EUJAP | VOL. 5 | No. 2 | 2009
ORIGINAL SCIENTIFIC PAPER
UDK: 159.9.01:165.7
165.7:159.9.01
165.21

FOLK PSYCHOLOGY IS NOT A METAREPRESENTATIONAL DEVICE

TAMÁS DEMETER

Max-Planck-Institut für Wissenschaftsgeschichte, Berlin

ABSTRACT

Here I challenge the philosophical consensus that we use folk psychology for the purposes of metarepresentation. The paper intends to show that folk psychology should not be conceived on par with fact-stating discourses in spite of what its surface semantics may suggest. I argue that folk-psychological discourse is organised in a way and has conceptual characteristics such that it cannot fulfill a fact-stating function. To support this claim I develop an open question argument for psychological interpretations, and I draw attention to the central role of rationality, the conceptual connections, and the essential evaluative content inherent in folk psychological ascriptions. As a conclusion I propose that a fictionalist account of the discourse would fit its characteristics better than a factualist-realist interpretation.

Key words: folk psychology, mental fictionalism, rationality, metarepresentation

1. Introduction

There seems to be widespread agreement that the function of folk psychology is predominantly social: that is, to facilitate navigation in the interactive world by representing the internal mechanisms causally responsible for the production of our fellows' behaviour. On the consensual view the value of the discourse hangs on its capacity to represent correctly the mental states of others that allows for the co-ordination of social behaviour. As these states are themselves frequently representational, e.g. beliefs and desires, the utility of folk psychology springs from the epistemic value of the *metarepresentations* it can deliver.

We have several discourses the semantic surface of which suggests that their sentences represent how things are, but if looked at more closely they turn out to be factually defective, i.e. not representing what they *prima facie* seem to represent. The most obvious cases to argue this point are moral and aesthetic talk ascribing evaluative properties whose re-

lation to descriptive ones is problematic to spell out. A similar problem emerges when considering rationality, another evaluative property central in our psychological interpretations. The rationality of an action does not follow from its behavioural description. It is relative to the agent's beliefs and desires, and can thus be judged only by ascribing various intentional states whose fit into a purely naturalistic description of the world is also problematic as the latter does not mention intentional states, i.e. mental states representing something else.

In this paper I offer four reasons for the thesis that folk psychology is a factually defective discourse, and therefore it cannot serve the purposes of metarepresentation either. My reasons derive from how folk psychological concepts are organised and used, and if I am right, they prevent the discourse relying on them from being genuinely fact-stating. So, the lesson I offer here is that in folk psychology we cannot represent agents' internal states despite the discourse consisting of propositions the semantic surface of which suggests that they are fact-stating.

2. Rationality

It is widely acknowledged that *rationality* plays a central role in our everyday understanding of human action, and that this is the feature that distinguishes it from the scientific treatment of behaviour. As Simon Blackburn puts it:

> In one way or another the fact that we need to theorise under a 'principle of rationality', or to see a proper point in people's doings in order to understand them, marks off this kind of theorising from anything found in the natural sciences. (Blackburn 1995, 282)

We understand others on the assumption that they are rational agents, meaning that their beliefs and desires form a relatively coherent system, and that they typically choose a course of action that follows from this system. Rationality thus operates on the set of beliefs, desires and other mental states we ascribe to agents, and given these we can predict and explain behaviour reliably, just because agents typically do what they rationally ought to do. This is what discriminates our understanding of the inhabitants of the social world from that of other objects around us.

Rationality is *normative*: given the agent's mental states, our reliance on it tells us what he ought to think or do, independently of what he actually does. He may not ever do what he ought: this would not matter at all in relation to what it is rational for him to do. The assumption of rationality is an assumption of conformity to a norm or a set of norms. Psychological explanation and prediction are made possible only by relying on this assumption. Without rationality there is no way of gaining cognitive access to an agent's behaviour as based upon his attitudes: inferences from attitude ascriptions to

behaviour remain ungrounded if a significant degree of rationality is not ascribed to the agents at the same time. Without presupposing rationality on the agent's part, having beliefs or desires does not establish any conclusion concerning his actions. Thus, lacking rationality, attitudes could not serve as *reasons* explaining the agents' behaviour.

Therefore, as Davidson (1980b) likes to say, rationality is a *constitutive principle* of psychological interpretation to which propositional attitude ascriptions must conform in order to be acceptable as good explanations of behaviour. It amounts to saying that psychological explanations *must rationalise* the event to be explained. If someone believes that the sky is blue, he must also believe that it is coloured – if he does not believe that, we need to reinterpret his former belief too, otherwise his system of beliefs is incoherent and the agent cannot be interpreted. If someone wants to start the car, and believes that this is to be done by turning the key, then he must form the intention of turning the key – otherwise the system of his propositional attitudes is irrational and the agent's behaviour cannot be interpreted.

Psychological interpretation is, therefore, *biased* in a peculiar way. As Davidson (1980c, 237) aptly emphasises, it is a business that commits us to finding a great deal of consistency in the agent's "pattern of behaviour". Rationality thus means both a constraint and a bias: psychological interpretations cannot do their job without it, and given that we, as a matter of course, understand one another through the looking glass of folk-psychological terms, we also have a natural tendency to look for rationality in behaviour.

Observing the principle of rationality in our interpretations yields *teleological* representations of behaviour (see Velleman 2003). Their teleological significance arises from beliefs and desires ascribed to the agent which specifies his aims and purposes. Psychological interpretation is thus a rationalising narrative which concludes in the action itself. Viewed from this angle, pieces of behaviour acquire meaning by obtaining some teleological significance in the network of these interconnections. This significance can be specified with reference to the conclusions towards which narratives gravitate. This is the framework in which psychological interpretations make sense of behaviour: with an attention to purpose and a commitment to finding coherence. This is the background against which the meaning and rationality of behaviour can be shown – and life can be seen as meaningful.

