required in order for us to make sense of given behaviour in a given environment. The ‘state’-talk is a mere *façon de parler* ...” (149) Sense-making, rather than reference, is central in the surplus meaning that exceeds our practices of mental state ascription, then: “[t]he basic motive for mental state attribution is to make sense of the subject’s behaviour and other mental states in mental terms.” (176)

But if the ascription of mental states is not really ascription of inner states, then in what sense can it be more than the mere appropriateness of applied mental predicates, as is required by intermediate ascriptivism? As I read Mölder, intermediate ascriptivism answers this question by allowing for the idea that there is a common meaning of mental state ascriptions that *enables* or *drives* our interpretive practices. By contrast, pure ascriptivism would simply claim that the meaning of mental state ascriptions is something that emerges *as a result* of our practices of mental state ascription.

But what would this common meaning consist in? What does it *mean* to ascribe *mental* states to people on an interpretivist account (if this meaning is not to be cashed out in terms of postulating inner states)? This is not just a question for Mölder. We can ask the same question to Daniel Dennett when he disconnects the meanings of intentional state ascriptions from our factual intentionalistic language games by allowing e.g. thermostats to have beliefs and desires as well (Dennett 1987, 22–34; see Slors 1996). Where Dennett does not even seem to recognize the problem, though, Mölder is clear about the fact that he considers this to be a problem that lies outside the scope of his book: “I . . . think that there is a serious task of working out a substantial theory of meaning. As concerns the ascription theory, however, all we need to assume is that terms already have some kind of meaning, but further details of the theory of meaning are not required.” (11) This is of course not entirely true, for we need to make sure that the details of such a theory do not contradict the central tenets of interpretivism. Mölder acknowledges this carefully: “meaning should not be explicated in terms of mental phenomena. For if the possession of such mental phenomena were in turn explicated along interpretivist lines, the account would be circular.” (152) Thus, the account that he envisages “would take those social roles that govern the use of our terms as external to our minds: they are thus abstract and objective patterns of usage.” (152)

But this indication of a theory of meaning that captures the meaning of mental state ascriptions is ambiguous. If the meaning of mental state ascriptions does indeed consist in abstract and objective patterns of usage, then why reject, as the intermediate ascriptivist does, that “one’s having mental states [i.e. one’s correctly being described as having mental states; M.S.] is exhausted by one’s being accordingly interpreted or interpretable”? (81) What I take Mölder to mean when he speaks of a theory of the meaning of mental state talk in terms of its patterns of usage, then, is best captured as follows: there is an understanding of the correct usage of mental state terms that *governs* the actual use of such terms and that hence precedes our practices of mental state attribution. And it is this *normative* understanding of the use of mental predicates that constitutes the surplus meaning of mental state ascriptions that is not to be reduced to the *factual* practices of ascription.

But even if this surplus meaning, this normative understanding of the use of mental predicates, need not hinge on *actual* reference to hidden internal states that cause our behaviour, there is still the option that it may hinge on *imagined* reference to inner states. Speaking of mental states is, in Mölder’s words, a *façon de parler*, but the question is how much this *façon de parler* contributes to a normative understanding of correct mental predi-

cate application that would govern our ascriptive practices. It seems at least possible to consider the surplus meaning to consist in the fact that the users of psychological language *take themselves* to refer to inner states that cause behaviour. That is, even if philosophers can explain this surplus meaning without reference to inner states in terms of the rules of a psychological language game, it may still be the case that the actual players of this language game only understand the rules of mental state ascription if they take themselves to be referring to inner states. In view of the fact that many people have more or less Cartesian ideas about the mental and hence about the meaning of mental state terms, this is a live option when we concede that the meaning of mental state terms is not exhausted by our practices of mental state ascription.

And that leaves the door open to an “as if” reading of interpretivism. Many philosophers take interpretivism to consist in the idea that we treat each other as if we have mental states (i.e. that ascribing mental states involves postulating inner states that cause behaviour) while in fact there are no such things (which is indeed what Mölder argues; cf. McCulloch 1990, Ravenscroft 2005). This depiction of interpretivism, however, turns the theory into a nigh-unacceptable option. It turns it into the idea that our day-to-day talk about mental states is characterized by reference to non-existing entities. Although this may prove to serve some purpose, it would nevertheless be the case that we merely *pretend* there to be mental states. If interpretivism is to be a defensible option, then in my opinion we should take care to rule out this reading. Paradoxically, as I will argue in the next section, it is by opting for what appears to be the more extreme theory of pure ascriptivism that we can do away with the idea of a surplus meaning of mental state ascription that allows for the option of interpretivism as an “as if” theory.

### 3. In defence of pure ascriptivism

Why does Mölder insist on intermediate ascriptivism? Why is pure ascriptivism considered extreme (82)? Mölder is not really explicit on this. But between the lines it can be read that what makes pure ascriptivism objectionable is that it appears to make *my* being in mental state *M* dependent upon someone else’s ascribing or being in a position to ascribe that state to me. And that sounds absurd. Mental states are not Cambridge properties. Whether or not I am in mental state *M* should be dependent upon interpreter-independent properties—however pleonastic or relational. It should not be up to an interpreter.

