Making Folk Psychology Explicit

The Relevance of Robert Brandom's Philosophy for the Debate on Social Cognition

Derek W. Strijbos · Leon C. de Bruin

Abstract One of the central explananda in the debate on social cognition is the interpretation of other people in terms of reasons for action. There is a growing dissatisfaction among participants in the debate concerning the descriptive adequacy of the traditional belief-desire model of action interpretation. Applying this model as an explanatory model at the subpersonal level threatens to leave the original explanandum largely unarticulated. Against this background we show how Brandom's deontic scorekeeping model can be used as a valuable descriptive tool for making folk psychology explicit. Following Brandom's non-formalist and non-mentalistic account of reason discourse, we suggest that the process of making sense of others is best captured as proceeding from a 'factive' baseline. According to this picture the ascription of beliefs and desires is not the default interpretation strategy, but rather the result of prior scaffolding of the agent's deontic score. We close by discussing Brandom's model in the light of empirical findings on the ontogeny of reason attribution.

Introduction

It has become customary among philosophers of mind to use the term 'folk psychology' to refer to the practical lore and know how that 'the folk' puts to use in

D. W. Strijbos (✉)
Department of Philosophy, Radboud University Nijmegen, Erasmusplein 1, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands
e-mail: d.strijbos@phil.ru.nl

L. C. de Bruin
Institute for Philosophy, Ruhr-University Bochum, Universitätsstrasse 150, 44801 Bochum, Germany
e-mail: lcdebruin@gmail.com

making sense of each other in everyday social life, their actions in particular. In the early days after its inception, 'folk psychology' was quite literally taken as folk *psychology*: the folk's theory of mind, consisting of a body of generalizations that functionally structures the relations between perceptions, mental states and behavior. Things changed when, independently of each other, Gordon (1986) and Heal (1986) proposed that our social understanding is not theoretical in nature, but rather proceeds by means of (practical) reasoning within a simulated or replicated context (of action). This ignited a new debate—under the now rather misleading name of 'folk psychology' or 'theory of mind'[1]—about the psychology of our social understanding. The theory theorists (TT) remained close to the original rendering of 'folk psychology', claiming that we are in some way or other predominantly *theorizing* about other people's states of mind when interpreting them. The simulation theorists (ST), by contrast, took our abilities to *simulate* others and replicate their thoughts as the central route towards social understanding. What was agreed upon on all sides, however, was that, as Goldman (1989) succinctly put it: 'No account of interpretation can be philosophically helpful […] if it is incompatible with a correct account of *what people actually do* when they interpret others.' (p. 162, italics added). Thus all parties took up the effort, to use Brandom's phrase, of making folk psychology explicit.

In answering this question, the majority of positions drew their inspiration from the belief-desire model of action theory, according to which having a desire towards some goal and a belief regarding the means is a necessary condition for the performance of an intentional action. This model concerning the conditions for intentional action was subsequently adopted as a model for the *folk psychological interpretation* of action. The last decade has witnessed a growing dissatisfaction with this belief-desire model of action interpretation among philosophers, psychologists and neuroscientists (e.g., Gallagher and Zahavi 2008; Ratcliffe 2007; Gallese 2005). Many of our everyday social interactions, these critics argue, don't seem to be adequately characterized as instances of belief-desire pair attribution. In response to this growing unease, adherents of the belief-desire model of action interpretation have taken their claims one step down, from the personal level to the sub-personal or cognitive level. This, however, poses the question whether the belief-desire model is still applicable at the personal, descriptive level. If it is conceded to the critics that it isn't, we are left with the rather surprising result that we are given a clear and well-defined *explanans* while there are serious questions about the *explanandum*. Against this background, we think that the debate is in need of a viable alternative personal-level characterization of folk psychology. It is in this process of making folk psychology explicit that Robert Brandom's philosophy may be of great value.

In the next section we sketch the current situation in the debate on folk psychology. We introduce Brandom's 'deontic scorekeeping model' of the practice of giving and asking for reasons in "Brandom's Deontic Scorekeeping Model". This model, we argue, has great potential for capturing an important *explanandum* of folk psychology: interpreting others in terms of reasons for action. In "Scaffolding Interpretation" we give an account of propositional attitude ascriptions based on Brandom's model. This account satisfies an important *logical* constraint on folk

---

[1] Premack and Woodruff (1978) coined the latter term.

psychology as put forward by the belief-desire model, but resists extending this constraint to the *psychology* of folk psychology. "Deontic Scorekeeping from a Development Perspective" offers some suggestions relating to the ontogeny of folk psychology in light of Brandom's model. Most importantly, we show that it invites a reinterpretation of some of the developmental evidence.