For these reasons, rationality – and psychological interpretation along with it – do not fit unproblematically into the causal network of events in the natural world. Causal laws are descriptive, they describe connections between events, and not some ideal order to which events must converge. By contrast, rationality is not a descriptive natural law or empirical generalisation to which agents conform, but a norm which behaviour can violate with the consequence, of course, that it turns out to be inaccessible by means of psychological interpretation. The teleological orientation of psychologi-

cal narratives has no parallel in descriptive causal histories: causes and effects can follow one another in an endless chain without ever concluding, so causal descriptions cannot lend teleological significance to events. Therefore rationality as its constitutive principle distinguishes psychological from physical discourse whose constitutive principles are causality, spatio-temporality, and measurement (see Kim 2003, 119).

Now to maintain the commitment that folk psychology is a discourse capable of representing the mental states as part of the causal structure of the world, one should bridge this gap between constitutive principles. However, I do not think it possible. I will consider three proposals and I draw some conclusions from them concerning the nature of psychological interpretation.

3. A non-starter

Robert Brandom rejects the idea of reducing normative to something non-normative in general, so he rejects naturalising rationality as well. The basis of Brandom's rejection is Wittgenstein's regress argument: a norm cannot be reduced to regularities of behaviour, as any piece of future behaviour can be fitted with the norm under some interpretation, because the norm itself does not tell the conditions of its correct interpretation. Brandom (1994, 21) concludes from this that a "pragmatist conception of norms", as he calls it, is required which is based on "a notion of primitive correctness of performance *implicit* in *practice*." This arises from the fact that practices can be carried out right or wrong. Norms are thus implicit in practice, and as rationality and intentionality are explained in terms of practice, norms are indispensable in their explanation too. It is, as Brandom says, "norms all the way down" (1994, 625).

The distinction between normative and non-normative facts is itself drawn in the normative vocabulary of our linguistic "scorekeeping" practices within which we keep track of the "commitments" and "entitlements" of those taking part of our discursive practices. Rationality is thus explained on three levels by reference to (1) norms implicit in our thinking and language, i.e. in practices governing the application of concepts, that (2) form the basis of our implicit attitude attributions, and which (3) we express in our attitude ascriptions explicitly (see 1994, 636ff). As the relevant social practices are public, they account for the objectivity of norms, which is crucial for understanding them adequately (1994, 63). Otherwise they could not count as standards of right and wrong. And this crucial point is missing from dispositionalist accounts that reduce norms to dispositions to behave in specific ways under specific conditions, i.e. to something non-objective and non-normative.

The problem with Brandom's account is that it does not explain norms; instead he takes them to be *primitive*. The obvious problem this account faces is this: How could norms, taken as primitive, be accommodated in the causal order of the natural world?

This poses a problem for Brandom because he holds the majority view that science is a descriptive enterprise, and it "will never run across *commitments* in its cataloging of the furniture of the world" (1994, 626). But if norms are not part of the natural world in their own right, how can they be causes of e.g. behaviour? Brandom's answer is that norms themselves are not causally efficacious, instead

> [w]hat is causally efficacious is our practical taking and treating ourselves and each other as having commitments (acknowledging and attributing commitments) – just as what is causally efficacious is umpires and players dealing with each other in a way that can be described as taking the score to include so many strikes and outs. (Brandom 1994, 626)

This account leads to a serious problem. If causal efficacy belongs to our "practical taking and treating" then why should we keep the talk about norms, why should not we just be contented with the talk about mutual attributions, given that only these are required for the explanation of behaviour? In this case presupposing norms in themselves, as implicit in practice, seems *superfluous*.

For our explanations we need only (2) and (3) from the above levels. Brandom, however, thinks that (1), i.e. norms implicit in practice, is the basis of our mutual attributions. He thinks that because he refuses to explain these attributions reductively in non-normative terms, and his reason for refusal is that he thinks reductive explanations threaten objectivity, accounting for which is an indispensable part of any explanation of norms. Now Brandom's view is *incoherent* as it stands. If norms implicit in practice are the basis of our attributions, it does not make sense to say that they are not causally efficacious and only their attributions are. Given that our attributions are based on norms implicit in practice, the causal efficacy of our attributions derives also from the norms themselves. If not, then there is no proper role for the norms to play in Brandom's account, and thus they are useless; and if yes, if they are the basis of our attributions, then it is hard to understand this relation if it is not causal, and therefore Brandom has no theory of how norms and rationality can be placed in the natural world.

4. An open question

Brandom articulates his position by contrasting his interpretationist account with that of Daniel Dennett (see Demeter 2009a). Dennett grounds his notion of rationality on the basis of evolutionary game theory, more specifically on the *optimal foraging theory* (see e.g. Dennett 1998). According to this theory animals should optimize the net amount of energy gained in a given period of time, i.e. to maximize the energy gained and minimize the energy invested. This is a strategy to follow under evolutionary pres-

sure. For example, if an animal is not disposed to leave a source of food if depleted, i.e. if it is not disposed to invest energy under certain circumstances, then its fitness suffers, or more concretely: the animal goes extinct eventually. And conversely, if it is eager to leave good sources of food without being forced by the circumstances, then its fitness suffers again, in this case by jeopardizing its own security. The optimal strategy, the pattern of behaviour to be followed here and in similar cases, can be mathematically modelled, which gives a list of the kinds of behaviour that can lead to evolutionary advantage for those following them. Due to the resulting advantage, these kinds of behaviour proliferate in the population, because those not following them simply go extinct, or have less offspring (and thus the behaviour goes extinct in the long run). This analysis can be extended to practical decision making in general: following highly complex calculations, this model can eventually give the "ideal order" to be approximated by the patterns of human behaviour too, that is the norms of rationality. And given that following these norms is beneficial, conformity proliferates, and most of the time agents will do what they rationally ought to do.