But pure ascriptivism need not be committed to the idea that what mental states I am in depends on the whims of some interpreter. As Mölder often acknowledges, the mental states a person is in need not depend on the ac-

tual interpretations that observers are inclined to attribute, but rather on a specific kind of *interpretability*. And it seems perfectly possible to cast interpretability in objective, i.e. actual interpreter-independent ways. Dennett (1987, 1991), for instance, is clear about how this should be done. Given behaviour is interpretable in terms of mental states  $M_1 \dots M_x$ , according to him, when it is predictively or explanatorily *useful* to do so; that is when such an ascription yields fruitful new predictions and ties in with other behaviour in an explanatory way. Defined in this way, the usefulness of ascriptions is an objective feature of them that is not up to the interpreter. Interpretability of my behaviour in terms of mental states  $M_1 \dots M_x$ , then, may well be regarded as a (relational) feature of my behaviour alone.

One crucial difference between Dennett and Mölder here is that instead of a distinction between pleonastic and natural properties, Dennett employs Reichenbach's distinction between *abstracta* and *illata*. *Illata* are like Mölder's natural properties and both agree that mental states cannot be understood in terms of them. *Abstracta* are pleonastic properties, but they are pleonastic properties of a very specific kind. *Abstracta* serve an explanatory or predictive purpose; their ascription allows us to access or disclose patterns in reality (behavioural patterns in the case of mental states) that would otherwise not be accessible or disclosed.

But here another difference between the views of Mölder and Dennett surfaces. In the description of mental state ascription Mölder gives in Section 2 of Chapter 5, it becomes clear that mental state ascription, according to him, is mainly retrospective. It is meant to make sense of behaviour, but not necessarily to predict behaviour (whether explanation falls under sense-making is unclear). Mölder: "not all ascriptions are made with the purpose of predicting behaviour and not all ascriptions need generate predictions." (176)

But here we need to precise. The second part of this claim might seem to be a repetition of the first but it is not. We can accept that mental state ascription does not always serve the purpose of prediction while still sticking to the claim that all ascriptions do generate predictions. In fact on all theories of the meaning of mental state ascription that hinge on the idea that the meaning of ascriptions hinges on the counterfactuals that it supports—which I take to be most current theories—all mental state ascriptions *should* involve predictions. When I ascribe the belief that Amsterdam is the capital of the Netherlands to John, this implies that I make certain predictions about John's behaviour in a range of hypothetical situations (he will utter the word 'Amsterdam' when asked what the capital of the Netherlands is, he will point to Amsterdam when he is asked to point out the capital of the Netherlands

on a map, etc.). Without making such predictions I would not be ascribing the relevant belief state.

The point is that whatever use we make of mental state ascription must, on a pragmatic, Dennettian view, be accounted for in terms of the counterfactuals and concurrent predictions that are connected with the ascription of mental states. This even goes for retrospective sense making: imagining a relevant range of counterfactual situations and concurrent predictions can actually perfectly well help to make sense of a piece of behaviour that is already performed. 'He did it out of jealousy' means 'his actions sprang from a state that would have produced behaviour  $B_1 \dots B_x$  under circumstances  $C_1 \dots C_x$ '. Two things are important to note: (1) Whether or not he was jealous really is an interpreter-independent fact—whether or not the relevant counterfactuals implied in the practices of our mental state attribution are true (which may well be something we will never know) is not up to the interpreter. (2) The use that we make of mental state ascription does hinge on prediction, even though this need not be overtly so.

Here it may be asked whether the set of counterfactuals associated with the ascription of a certain mental state in a certain situation does not in fact constitute a surplus meaning that exceeds our practices of mental state ascription. I think this would be the wrong conclusion. For the set of counterfactuals that are implied by the application of a certain mental state term is fixed by the usage of that term in our actual practices. Nothing would explain further counterfactuals that are not associated with our use of mental state terms, nor is there any explanatory purpose that would be served by invoking such extra counterfactuals.

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. On this view, then, which is reminiscent of the Rylean strands in Dennett's Intentional Stance theory but which may just as well be compared to Lynne Baker's practical realism, there is no surplus meaning of mental state ascriptions over and above our practices of mental state attribution. And without such surplus meaning, there is simply *no sense* to be made of the notion of *pretending* to ascribe mental states; the idea of "as if" ascription is empty. The point here is akin to Ryle's (1949, Chapter 1) criticism of the idea that intelligent behaviour is behaviour that is caused by inner mental states: in order to be in a position to invoke inner causes for intelligent behaviour, we need to be able to identify intelligent behaviour as such and distinguish it from, say, sneezes and hiccups. But once we are in that position, invoking inner causes is completely redundant, non explanatory and hence meaningless as

a means of distinguishing intelligent from other behaviour, for we already have that distinction. The point here is identical: there is no use for, and hence meaning to be given to “as if” ascription. Pure ascriptivism does not allow for an “as if” interpretation.

Paradoxically, then, given that an “as if” interpretation depicts interpretivism as a radical theory, a stricter pure ascriptivism can be defended as a less radical option than the less strict intermediate version.