The Belief-Desire Model of Folk Psychology

One of the central tenets of mainstream action theory is that an agent performs an action of a certain type intentionally only if she has a motivating reason that causes the action. A motivating reason consists of a desire toward some end and a belief regarding the means (cf. Smith 1987). This Belief-Desire Model (BD-Model) of intentional agency has dominated action theory ever since Davidson defended his thesis that 'rationalization is a species of causal explanation' (1963, 2001, p. 3). According to Davidson, singling out the reason why the agent acted (rather than merely a possible reason why she could have acted) is singling out the *cause* of the action. In order to do the latter, he argued, it is necessary 'that we see, at least in essential outline, how to construct a primary reason.' (p. 4). Primary reasons consist of a desire toward performing an action of a certain type and a belief that the intended action is of that type. Although Davidsonian primary reasons lack the means-ends structure of Humean motivating reasons, the general message of both accounts is often taken to be the same: the reason for an action is an interlocking belief-desire pair that brings about the action.

The BD-Model has been widely adopted as a model of folk-psychological action interpretation. It is now in fact common sense among many theorists of social cognition that ascribing reasons to others in everyday social practice requires the attribution of an appropriately structured BD-pair. Consider Currie and Sterelny (2000, pp. 145–146), for example, who claim that 'our basic grip on the social world depends on our being able to see our fellows as motivated by beliefs and desires we sometimes share and sometimes do not […]' In similar fashion, Frith and Happé (1999, p. 2) state that 'in everyday life we make sense of each other's behaviour by appeal to a belief-desire psychology.' Or take Scholl and Leslie (1999), who consider it a non-controversial example of everyday sense-making that 'If you see a person running to catch up a just-departing train […] you interpret the person as an intentional agent, who believes that there is a just-departing train, and who wants to get on it.'

In its strongest formulation, the BD-Model of folk psychological action interpretation takes the interpretation process to require the (re)construction of the action under consideration in terms of a constellation of beliefs and desires, minimally a desire toward some goal and a belief regarding the means. The interpretation process thereby obeys the principles of belief-desire psychology, chief amongst them being the 'central action principle' 'if A wants P and believes that doing q will bring about p, then ceteris paribus, A will q.' (Borg 2007, p. 6)

There are two dominant explanations of how this works: theory theory (TT) and simulation theory (ST). According to TT, the folk's understanding of other people's reasons for action requires making use, in some way or other, of an innate or

acquired theory that specifies the functional roles of beliefs and desires (and other psychological states).[2] This theory of mind, as Gopnik and Meltzhof (1997, p. 126) point out, '(...) has many complexities but also a few basic causal tenets (...). These tenets are perhaps best summarized by the "practical syllogism": if a psychological agent wants event y and believes that action x will cause event y, he will do x.'[3] Thus, the aforementioned action principles of the BD-model are considered to be part of the interpreter's folk psychological theory.

The simulationist account of the BD-model replaces the 'theory-driven' interpretation process postulated by TT with a 'process-driven' mindreading procedure.[4] On this version of ST, interpretation proceeds by selecting appropriate target beliefs and desires (of the agent to be interpreted), pretending to have those beliefs and desires oneself (while quarantining some of one's own beliefs and desires, when appropriate), deciding offline what to do on the basis of those beliefs and desires, and finally attributing the resulting decision to the agent under consideration. When the interpretation is not an instance of predicting a future action, but of explaining a past action, one generates target beliefs and desires and tests them in pretend-mode until one finds an appropriate match with the decision (and corresponding action) to be explained. The simulator subsequently attributes the relevant BD-pair to the agent who performed the action (cf. Goldman 2006). Thus on both the TT and ST accounts of the BD-Model, interpreting an agent in terms of her reasons for actions consists in aiming to reconstruct (by theorizing and/or simulating) the BD-pair that causes the action under consideration.

In its early days, the debate was mainly concerned with the folk psychological counterpart of Davidsonian 'rationalizations', i.e. with action interpretation in terms of reasons for action (in the light of which the agent acted). But this changed dramatically over the years. Since the discovery of mirror neurons in monkeys (e.g., Rizzolatti et al. 1996, 2001), the search for mirror neuron systems in humans and the questions regarding the theoretical and philosophical underpinnings of these findings slowly but surely drew attention away from paradigmatic instances of reason attribution to more basic or 'low-level' forms of social interaction.[5] Many philosophers, psychologists and neuroscientists now argue that phenomena such as 'empathic resonance' (Iacoboni 2005; Slors 2007), 'face-based emotion recognition' (Goldman 2006), 'primary' and 'secondary intersubjectivity' (Trevarthen 1979; Trevarthen and Hubley 1978; adopted by e.g. Gallagher and Hutto 2008) are not adequately characterized as instances of 'interpretation' in the traditional sense, i.e. in terms of propositional attitudes. They claim that the bulk of our embodied engagements with others can be adequately explained without reference to

---

[2] There are various TT-construals, for example 'modular TT' (Fodor 1992; Leslie et al. 2005), 'scientific TT' (Gopnik and Meltzoff 1997), 'modal TT' (Maibom 2003), 'external TT' (Braddon-Mitchell and Jackson 2007). Despite many differences, all conform to this general idea.