Prima facie, this approach naturalizes rationality by showing its proper home in the selective processes of evolution. Indeed, it makes plausible that the rationality of *some* kinds of behaviour may be explained this way. However, it seems all too optimistic that this explanation can be generalized so as to encompass 'rationality' in its really intriguing uses, that is, in the explanation of *human social behaviour*. The problem is that in these cases, where rationality is most interesting to us, we face *psychologically* complex situations where, given Dennett's account of how we attribute intentional states, the evolutionary story must fail in this context.

For Dennett (1991), when giving a folk-psychological interpretation we look at the agent's behaviour from the *intentional stance*. From this stance we calculate, on the basis of real patterns of behaviour, abstract entities referred to in psychological descriptions. By giving a folk-psychological description we pick out a behavioural pattern, and the description can be true if the agent produces relevant behavioural patterns. So, folk psychology is ultimately a fact-stating discourse because the truth of its interpretation-bound descriptions is rooted in behavioural patterns that are independent of the interpreter. These patterns are caused by the agent's internal, e.g. neurological, mechanisms which are independent of interpretation and so are the patterns themselves. Therefore, folk psychology gives interpretation-bound descriptions on the personal level that have truth-makers on the subpersonal level. Although folk-psychological descriptions do not map onto the agent's internal mechanisms isomorphically, personal-level predictions and explanations are still causal because they are inferred from real behavioural patterns caused by internal mechanisms.

What is rational to do or to believe in social situations depends, at least partly, on interlocking systems of values, beliefs, desires, etc. that is, on the interpretation of other parties to the situation. These attitudes are attributed to the agents on the basis of their

behaviour, thereby discerning real behavioural patterns in it. However, as Dennett willingly acknowledges, behaviour allows for radically different interpretations and predictions of an agent, interpretation picks out patterns of behaviour and there is no "deeper fact of the matter" that could decide which one of the possible and incompatible interpretations is true. Nothing intrinsic in the agents' behaviour determines which pattern is to be picked out:

> I see that there could be two different systems of belief attribution to an individual which differed substantially in what they attributed – even in yielding substantially different predictions of the individual's future behavior – and yet where no deeper fact of the matter could establish that one was a description of the individual's real beliefs and the other not. In other words, there could be two different, but equally real patterns discernible in the noisy world. The rival theorists would not even agree on which parts of the world were pattern and which noise, and yet nothing deeper would settle the issue. The choice of a pattern would indeed be up to the observer, a matter to be decided on idiosyncratic pragmatic grounds. (Dennett 1991, 49)

Given all this there is no way of specifying the "ideal order" to which behaviour should converge in any given social situation, because what the situation *is* depends on our interpretations. There are no social situations independent of interpretation: to the extent they are independent, they are not social (but behavioural, neural, etc.).

And the same applies to the outcome: our judgement on the rationality of the agents' behaviour depends on our interpretations. The course of behaviour eventually followed in a social situation will be rational under some interpretations, and irrational under some other. Now it seems that evolutionary game theory can be useful in explaining rationality where "facts of the matter" determine an optimal strategy, but in social situations, as they depend essentially on interpretations which are not made true or false by relevant facts, this cannot be the case. The optimal strategy to be followed in a social situation inevitably depends on interpretation, and as there is no optimal (uniquely true) interpretation there is no optimal strategy either. The benefits of Dennett's account are thus dubious. We might gain the possibility of calling 'rational' animals of the kind that we do not even think of as rational, but we are not a step closer to explaining what rationality is where it plays its proper part, namely in the social world, in the space of reasons. So Dennett's account of rationality fails, but its failure is full of lessons.

Even if ideal epistemic access is granted, it is still possible to give coherent interpretations of an agent's behaviour and internal states with radically different sets of propositional attitudes – and it remains an *open question* which one of them is true. And this is due not to less than ideal access to the relevant evidence, but to the *lack of facts independent of the psychological discourse*. The case is not that there are potentially relevant but verification-transcendent facts underlying folk-psychological discourse that

are for some reason inaccessible (for example, because of the limits of human experience, or the theory-ladenness of experience). I would rather say, *due to its nature* folk-psychological discourse is incapable of stating facts about an objective – i.e. discourse-independent – order which can be treated as mental reality. Saying that there can be disagreement about which parts of the world belong to patterns and which are noise seem to suggest that behavioural patterns cannot be identified without psychological concepts. So, without the threat of circularity, they cannot be used for grounding them.

The evidences relevant for psychological interpretation, mostly behaviour that is, are as a matter of course seen through the concepts of folk-psychological discourse. Folk psychology organises evidence into a rational and coherent system of propositional attitudes – and it cannot do otherwise. One cannot step back from the intentional stance and weigh evidence independently of it – that would entail not giving psychological interpretation at all. One can give alternative interpretations but with them evidence changes as well: some parts of the world cease to be noise and begin to make sense by fitting into a pattern, other parts become noisy. But even then, alternative interpretations remain within the framework set by the intentional stance or the constitutive principles of psychological interpretation. The case is thus not that we have different theories organising and weighing pieces of evidence differently, rather the same "theory" (i.e. folk psychology) allows for creating evidence and interpretations in divergent and incompatible ways.

This arises from the lack of independent criteria for determining the epistemic value of psychological interpretations. The extent to which an interpretation is precise, satisfactory, etc. can be judged only by relying on the psychological background which the interpretation presupposes. Classifying behaviour – which bodily movement counts as an action and is relevant to which mental states – belongs to the realm of folk psychology. Behavioural evidence counting in the justification of psychological interpretations is already filled with folk-psychological content and interpretation, and it cannot be otherwise as in this context evidence would not count as evidence without it. In order to use some bodily movement as evidence in an interpretation, I need to specify its meaning and significance; and *vice versa*, by ascribing mental states to an agent, I give meaning to some of his bodily movements. There are no discourse-independent relevant facts in the business of folk-psychological interpretation. It seems then that, on an interpretationist account, the way the discourse works, its nature, leads to its antirealist interpretation.