#### 4. Pure ascriptivism and the metaphysics of mental explanation

One asset of Mölder’s ascriptivism is that it does not position itself as a view that completely rules out other positions in the philosophy of mind, such as functionalism, dispositionalism or practical realism. I very much think that this is the right way to go. The version of pure ascriptivism described in the previous section also allows for the compatibility of interpretivism and other positions in the philosophy of mind. In this semi-final section I want to compare Mölder’s views on this compatibility with the views implied by pure ascriptivism (in the form outlined above). I will argue in favour of pure ascriptivism.

As a means of introducing his ascription theory, Mölder first discusses traditional views on the mind-body relation (or the mind-brain relation) such as functionalism, dispositional accounts and practical realism. He argues that these views “are compatible with the interpretivist approach in the sense that it is possible to plug an interpretivist component in these views ... [O]ne can argue ... that instantiating a functional role, having an appropriate disposition and the truth of relevant counterfactuals are a matter of interpretation.” (154)

Given what is said about the interpreter-independent status of intentional interpretability of behaviour in the previous section, it should be clear that this option is not the one I would favour. In my view, the “relevant counterfactuals” *are* what we ascribe when we ascribe mental states and the interpretability in intentional terms of behaviour consists in the fact that highlighting those counterfactuals serves practical purposes, such as predicting further behaviour. The idea that these counterfactuals themselves would be dependent upon interpretation seems to point clearly to the idea that the meaning of mental state ascription is *not* exhausted by exploiting existing counterfactuals. This is intermediate interpretivism, not pure ascriptivism.

Pure ascriptivists can agree with Mölder, however, that ascriptivism is not incompatible per se with functionalism, dispositionalism or a counterfactual approach to the nature of mental states and mental state ascription. Rather than inserting an element of interpretation in functionalism, dispositionalism and counterfactual approaches, however, I would suggest assim-

ilation in the case of dispositionalism and counterfactual approaches on the one hand and a division of theoretical labour in the case of functionalism. The version of pure ascriptivism outlined and defended above incorporates dispositionalism and counterfactual approaches. There are differences in emphasis but the metaphysical commitments of these theories to overlap. This is not the case with functionalism. Functionalism makes claims about the nature of mental states (they are causal role states) and hence makes certain demands on the implementation bases of our mental states. Interpretivism does not necessarily share these metaphysical commitments. However, if functionalism turns out to be true, the instantiation of causal role states would be what explains the fruitfulness of intentional interpretations of behaviour. It would explain why certain counterfactuals hold and why people have the behavioural dispositions they have.

So pure ascriptivism is compatible with functionalism. But it is not committed to it. It is simply committed to whatever explains the truth of the sets of counterfactuals that we postulate in ascribing mental states to each other. This may be a functionalist explanation, but it might just as well be a non-mental neural explanation. As long as there is something that explains why the counterfactuals hold. Pure ascriptivism is merely committed to the claim *that* when it is true that person *P* is in mental state *M*, a specific set of counterfactuals holds, but it can accept any reasonable explanation of the fact that these counterfactuals hold.

This division of labour view on the compatibility of interpretivism and functionalism seems to me to more transparent than allowing for an element of interpretation in assigning causal roles. The interesting thing is that it seems to fit extremely well with Mölder's own ideas on mental causation from an interpretivist point of view. Here's a very telling quote (230-231)

Giving causal explanations in mental terms need not imply token identity or other kinds of reductionism. But it is still the case that when an action occurs, some physical process must take place. Its explanation in folk-psychological terms does not point out the relata of this process, but it does not follow that there are no efficacious entities involved. Similarly, when we make a successful prediction of an action, which is confirmed by the further course of events, there must again be something causally efficacious occurring or being instantiated. ... The reason why it was hard to explain how mental entities can be causes is that this task was previously taken to consist in explaining how mental entities can have causal efficacy, which indeed they cannot have.

I could not agree more (see Slors 2007).

## 5. Concluding remark

This brings me to a final remark that needs to be made. It is quite possible that I have exaggerated the differences between Mölder's views and my own for the sake of argument. If I have, however, I think the distinction between pure and intermediate ascriptivism is the source of confusion—it had better been given up.

## Bibliography

- Dennett, D. (1987). *The Intentional Stance*, MIT Press, Cambridge, MA.
- Dennett, D. (1991). Real patterns, *The Journal of Philosophy* **88**: 27–51.
- McCulloch, G. (1990). Dennett's little grains of salt, *The Philosophical Quarterly* **40**: 1–12.
- Mölder, B. (2010). *Mind Ascribed: An Elaboration and Defence of Interpretivism*, John Benjamins, Amsterdam.
- Ravenscroft, I. (2005). *Philosophy of Mind: A Beginners Guide*, Oxford University Press, New York.
- Ryle, G. (1949). *The Concept of Mind*, Hutchinson, London.
- Slors, M. (1996). Why Dennett cannot explain what it is to adopt the intentional stance, *Philosophical Quarterly* **46**: 93–98.
- Slors, M. (2007). Intentional systems theory, mental causation and empathic resonance, *Erkenntnis* **67**: 321–336.