[3] See also Fodor (1992, p. 283), who suggests that 'normal cognitive development eventuates in the child's internalization of a tacit "metacognitive" intentional psychology: specifically, in the internalization of some version of the folk psychological theory that an agent's behavior is normally caused by his beliefs and desires'.

[4] The non-cognitivist versions of simulation theory put forward by Heal (e.g. 1986, 1998) and Gordon (e.g. 1986, 1992) do not take the BD-model as their starting point.

[5] Cf. Goldman (2006, p. 113), who characterizes 'low-level mindreading' as 'comparatively simple, automatic and largely below the level of consciousness.'

'mindreading', and instead put forward other capabilities to do this job, such as imitation, intentionality detection, eye-tracking, the perception of intentional or goal-related movements, and the perception of meaning and emotion in movement and posture (cf. Gallagher 2001; Gallagher and Zahavi 2008; Gallese 2005; Hutto 2008; Ratcliffe 2007).[6] These embodied practices constitute the baseline for social understanding, what Bruner and Kalmar (1998) called the 'massively hermeneutic' background that is required for the more advanced accomplishments of interpretation.

We think this renewed interest in low-level social cognition is of crucial importance for an inclusive picture of interpretation, and tend to agree with many of these authors that appreciation of these forms of interpretation in daily human interaction severely restricts the scope of 'high-level' interpretation.[7] However, for the purpose of this paper we wish to put these issues to one side. In what follows, our focus is exclusively on these more advanced or high-level accomplishments of interpretation—the interpretation of other people's actions in terms of reasons for action. And here the BD-model is still the common denominator in the debate.

This is not to say that consensus has been complete, however. To many advocates of low-level interpretation the BD-model appeared to be seriously limited, and therefore it seems only natural that some of them started to question its adequacy at the reason-level as well. Thus Ratcliffe (2007) observed that

'FP [folk psychological] and situational explanations have a similar structure. Just as one can say 'if B believes p and desires q, all things being equal, B ought to do r', one can say 'if p is the case and q is the case, all things being equal, B ought to do r'. Norms are integral to the relationships that comprise situations, just as many proponents of FP claim that they are integral to the relationships between beliefs, desires and actions. The systematic structure we require in order to interpret people is out there in the shared world. So the burden need not be carried by a complicated understanding of the relationships between mental states.' (Ratcliffe 2007, p. 97–98.)

Or consider Gordon (2001), who claimed that a reason explanation

'explains action in terms of the reason or reasons for which the agent acted, where "a reason" is understood as a reason for, a reason in favour. This is a reason in the strict sense: a favourable consideration, something about the world—a fact—that, at least to the agent's eyes at the relevant time, favored, or argued in favor of, doing what he did.' (p. 178).

Such 'factive' interpretation proceeds by calling upon states of affairs or events occurring in the world that justify or make intelligible why the agent acted. The

---

[6] Gallese remarks that 'Whenever we face situations in which exposure to others' behavior require a response by us, be it active or simply attentive, we seldom engage ourselves in an explicit, deliberate interpretive act. Our understanding of a situation most of the time is immediate, automatic, and almost reflex like' (Gallese 2005, p. 102).

[7] Cf. Goldman's (2006, p. 147) characterization of 'high-level simulation': it 'includes one or more of the following features: '(a) it targets mental states of a relatively complex nature, such as propositional attitudes; (b) some components of the mindreading process are subject to voluntary control; and (c) the process has some degree of accessibility to consciousness.'

proposal is that the *default* folk-psychological interpretation strategy is to rely on such states of affairs or events when trying to make sense of someone else's action—those states of affairs or events that the agent responded to in performing the action. This is not to say that factive interpretation neglects the perspective of the agent. Rather, it means that the interpreter on default takes the perspective of the agent to be given by *looking out into the world*. When agent and interpreter *share* certain beliefs and desires, ascription of these attitudes can usually be suspended, without loss of predictive or explanatory adequacy, in virtue of the worldly offerings these attitudes are directed at. It is not denied that beliefs and/or desires are sometimes attributed in the interpretative act (see Scaffolding Interpretation), only that doing so is not necessary for reason attribution *simpliciter*.

What the above authors take issue with is a *descriptive* claim on the part of the proponents of the BD-model of action interpretation. It regards *what* people are supposed to be doing at the personal level, when they interpret one another in terms of reasons for action. On the BD-model, as we have seen, this question should be answered in terms of the ascription of appropriately structured BD-pairs. The counterclaim here is that in many everyday cases of sense-making, it suffices to attribute 'factive' reasons—reasons in terms of normatively salient aspects of the (social) environment. So *what* interpreters are doing in these default cases, should *not* be coined in terms of mindreading or mentalizing.