To draw the conclusion of this section: Dennett's account is *unstable*. It should either give up the idea that there are components of the world independent of folk-psychological discourse that are described in this discourse, or that folk-psychological interpretation is based on the constitutive norms of rationality. I think this latter option should be avoided, as Blackburn's dictum quoted at the beginning of this section seems convincing. Folk psychology allows us to conceive of ourselves and others as *persons*,

a special group of agents looking at the world from a subjective perspective and being responsible for their behaviour. It provides the conceptual resources to understand ourselves and fellow humans in a special way, quite differently from other things in the world, and responds to the needs of social interaction and not of disinterested cognition.

If the implications of Dennett's interpretationism are accepted, then the idea that folk psychology is a fact-stating discourse should be given up. This conclusion is almost inherent in interpretationist accounts, but not drawn. The insight that folk psychology, due to the constitutive role of rationality, is a discourse fundamentally different in kind from discourses about other regions of the world suggests it. But even if this insight is readily available, and the conclusion is at hand to draw, it is overshadowed by the conviction that despite being interpretative, folk psychology is still a descriptive, explanatory and predictive device.

5. Conceptual connections

As an interpretation must rationalise the agent's behaviour, it must portray its causal background as if it was a logically coherent system of attitudes. This threatens the very idea of folk psychology being causally predictive and explanatory, because the *relata* of folk-psychological explanations are not conceptually independent entities.

"To explain an event is to provide some information about its causal history" – as Lewis's (1986, 217) sensible dictum has it. As is commonly held, the epistemic value of folk psychology derives from its capacity to describe causal connections between mental states and behaviour. The entity mentioned in the *explanans* can only be relevant for the one in the *explanandum* if, as Hume's famous criterion has it, they are independent entities. Otherwise there can be no causal connection between them:

> All those objects, of which we call the one *cause* and the other *effect*, consider'd in themselves, are as distinct and separate from each other, as any two things in nature, nor can we ever, by the most accurate survey of them, infer the existence of the one from that of the other. (Hume 2002, 2.3.1.16)

In folk-psychological explanations this criterion is typically not met, as more often than not there are conceptual connections between *explanans* and *explanandum*, that are knowable *a priori* while "no connexions among distinct existences are ever discoverable by human understanding." (Hume 2002, Appendix, 20) So, if a connection is discovered *a priori*, then it cannot hold between two entities existing independently. As psychological explanations are based on existing conceptual connections, they do not describe the relation of independent entities, they are therefore not causal explanations and cannot serve as the basis of causal predictions.

As Norman Malcolm (1984, 88) says, even if the inferences from attitudes to actions are not entirely *a priori*, the conceptual connection "is strong enough to rule out the possibility of there being a merely contingent connection." Folk-psychological interpretations are not supported by empirical generalisations but by conceptual connections. Correspondingly, putative psychological explanations do not state facts: as there are conceptual connections between mental states and actions caused by them, there are no logically independent entities whose relation could make true the explanation relating them. And as Davidson explains, this cannot be otherwise as long as we are in the business of psychological interpretation:

> these obvious logical relations amongst beliefs; amongst beliefs, desires, and intentions; between beliefs and the world, make beliefs the beliefs they are; therefore they cannot in general lose these relations and remain the same beliefs. Such relations are *constitutive* of the propositional attitudes. (Davidson 1985, 196)

The logical interconnections among propositional attitudes, among mental states and actions are peculiar to the mental, and it has no analogy in the physical world. There are no logical or conceptual connections, for instance, between lightning and fire. The truth of 'lightning causes fire' is grounded in, and therefore testable by, experience and does not state a conceptual connection immune to experience. At first, it seems, one might argue that there are conceptual connections here, as nothing can be lightning that cannot cause fire. This is sophistry. The meaning of lightning does not entail the disposition to cause fire. This disposition derives from the intrinsic properties of lightning. The definition of lightning might run: 'a high voltage electrical discharge caused by atmospheric phenomena', whose properties explain its disposition to cause fire. But to say that 'a high voltage electrical discharge caused the lightning' would not count as a causal explanation as 'electric discharge' is just part of the concept of lightning.

The approach advertised here is sometimes argued against thus: 'The sun caused sunburn' is a well-formed causal sentence, and it is true by definition that sunburn is caused by the sun. But this smells like sophistry again. This is the reverse case of lightning and fire: it seems that there is a conceptual connection here, while in fact there is not. When we talk about sunburn we talk about a certain inflammatory condition of the skin, which is a natural kind *and* hint at its causal prehistory. There is no conceptual connection between the condition and its alleged cause. What makes it seem like one is that we are attentive to a subclass of these conditions with a specific causal history, for example because of their frequency or the initial act of dubbing, and this attention is reflected in the name we have given to this phenomenon. But the condition itself can result from other causes, for instance in a solarium, and thus the condition's causal history is an entirely contingent feature. Therefore it is as irrelevant to the real nature of the phenomenon as its causal history is to a piece of gold: it is gold no matter how it came about. And it would be irrelevant even if, due to some historic accident,

we had chosen to indicate in their names the different causal histories of pieces coming from mines and from the laboratories of alchemists. And the case is just the same with 'sunburn'.

One could think that this argument could be applied to save the causal character of psychological explanations, if it was shown that conceptual connections in them are similarly illusory. But the prospects are not good for an attempt like this. While the extension of 'sunburn' does not contain anything of conceptual nature (only the condition itself plus its causal history), the extension of psychological concepts, by contrast, contains essentially conceptual entities (like beliefs, desires, etc.). More often than not, psychological propositions imply entities whose conceptual connections, as we saw in Davidson above, are constitutive of their identity. Therefore it is not true that mental events can be logically independent of one another, and that they do not presuppose anything conceptually (see also Davidson 1987, 59).