Another point is often made as well. On the BD-model, folk interpreters are explaining other people's reasons in causal terms. It will be remembered that Davidson put forward his causal analysis in order to account for the motivational (as opposed to *merely* justifying) character of the reasons cited in successful reason explanations. If reason explanations are causal explanations, then such explanations had better mention some psychological properties of the agent that suffice (ceteris paribus) to bring about the action. Again, this is a claim on the descriptive level, regarding the *explanandum*. On the alternative reading of the explanandum mentioned here, this appears to be problematic: action explanation in terms of facts (he bought her a present because it's her birthday today), future-dated facts (I'll buy the tickets right away because they will be sold out by tomorrow) or moral truths (she helped the old lady because it was the right thing to do) are not easily squared with such causal analysis (cf. Gordon 2001). But there are other ways to deal with the 'many reasons problem' that inspired Davidson's analysis. Gordon (2000), for example, proposes a counterfactual analysis on which the counterfactuals are not backed up by (causal) laws but by the agent's *policies*. Such policies can easily be issued in factive terms, rather than in terms of explicit mental states. Hutto (forthcoming) speaks of the *informational relevance* of reasons given in folk practice and asserts that "It is entirely possible […] that the factors cited by the folk are worth mentioning does not entail that they pick out (or attempt to pick out) causally relevant properties per se." Rather, he proposes to analyse such relevance in terms of the role reasons given play in 'folk psychological narratives' (2008).

We think that the target of these alternative views is the BD-model as it was originally intended: a proper, personal-level description of *what people do* when they interpret one another in terms of their reasons for action. At this point it might be objected that this is not entirely fair to those defenders of the BD-model who use it

primarily as an *explanatory* model targeted at the sub-personal level.[8] On this rendering, the BD-model is not supposed to tell us *what* it is that we do when we interpret each other, but *how* we do it. Answering this how-question involves uncovering the cognitive mechanisms and processes that underlie our socio-cognitive capacities and cause our judgements about other people's reasons for action. Consider Nichols and Stich's (2003, p. 94) ST/TT hybrid model of 'the full third-person mindreading system' that is supposed to be operative in adult social cognizers. This system contains mechanisms for the attribution of percepts and desires and a complex subsystem of mechanisms for the attribution of default and discrepant beliefs. The operations of these mechanisms need not be phenomenologically transparent or practically salient, so the aforementioned objections appear to be largely irrelevant.

Interestingly, Carruthers (2006) adopts Nichols and Stich's model in order to explain why his cognitive, sub-personal hypothesis of 'massive modularity' is counterintuitive at the phenomenological, personal level. For 'this human mind-reading module operates with a simplified model of the mind and its operations, included in which is the idea that the mind is transparent to itself.' (p. 215). The hypothesis of massive modularity predicts 'that the mind decomposes into far more components than would generally be recognized … by common sense psychology.' (p. 213). We agree that phenomenology has no special authority when it comes to the sub-personal cognitive mechanisms and processes of social cognition. Moreover, it is primarily what interpreters *do*—their behavior—that such sub-personal goings-on need to explain, not what people *experience* in the act of interpretation. But when it comes to explaining the behavior of social cognizers, both Nichols and Stich and Carruthers take as their starting point a mentalized picture of the *personal level explanandum*. Thus Nichols and Stich tell us they will present their account 'of the mechanisms underlying third-person mindreading—*the attribution of mental states to others*.' (p. 60, italics added). And Carruthers gives a picture of *common-sense psychology* according to which 'percepts give rise to beliefs and serve to inform practical reasoning; beliefs and desires interact in practical reasoning to create intentions and actions…' (p. 213). He then claims that common sense's only failing is that 'it doesn't postulate enough perceptual mechanisms, nor nearly enough mechanisms for producing new beliefs, new desires, and new actions' (ibid). Carruthers thus deals with phenomenological objections to his massive modularity hypothesis by invoking a sub-personal mindreading model that itself appears to be *derived* from a personal level characterization of common sense psychology that fits the BD-model. But it is exactly this characterization that is at issue here.

There is nothing in principle wrong with switching from a descriptive to an explanatory use of the BD-model. Yet it is not obvious that this underground movement is helpful in the present case. For *if* it is agreed that *what* interpreters do is best captured *not* in mindreading terms (i.e. BD-pair attribution), then it is not at all

---

[8] We thank an anonymous reviewer for this suggestion. In the remainder of this section we talk of 'descriptive' use of the BD-model when it is targeted at the personal-level *explanandum*. The explanandum can be an instance of folk-psychological reason explanation or prediction. In its 'explanatory' use, the BD-model is targeted at the subpersonal-level *explanans*: some cognitivist story that causally explains the behavior of the interpreter. The difference becomes relevant when the explanans of the interpretative act is framed in terms of the BD-model, but the explanandum itself is not.

clear why we should invoke a 'full third-person mindreading system' to explain *how* they do this. It seems strange to hold on to an explanation when its target explanandum turns out to be rather different. Doing so would need more argumentation. If indeed the default strategy of folk-psychological interpreters is to attribute 'factive' reasons to others, then it needs to be shown why an explanation of such attributions needs to invoke metarepresentational processing in terms of BD-pairs.[9]

Of course, it is also open to the defenders of the BD-model to argue that it is both descriptive and explanatory, or rather, that it is descriptively adequate *because* it is a successful model at the explanatory level. There is nothing in principle wrong with this move: explananda may turn out different than expected when new explanatory theories surface. Yet this move seems a little suspect in the current context: the BD-model is not a new explanatory theory—it is derived (historically, at least) from its use as a descriptive account.

We certainly do not mean to make a decisive argument against the BD-model of action interpretation here. We merely intend to show that: 1) as a descriptive model it is not uncontroversial, and 2) as an explanatory model it needs to be clear about the phenomenon to be explained. Both points reveal the importance of answering the what-question at the personal level of description: what it is that people do when they interpret one another in terms of reasons for action.

At this point, it is interesting to turn back to Davidson's original thesis for a moment. He claimed that 'In order to understand how a reason of any kind rationalizes an action it is necessary and sufficient that we see, at least in essential outline, how to construct a primary reason.' (1963, 2001, p. 4). There is an important question to ask here. Who is the 'we' in the above quote, who is supposed to 'see how to construct a primary reason'? As far as Davidson's claims go, this is the *action theorist*. All Davidson's proposal strictly amounts to is that the reason for an intentional action can always be described correctly in terms of an appropriately structured belief-desire pair. For an action theorist interested in the conceptual connections between our concepts of reason, intention, belief, desire and the like, this is an interesting and substantial thesis. It suggests that having a reason for action entails having a belief and a desire that are syllogistically structured. The attribution of a reason would thus imply the presence of such a belief-desire pair on the part of the interpreted agent. But this is something the action theorist concludes on reflection, not something the interpreter should necessarily consider in real-time social interaction. In other words, from the claim that a reason explanation *implies* that the agent has a motivating desire and an appropriate instrumental belief, it does not follow that making sense of the action requires the interpreter to *actually ascribe* such a belief-desire pair. All that follows is that an important part of folk psychological reality can be (philosophically) reconstructed in terms of the BD-model. This is a claim we do not wish to argue against.

---

[9] Of course, there are situations in which it *does* matter to take into account the *particular* perspective of the agent. This is especially important if the agent's perspective diverges from the interpreter's take on the world. We explain this is more detail in "Scaffolding Interpretation". Here, however, we strictly deal with cases of factive interpretation, in which the perspectives of the agent and the interpreter do *not* diverge in this sense.

What we hope to have shown in this section is that in the debate on social cognition, much hinges on the current status of the BD-model. The present situation appears to be that opponents criticize the BD-model at the personal level of description, whereas proponents focus their attention on the sub-personal level of explanation. Proponents need to ask themselves: Is the BD-model still a descriptive model or has it turned into an explanatory model? If the former, then it should be acknowledged that there is serious *disagreement* about the explanandum and that additional resources are needed to settle the matter. In case of the latter: we can't properly evaluate the BD-model until we are clear about what it is that the model is supposed to explain. Against this background, we wish to introduce Robert Brandom's 'deontic scorekeeping model' (DS-Model) in the next section. The DS-model provides a detailed description of our practice of giving and asking for reasons that is at odds with the BD-model. We argue that if this model can earn its keep at the descriptive level, this will have consequences at the explanatory level as well.

Brandom's Deontic Scorekeeping Model

The deontic scorekeeping model attempts to characterize what people do when they take part in the game of giving and asking for reasons. On the DS-model there are two kinds of *deontic statuses* that are central to our reason-giving practices: being *committed* and being *entitled* to say or do something. There are furthermore different sorts of attitudes that can be adopted towards the deontic statuses of commitment and entitlement. Let us start with 'theoretical' reasons: reasons for claims made rather than actions performed. First, in making a claim that p, a speaker is *undertaking* a 'doxastic' commitment to the effect that p and normally also *acknowledges* the commitment undertaken. An interlocutor will *attribute* the commitment undertaken to the speaker and hold her responsible for it—even if the commitment undertaken is *not* acknowledged by the speaker. This can happen when in committing herself to p, the speaker, unbeknownst to herself, also undertakes a commitment to the effect that q (entailed by p). In that case, fellow interlocutors may still attribute the entailed commitment to her. This will come out when the speaker undertakes further commitments that the interlocutor deems incompatible with q. He may then challenge her and confront her with her commitment to q, upon which she may defend, retract or refine her commitments by undertaking (and acknowledging) new ones, starting a new round in the game of giving and asking for reasons.