Due to its constitutive principle and the conceptual connections in it, the logic of folk-psychological discourse is of a different kind from that of other discourses about the inanimate world in which causality plays the role of a constitutive principle. It is just one step further to argue that causality is therefore alien to the logic of the discourse about the mental. Melden incites suspicion thus:

> Where we are concerned with causal explanations, with events of which the happenings in question are effects in accordance with some law of causality, to that extent we are not concerned with human actions at all but, at best, with bodily movements or happenings; and where we are concerned with explanations of human action, there causal factors and causal laws in the sense in which, for example, these terms are employed in the biological sciences are wholly irrelevant to the understanding we seek. The reason is simple, namely, the radically different logical characteristics of the two bodies of discourse we employ in these distinct cases – the different concepts which are applicable to these different orders of inquiry. (Melden 1961, 184)

The most influential critique of this view comes from Davidson (1980a). For Davidson folk-psychological explanations are as causal as physical explanations. Davidson's argument is based on the insight that causal relations are *extensional*, relating events, while explanations are *intensional*, relating descriptions of events. We can talk about causal explanations where the events mentioned in the descriptions are causally connected. Causal connection is possible where the relation of events can be subsumed under causal laws, but this cannot be done in the psychological idiom, only in the language of physics. Nevertheless, a psychological and a physical description can be the description of one and the same event, because events are spatio-temporal particulars that can be described in various ways. Psychological explanations can thus be causal because the events mentioned in them can be causally connected, though this is not

transparent if seen through psychological descriptions. Seeing causal connections requires a physical description of events.

Davidson's argument, albeit ingenious, cannot save the causal interpretation of psychological explanations. The problem remains because the physical idiom is fitted to talk about conceptually independent events and to mention them in explanations. But we cannot do this in the psychological idiom because we cannot represent events as conceptually independent. The semantic surface leads us astray if we understand 'because' as a causal connective. The physical description of the same situation, however, portrays an *a priori* inaccessible connection of two conceptually entirely separate events: e.g. the connection between some activity in the agent's neural network and then a series of bodily movements. This suggests that while we can represent a relation by a physical description as a causal connection between two distinct events, we cannot do the same by a psychological one. The two kinds of description carve events in qualitatively different ways, portray them from incompatible perspectives.

The consequence is not only that the two discourses, physical and psychological, are mutually irreducible to one another, as Davidson is happy to acknowledge. It is also that it makes no sense to say that the same event can be described in a psychological and a physical vocabulary. This should be hardly surprising: if one admits, as Davidson does, that the two discourses are organised by incompatible constitutive principles and logically differently, then there will be no mapping between them to substantiate the claim that psychological and physical descriptions can count as descriptions of the same event. As quoted above, Davidson considers logical relations among mental events as constitutive of their identity. Nothing similar can be found among physical phenomena; it makes no sense to say that they have logical connections. As *identity conditions* formulated in psychological or physical vocabulary are of a different nature, there are no pairs of identity conditions, put forward in psychological and physical terms respectively, that could identify the same event.

This can take us to the idea that psychological concepts are special in their way. It seems to me promising to argue that they play a *constitutive*, as opposed to a merely descriptive, role in relation to the phenomena that they seem to describe if their declarative semantics is taken at face value. This explains why 'because' cannot be a causal explanatory connective in psychological contexts. These are similar to social contexts as Peter Winch explains them:

> The conceptions according to which we normally think of social events are logically incompatible with the concepts belonging to scientific explanation. An important part of the argument was that the former conceptions enter into social life itself and not merely into the observer's description of it. (Winch 2008, 89)

Social phenomena *qua* phenomena cannot be detached from the concepts possessed by those taking part in them. This is a property of social phenomena that has no counterpart in the natural world, as "the concept of gravity does not belong essentially to the behaviour of a falling apple ...: it belongs rather to the physicist's *explanation* of the apple's behaviour" (Winch 2008, 119). More recently, Ian Hacking seems to concur with Winch while discussing mental phenomena:

> I am urging caution in projecting results of trauma produced by impersonal conditions onto trauma produced by human actions. This is not because some different kind of memory is involved, but because of a logical difference between the events remembered. We describe earthquakes, but it makes little sense to talk about an earthquake under a description. It is just an earthquake. (Hacking 1995, 248)

Earthquakes, physiological and neural events, etc. *as* phenomena are untouched by the way they are described, by the concepts that are used in their descriptions. But by introducing new psychological concepts we introduce, as Hacking says, new ways of being a person, new ways of acting intentionally, of having mental states – we create, as it were, new mental phenomena. If a given psychological concept is out of use, then it is not possible to think and act intentionally by this concept (Hacking, 1995, 239) – it is thus not only the possibility to talk about it, but the phenomenon itself that is missing.

Now, in descriptive discourses concepts belong to the descriptions only but not to the phenomena themselves as in our discourses on social and psychological phenomena. One could argue (see e.g. Kusch, 1997 and 1999) that mental phenomena are a subclass of social phenomena in the sense that they presuppose the *institution* of folk psychology, and it makes sense to talk about the mental only against the background of this institution. There are thus no psychological phenomena independent of psychological concepts. By choosing different concepts to interpret an agent's behaviour the interpreter puts emphases on different aspects of the agent's circumstances and behaviour. This is what Dennett seems to suggest, too, in emphasising that different systems of attitude ascriptions reveal different patterns in an agent's behaviour. And Davidson, too, when he points out that psychological interpretation entails the commitment that there is a significant degree of coherence and rationality in the behaviour of the agents to be interpreted. The patterns of behaviour revealed by interpretation depend for their very existence on the concepts with which we try to make sense of it. Deploying different concepts results in different patterns to be revealed, and no evidence can favour one set of concepts over another.

Furthermore, for our interpretations we can rely only on evidence that is tailor-made by folk-psychological concepts. Without relying on the conceptual resources folk psychology provides, stimuli and responses are unstructured; they can be organised precisely by folk-psychological concepts. The perceptual states relevant in the formation

of a belief cannot be specified without knowing which belief it is that they are relevant for – there are conceptual connections between the contents of perceptual states and the propositional attitudes that rely on them. Similarly, the way we identify an agent's behaviour conceptually depends on the psychological states we ascribe to him or her. Owing to these conceptual connections no priority can be assigned to perceptual states and behaviour in an explanation of the origin of psychological terms.