The speaker may thus attempt to *justify* commitments undertaken. This is where the deontic status of *entitlement* comes in. In questioning or challenging a claim made, the interlocutor is *suspending* or *withholding* entitlement to the commitment undertaken. And in subsequently giving a reason for p, say r, the speaker is undertaking a further commitment in order to *vindicate* entitlement to the original commitment undertaken (p). If the interlocutor accepts r as a reason for p, he will attribute entitlement to the new commitment undertaken (r) and judge p to either follow from or be compatible with r. As a result, the interlocutor grants authority to the commitment originally undertaken: that p. But the interlocutor may also fail to see r as a justification for p, even if he does attribute entitlement to the speaker with

respect to r. He will then withhold entitlement to the original commitment and will probably challenge the speaker to make herself better understood, asking her for further reasons.

When an interlocutor asks for reasons for *action*, he is withholding entitlement to a *practical* commitment—an intention (and resulting action), a commitment to act.[10] In giving her reasons, the agent then attempts to vindicate entitlement to this commitment by giving a practical reason that permits (and perhaps commits her to) the action (to be) performed. The practical reason is often a 'doxastic' commitment, an assertion concerning some (future) state of affairs, value, rule, etc. If successful, the interlocutor will judge the reason given to warrant the original practical commitment to act.

According to the DS-model, people participating in reason discourse are *keeping score* of each other's deontic status, of what each is committed, entitled and not entitled to say and do in light of claims made and actions performed. These scores are brought to bear in assessing new actions and speech acts. On the DS-model, making sense of someone's action precisely consists in such assessment: in judging whether the action performed is compatible with (and perhaps follows from) the agent's deontic scores as kept by the interpreter. In saying or doing something participants *change their deontic scores* by altering the constellation of moves they are committed, entitled and not entitled to make. This may enable them to provide a reason for a claim made or an action performed, but it will also affect which moves one is committed, permitted or not permitted to make in the future.

Consider the following short conversation:

A: "Why did you get up so early?"
B: "Got an early shift this morning."
A: "Oh, right, of course."

Using the DS-model, we could start describing what is going on between A and B as follows. A withholds attribution of entitlement to B's practical commitment to getting up early. B responds by explicating a second commitment that should vindicate entitlement: that she has an early shift this morning. A attributes entitlement to this second, doxastic commitment and commits himself to it. In the light of this commitment, he grants entitlement to B's original commitment to getting up early. Importantly, this way of characterizing reason discourse still leaves open the question whether the interpretation process involves the attribution of *mental states*.

According to the BD-model, the above piece of reason discourse should be regarded as an essentially *truncated* expression of the rationalizations A has to think through in order to make sense of B's action. On this picture, genuine action explanations always include a proper belief-desire pair. So if one or both of these attitudes are absent in the conversation, the interpreter must insert them himself through some or other mindreading routine. In order to really understand the answer

---

[10] Here and throughout, we take it that intentionally performing an action at t is typically the reliable result of an intention to perform an action of that type at or prior to t. Explaining the intention with which one performs a certain action is therefore normally sufficient to explain the action. Brandom treats actions as 'language exit transitions': reliable dispositions to respond differentially to the acknowledging of commitments by bringing about various kinds of states of affairs. (see 1994, p. 235)

given, to appreciate it as a genuine reason explanation, A needs to 'read between the lines' and filter out just the right belief-desire pair: B wants to arrive at work on time, she believes that by getting up early she will arrive at on time, so she will get up early (or something like this). On the BD-model, A's mindreading routine thus has to mirror the formal structure of a kind of practical syllogism.[11]

This is reminiscent of what Sellars (1953) called 'the dogma of formalism', according to which 'the inference which finds its expression in 'it is raining, therefore the streets will be wet' is an enthymeme' (p. 313). According to this 'dogma', whenever an inference such as 'it is raining, therefore the streets will be wet' is endorsed, it is because of a belief in the conditional 'If it rains, then the streets will be wet'. With this tacit premise supplied, the inference is an instance of conditional detachment of the formally valid inference 'If it rains, then the streets will be wet. It is raining, therefore the streets will be wet.' Thus, appreciation of the proprieties of inference requires mastery and application of the principles of logic that are supposed to guide our inferential activities. Inferences like the above 'it is raining, therefore the streets will be wet' are consequently treated as derivative.

Drawing from Sellars' observations, Brandom wants to reverse the order of explanation: he proposes to treat inferences like 'it is raining, therefore the streets will be wet' as primitive, and their formal renderings as derivative.[12] Therefore, he starts with *materially* correct inferences, inferences that are correct in virtue of their *content*, not their form. The inference 'it is raining, therefore the streets will be wet' is a material inference, and will normally be treated as correct *because* streets get wet under conditions of rain. Brandom extends Sellars' account of 'theoretical' material inferences to practical ones. As we read Brandom, he treats practical inferences as transitions from reasons to intentions (and actions), from doxastic commitments to practical ones.

Now the important question is: what determines the specific inferential relations of commitment, entitlement and incompatibility associated with a speech act or an action? On the DS-model these inferential roles are played by *material* inferences. Thus when A treats B's answer 'got an early shift this morning' as a proper reason explanation in response to his question 'why did you get up so early', this is in virtue of the material inference 'she's got an early shift this morning, therefore she should get up so early'. The reason provided serves as the antecedent and the (commitment to perform the) action as the consequent. In attributing entitlement to the agent's practical commitment to get up so early, the scorekeeper *endorses* this practical material inference, thereby treating the material inference as a proper one, a

---

[11] Cognitivist ST (e.g. Goldman 2006) is no exception here: even though the central simulation routine might not require tacit knowledge of central action principles relating beliefs, desires and intentions/actions, it still needs to *follow* the cognitive route marked by such principles: no pretend-decision and hence no action interpretation without integrating the pretend-belief with the pretend-desire. Therefore, as Goldman admits, even if we think of simulation as being process-driven, such a process still requires that 'some elements inside the attributor causally mediate between his explicit premises and conclusions, and that the causal structure of these elements mirrors the logical structure of psychological theory.' (2006, p. 33)

[12] Brandom treats the formal principles of inference as expressive tools that enable us to make explicit the proprieties of material inferences, by *saying* what was *done* in endorsing the inference. Thus the conditional 'whenever it rains the streets will be wet' is considered as a means of saying what is being done in endorsing the material inference 'it is raining, therefore the streets will be wet'.

permitted move on the part of the agent in this game of giving and asking for reasons.

The propriety of this material inference can be assessed *without reference to the agent's mental states*: it is *because* she has early shifts on Mondays that the agent should get up early today. Material inferences are correct in virtue of their content: some pattern in the world that reflects the inferential relation endorsed. In this case it is a normative pattern in B's practical life, namely that she should get up early on days that she runs early shifts. Its normativity reflects the fact that the corresponding material inference is always subject to assessment by fellow scorekeepers (and oneself). Hence it is an essentially social pattern. But it is not a strictly psychological pattern in the agent's mental states.

The practical commitment to get up early is of course agent-specific, as is the reason for undertaking it. It is the *agent's* commitment and *her* reason, not the interpreter's. In this respect, practical material inferences are different from theoretical ones. In theoretical material inferences, the consequent expresses some fact about the world. As long as entitlement is attributed to the antecedent, commitment to the consequent will not be restricted to the person making the claim. In general: the soundness of theoretical material inferences does not depend on who undertakes the relevant commitments. In case of practical material inferences, by contrast, the consequent contains reference to a specific individual or group of individuals.

Consider our example again. The scorekeeper asks the agent why she got up so early and the agent replies that she's got an early shift this morning. By accepting the agent's answer as a piece of knowledge and moreover as a valid reason, the scorekeeper will not thereby have committed himself to getting up early. The inference in virtue of which the agent's answer vindicates her practical commitment is agent-relative: 'she's got an early shift this morning, therefore *she* should get up early'. Practical material inferences do depend for their validity on the person who undertakes the commitment featuring in the consequent. In general, the soundness of practical material inferences is sensitive to the conditions of the relevant agent.[13]

But it doesn't follow that this agent-relativity should be expressed in *mentalistic* terms. Again, the endorsed material 'she's got an early shift this morning, therefore she should get up early' contains no references to mental states. The DS-model allows us to articulate each move in this piece of reason discourse without having to add further silent psychological premises. This fits nicely with Ratcliffe's remark in the quote above: 'The systematic structure we require in order to interpret people is out there in the shared world. So the burden need not be carried by a complicated understanding of the relationships between mental states.' (cf. The Belief-Desire Model of Folk Psychology)

The DS-model moreover appears to be at odds with a causal analysis of reason explanation. Of primary interest to the interpreter are material inferences proposed by the agent, inferences that the interpreter himself needs to assess in order to either

---

[13] The agent relativity here only concerns the practical situation the agent is in. Had the interpreter been in the same practical situation, he would probably have acted similarly. This *inter*personally robust action pattern precisely reveals the factive nature of the reason attributed by the interpreter. Mere *intra*personally robust action patterns often call for mentalized ascriptions. See Scaffolding Interpretation.