There is thus no *independent evidence,* uninfluenced by the discourse's conceptual apparatus, against which to test the truth of psychological narratives, as there are no relevant phenomena independent of folk-psychological interpretation. The criteria of what counts as relevant stimuli and behaviour are set by the actual interpretation itself. Therefore, a difference in the concepts used for interpretation is a difference in psychological phenomena at the same time. The case is not that different interpretations describe the same phenomena differently, rather it is that with different concepts we talk about different phenomena too. It is not that interpretation-independent evidence underdetermines the acceptance of psychological propositions, rather it is that the range of possible items of evidence is itself conceptual in nature: psychological concepts are not tools for descriptions, but constituents of psychological phenomena themselves.

In this sense folk psychology creates its own phenomena. If we acknowledge that there is no way of choosing from among incompatible interpretations on the basis of independent items of evidence like pieces of behaviour or neurological information. As different interpretations reveal different behavioural patterns, and it seems to suggest that it is interpretations themselves that bring to light the evidence relevant from their own angle – and not *vice versa*.

6. Evaluation

The possibility of moral evaluation is based on psychological interpretation. Hume rightly says:

> Tis evident, that when we praise any actions, we regard only the motives that produc'd them, and consider the actions as signs or indications of certain principles in the mind and temper. The external performance has no merit. We must look within to find the moral quality. This we cannot do directly; and therefore fix our attention on actions, as on external signs. But these actions are still consider'd as signs; and the ultimate object of our praise and approbation is the motive, that produc'd them. (Hume 2002, 3.2.1.2)

Only by the conceptual resources of folk psychology can agents be represented as agents moved by *evaluable motives*. This chance is not given by the physical – or in general: scientific – description of agents. In this idiom one can describe the causal

chain resulting in behaviour, but this will be an impersonal description representing the agent as if events were happening to him rather than him doing something. To wit: by a scientific description his behaviour cannot be described *as his action*.

When behaviour is represented as resulting from mental states, then, if my diagnosis is correct, its real causes are not represented thereby. This should not prevent us, however, from *representing agents as persons*, as suitable objects of psychological understanding and moral evaluation: these representations do express affective reactions relevant in these two senses. Character traits or reasons are elements in representations connected to motivationally relevant feelings: these influence how we behave towards those whom they are ascribed to. Psychological narratives are suitable means of moral orientation: they convey implicit moral evaluations, and thus configure our moral sensibility. In the process of socialisation we acquire paradigmatic narrative structures that induce affective responses that are stabilised by repeated encounters. Some sorts of motivation, if ascribed, shed favourable, while others unfavourable, light on the agent or behaviour. It is through these interpretations that we understand an interpreter's reactions and the object of interpretation. Affective reactions as conveyed by psychological narratives are partly responsible for how we navigate in social situations, and normative ethics is precisely about how to regulate moral feelings and interpersonal behaviour appropriately (see Frankfurt 1988, 80).

Representing an agent (or anything, for that matter) as a person entails the acknowledgement that it is appropriate to apply to her the categories of *freedom* and *responsibility* – categories that have no counterparts in scientific dictionaries. Agents can be seen as free if they can be interpreted psychologically, otherwise their behaviour can be at most indeterminate. Without psychological ascriptions behaviour can be explained physiologically in causal terms, but from this one cannot see the contribution a person as a person makes to the situation – on the contrary: causal explanations reveal determining factors and thereby exempt agents from responsibility. They are thus incompatible with the idea of a personal decision which follows from freedom and is presupposed by responsibility. Judging responsibility is based on the psychological narratives that serve evaluative purposes along with hermeneutic ones.

Due to the seemingly causal connections between reasons and actions, persons can be treated as the sources of their actions with deliberative capacities – precisely because conceptual connections take some philosophical effort to be revealed. But this "causal" connection is of a different kind from those of physiological explanations. In psychological narratives we cite special "causes" – character traits, motivations, purposes, etc. – pertaining to persons over whom they have authority, as they can decide on them, influence them, and are therefore responsible for them. Because of this authority over (at least some of) their mental states and character traits, persons are responsible also for actions that are understood as springing from these "causes." These "causal connections," as opposed to physiological ones, do not obliterate responsibility – on the

contrary: they are presupposed by it because they belong to the realm of a person's special authority.

Beyond that, moral evaluation does not presuppose freedom in any robust metaphysical sense. Our behaviour may be subject to natural laws without our knowledge, but the narratives we tell to and about ourselves must be based on the illusion of free will – without it there is no way of understanding behaviour as autonomous action. *Will*, as Daniel Wegner (2002, 325 ff.) argues, can be seen as an affective shadow of some physiological processes (a somatic marker), which reminds us that some events are attributable to us. If we interpret a piece of behaviour as an intentional action we ascribe the same to the agent so interpreted. This plays a central role in judging moral responsibility and in deciding who deserves what, i.e. in the context of praise and punishment, and not in an exclusively social sense: bad conscience is a typical example of moral self-punishment. Growing into a competent user of folk psychology we learn which narratives to interpret as expressions of bad conscience, and which ones can incite this feeling – we learn the relevant paradigmatic narrative structures. If one has got bad feelings about his past behaviour then it can be interpreted by folk-psychological concepts, and understood as bad conscience. And vice versa: narratives about our behaviour told by others, or our own self-interpretation can also incite similar bad feelings. But in the case of events that we do not feel to be attributable to us at all, we cannot have bad conscience.

As understanding behaviour and the ascription of responsibility go hand in hand, *psychological interpretation and moral evaluation spring from the same sources*. Kathleen Wilkes (1998, 155) rightly points out: an episodic life – whose events are not connected by the concept of responsibility, are not understood as appropriate objects of praise or blame, and in which emotions are mere feelings without a proper history, etc. – cannot be moral. Moral evaluation presupposes that we treat ourselves as fairly stable intentional systems, to represent ourselves as such by psychological narratives. They bestow moral features upon agents by attributing *reasons, motivation, character,* etc. to them – that is properties that have essential implications for moral evaluation. Therefore, our psychological sensitivity is not merely the basis of our moral sensitivity, but it is partly moral in itself. Concepts deployed in psychological understanding are typically evaluative concepts, or at least have evaluative implications. As they all belong to the same ballpark it is hardly surprising that in understanding actions we frequently rely not on propositional attitudes but on character traits, virtues or imperfections that have moral overtones (see Morton 2003, 43 ff.).

So, psychological narratives bestow upon agents evaluative properties that we ascribe because we are sensitive to certain aspects of the surrounding world, namely to the contribution of agents similar to us (see Mackie 1977, 31 ff.). Agents may not have a grounding property, or a set of properties, in common that warrants the ascription of a given evaluative property. What is common to them is that they are ascribed the given property which cannot be identified independently of the act of evaluation, e.g.

by means of a descriptive science. Goldie (2000, 30) adduces 'dangerous' as an example: "bulls with long horns, dogs with rabies, exposed electric fires, icy roads, strangers with sweets, certain ideas, Lord Byron" have nothing in common that grounds this evaluative property which they share. The possession conditions of evaluative concepts cannot be based on our ability to recognise common features in things subsumed under them – instead they may be grounded in some sort of response, e.g. in affective reactions. And arguably, this is what we do in our psychological interpretations and moral evaluations: we subsume fuzzy affective reactions under concepts thereby structuring and making them communicable.

The essentially evaluative character of folk psychology explains one of its peculiar features, namely that, as Lewis (1991, 209) puts it, folk-psychological terms stand or fall together. Lewis takes these terms as belonging to a causal-descriptive folk theory whose predicates stand for intermediary states necessary for an inference from a perceptual state to a behavioural one. The meaning of folk-psychological terms is thus granted by the inferential role they play in publicly-observable experience, and they refer to the internal functional states that ensure the causal-inferential connections between external stimuli and behavioural responses to them. So, folk-psychological terms refer to the entities, whatever they are, occupying the causal roles specified by the theory. As the entities are defined exclusively by an implicit functional definition, the theory has nothing to say about them apart from the causal role they play in producing behaviour. This definition is a definite description specifying the meaning of the term and thereby its reference. If empirical research reveals that there are no entities corresponding to the descriptions, then the theory, if it is one, turns out to be false. And what is more: potentially a single empty description can have devastating effects for the theory as a whole. If there is a single description that does not pick out some entity, then the term it defines turns out to be lacking reference, and so do all the terms whose definition relies on it.

On Lewis's account with the introduction of new psychological terms or exclusion of existing ones the meaning of others changes too. And folk psychology is subject to historical transformations. Changes in folk-psychological terminology, let it be the introduction of a new term (like the 'unconscious') or the elimination of an older one (like 'demonic possession'), threatens the truth of our previously accepted explanations and predictions as it changes the interrelations of descriptions that give folk-psychological terms meaning. And this contradicts our psychological practice: despite semantic changes we do not consider all previous explanation false, even if the relevant descriptions are now substantially different. It seems to be a serious challenge to Lewis's construal of folk psychology to account for meaning change while preserving the truth value of previous explanations and predictions.

However, this problematic feature is not at all surprising or disturbing if folk psychology is understood as an essentially evaluative discourse and not as a descriptive one.

Evaluative concepts are defined *contrastively*, and the contrasts specifying their content are not binary but obtain in several different directions and to various degrees. As Charles Taylor puts it:

> No one can have the idea what courage is unless he knows what cowardice is, just as no one can have a notion of 'red', say, without some other colour terms with which it contrasts. It is essential to both 'red' and 'courage' that we understand with what they are contrasted. And of course with evaluative terms, as with colour terms, the contrast may not just be with one other, but with several. And indeed, refining an evaluative vocabulary by introducing new terms would alter the sense of the existing terms, even as it would with our colour vocabulary. (Taylor 1985, 19)

If the network of various contrasts constitutive in the meaning of psychological terms changes, their meaning changes too. By introducing new psychological concepts the concept of a person changes too, and our ethics along with it. Just think of how reference to the unconscious (in Freud's sense) can transform the concept of an autonomous, deliberating, responsible etc. person; or of how the concepts of desire, motivation, reason, etc. can thus be reshaped too; and also of the extent to which moral evaluations change if behaviour is understood by it. The time for conceptual transformations comes when we feel that our established concepts are incapable of expressing our relevant affects, or if we find that arbitrarily introduced ones can serve well the purposes of expression and understanding. This is a two-way process: transformations of psychological concepts both imply and arise from changes in morality and social imagery.

The contrastive nature of evaluative concepts also explains how psychological interpretations bring coherence to behaviour. Action explanations are always contrastive: they explain why an agent did what he did as opposed to something else (see Lewis 1986, 229). While actions understood as free events can be explained only contrastively, other indeterminate, but not free, events (like radioactive decay) cannot be contrastively explained – only the causal chain resulting in the given event can be described. Contrastivity can be seen as being in close harmony with the need for psychological interpretation arising in unfamiliar situations, where we need to know why someone did what he did as opposed to something more familiar.

The evaluative aspect of folk psychology is a feature typically overlooked in theories of action and motivation, frequently called Humean, focusing on rational calculation with propositional attitudes. If it is admitted that the evaluative content cannot be separated from folk-psychological concepts then the makes it doubtful whether they can be used in purely descriptive contexts – if no further provisos are added. Philip Pettit (2002, 229 f.) points out this problem in relation to economics. He sees a tension between the rich moral and quasi-moral idiom of folk psychology and the way folk-psychological concepts are used in economic theories. The anthropology of economics

is different from that of folk psychology; they portray motivations differently. While we understand ourselves in the evaluative idiom of folk psychology, economists work with rational calculation without moral overtones in which this aspect has no role to play. This tension, however, is only superficial: in the context of social science, psychological concepts do not play the role they play in the natural interplay of everyday life – they are used as technical terms. And precisely this is the proviso we need to add. In technical contexts the meaning of psychological terms does not depend, or depend only partly on what folk-psychological discourse is like, but they depend more on the theory that exploits them. And thus the tension disappears: terms of an evaluative discourse can be technical terms in descriptive ones – but one has to bear in mind that their meaning is thus changed.

7. Conclusion

I intended to show that due to its nature, folk psychology is incapable of metarepresentation. Psychological narratives do not communicate knowledge about the internal world of the agents they are told about, because their logic prevents them from doing so. So their real function in everyday life should be looked for in non-epistemic contexts. I gestured toward a possible alternative elsewhere suggesting that a *fictionalist* account of folk psychology could be more appropriate than a factualist one emphasising the discourse ability to deliver true representations of an agent's internal functioning (Demeter 2009b).

One possible route a fictionalist account can take is to argue that instead of providing knowledge of internal mechanisms grounding explanations and predictions, folk psychology conveys how one feels about those interpreted. So understood, the truth value of psychological ascriptions is just irrelevant. Instead of aiming at truth, interpretations proposed for acceptance aim at configuring affective sensibilities of others so as to feel similarly about the object of interpretation. Social sensitivity is thus both the basis and the target of psychological narratives: they are expressed, accepted or rejected on this basis, and they tune how we feel about the components of the social world. Our narratives *reflect* the way we feel about others and ourselves, and via our psychological sensitivity they *influence* our navigation in the social world. Interpretation and evaluation concur in this process. Let me offer this paper as a motivation for the elaboration of this fictionalist account.

REFERENCES

Blackburn, S. 1995. Theory, Observation and Drama. In *Folk Psychology: The Theory of Mind Debate*, ed. M. Davies and T. Stone. Oxford: Blackwell.

Brandom, R. 1994. *Making It Explicit*. Cambridge: Harvard University Press.

Davidson, D. 1980a. Actions, Reasons and Causes. In *Essays on Actions and Events*. Oxford: Clarendon.

Davidson, D. 1980b. Mental Events. In *Essays on Actions and Events*. Oxford: Clarendon.

Davidson, D. 1980c. Psychology as Philosophy. In *Essays on Actions and Events*. Oxford: Clarendon.

Davidson, D. 1985. Incoherence and Irrationality. In *Problems of Rationality*. Oxford: Clarendon, 2004.

Davidson, D. 1987. Knowing One's Own Mind. In *Self-Knowledge,* ed. Q. Cassam, Oxford: Oxford University Press.

Demeter, T. 2009a. Where Rationality Is. In *Verstehen nach Heidegger und Brandom,* ed. B. Merker. Frankfurt: Meiner.

Demeter, T. 2009b. Two Kinds of Mental Realism. *Journal for General Philosophy of Science* 40: 59-71.

Dennett, D. 1998. Cognitive Ethology: Hunting for Bargains or a Wild Goose Chase. In *Brainchildren: Essays on Designing Minds*. Cambridge, Mass.: MIT Press.

Dennett, D. 1991. Real Patterns. *Journal of Philosophy* 88: 27-51.

Frankfurt, H. G. 1988. *The Importance of What We Care About*. Cambridge: Cambridge University Press.

Goldie, P. 2000. *Emotions: A Philosophical Exploration*. Oxford: Clarendon.

Hacking, I. 1995. *Rewriting the Soul*. Princeton: Princeton University Press.

Hume, D. 2002. *A Treatise of Human Nature*. Oxford: Oxford University Press.

Kim, J. 2003. Philosophy of Mind and Psychology. In *Donald Davidson*, ed. K. Ludwig. Cambridge: Cambridge University Press.

Kusch, M. 1997. The Sociophilosophy of Folk Psychology. *Studies in History and Philosophy of Science* 28: 1-25.

Kusch, M. 1999. *Psychological Knowledge: A Social History and Philosophy*. London: Routledge.

Lewis, D. 1991. Psychophysical and Theoretical Identifications. In *The Nature of Mind,* ed. D. M. Rosenthal. New York: Oxford University Press.

Lewis, D. 1986. Causal Explanation. In *Philosophical Papers*, vol. 2. Oxford: Oxford University Press.

Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin.

Malcolm, N. 1984. Consciousness and Causality. In *Consciousness and Causality: A Debate on the Nature of Mind*, eds. N. Malcolm and D. Armstrong. Oxford: Blackwell.

Melden, A.I. 1961. *Free Action*. London: Routledge and Kegan Paul.

Morton, A. 2003. *The Importance of Being Understood: Folk Psychology as Ethics*. London: Routledge.

Pettit, P. 2002. The Virtual Reality of *homo economicus*. In *Rules, Reasons, and Norms*. Oxford: Clarendon.

Taylor, C. 1985. What Is Human Agency? In *Human Agency and Language: Philosophical Papers 1*, Cambridge, England: Cambridge University Press.

Velleman, J. D. 2003. Narrative Explanation. *Philosophical Review* 112: 1-25.

Wegner, D. M. 2002. *The Illusion of Conscious Will*. Cambridge: MIT Press.

Wilkes, K. 1998. ΓΝΩΘΙ ΣΕΑΥΤΟΝ (Know Thyself). *Journal of Consciousness Studies* 5: 153-65.

Winch, P. 2008. *The Idea of a Social Science and Its Relation to Philosophy*. 2nd ed. London: Routledge.

Received: June 20, 2009
Accepted: September 15, 2009

Max-Planck-Institut für Wissenschaftsgeschichte
Boltzmannstrasse 22
Berlin 14195 Deutschland
tdemeter@mpiwg-berlin.mpg.de