endorse or reject them. These inferences feature the agent's considerations in favor of performing the action, the reasons in the light of which the agent acted or in terms of which she can rationalize her action. The interpreter has to assess whether these considerations *make sense or not*, whether it is correct for the agent to follow this line of thought or not within her practical context. It is not at all obvious what notion of causation, if any, could play a role here in structuring this essentially normative dimension of action interpretation.[14] But the proposals by Gordon and Hutto mentioned in "The Belief-Desire Model of Folk Psychology" would fit in nicely: the normative patterns of action reflected in practical material inferences correspond with the agent's policies and are manifested in the stories that can be constructed about her in folk psychological practice. The central role in the DS-mode for the construction, endorsement and rejection of material inferences will also remind the reader of Heal's (1998) notion of 'co-cognition': 'a fancy name for the everyday notion of thinking about the same subject-matter' (p. 483). On many occasions, *what* the agent is considering and responding to by performing his action is some (future) fact or value. Understanding the agent's action then involves co-cognition of the relevant material inferences, understanding what he responded to, what it was that he took into consideration that provided a reason to act upon.

The DS-model is a descriptive model, meant to characterize the explanandum of action interpretation: what it is that we do when we interpret other people in terms of their reasons for action. Is it superior to BD-model in this respect? We don't have the space here to argue for this. Yet we hope we have presented a way of thinking that points to an affirmative answer. But then again, perhaps this question needs no answer. For it seems that some proponents of the BD-model have already opted out of the descriptive level and intend their model to be used only at explanatory, sub-personal level. If this is case, however, they should be aware of the consequences of this neglect. If the DS-model is close to being an adequate description of the explanandum, an explanatory BD-model seems unnecessary. For an explanation of how we construct and endorse material inferences that lack reference to mental states does not require reference to metarepresentational processing.[15]

Of course there are instances of action interpretation in which interpreters do make reference to the agent's mental states. People do think about other people's mental states in folk psychological practice. But this does not mean that such reference is the *default* interpretation strategy. In fact, the DS-model suggests the

---

[14] Davidson famously tried to reconcile the normative and the causal dimension of reasons for action by making a distinction between essentially interest-relative and hence intensional causal explanations and extensional causal relations between particular events (e.g. 1970). For Davidson, rationalization is normative, if only because 'there is a certain irreducible—though somewhat anaemic—sense in which every rationalization *justifies*: from the agent's point of view there was, when he acted, *something to be said for the action*.' (1963, 2001, p. 9, italics added). Yet if successful, such normative explanations refer to the mental event that is causally related to the action. This reconciliation faces notorious problems (see e.g. Hutto 1999), primary among which is that Davidson does not seem to be able to tell us *why* we should consider the normative reason explanation as a causal one if it is not in virtue of citing causally efficacious mental properties. But Davidson's worry was exactly that we couldn't make sense of the normative dimension of reason explanation in this way. We think that Davidson was right on this latter point.

[15] At the very least, this shifts the burden of proof to proponents of the BD-model. That is, they need to explain why the personal-level descriptions provided by the DS-model have to be explained via sub-personal B-D models.

exact opposite. People can start off on the simple assumption that they *share* much of the inferential web of commitments and entitlements that form the background of their conversations about their reasons for action. The material inferences that are constructed and endorsed throughout everyday conversations suggest they do. As pointed out above, these inferences often reflect patterns in people's practical lives, not patterns in people's mental states. In the next section we will indicate how mentalistic interpretation, when it does occur, can be viewed as *the result* of a piece of reason discourse, rather than its *foundation*.

A brief note before we go to the next section. Phenomenology has recently played an important role in the debate on social cognition by questioning the BD-model of reason interpretation at the level of our everyday experience. In response to phenomenologists' critiques (e.g. Gallagher 2001, 2007; Gallagher and Zahavi 2008) defenders of classic mindreading theories have taken Carruthers' line of argument (see The Belief-Desire Model of Folk Psychology) and were quick to point out that their models were explanatory ones, exclusively directed at the sub-personal level (e.g. Herschbach 2008; Spaulding 2010). Attribution of mental states, they pointed out, need not be phenomenologically transparent. We would like to stress that the DS-model is not intended as a phenomenological model; it does not try to capture the *experiential* level of action interpretation. Again and again, the central question throughout this paper has been: *what is it that people do* when they interpret each other in terms of reasons for action? What it is that people do—characterized at the personal level of description—also need not be phenomenologically transparent. Nor is the vocabulary of commitment and entitlement supposed to be 'common sense'. Brandom's DS-model is a *theoretical model* that attempts to make explicit some important features of our social practice. Saying what it is we do in social practice is not easy and should not be expected to be explicit in common sense. As a theoretical model, it provides an alternative to classical mentalistic descriptions of folk psychological action interpretation. It thereby reveals the controversial status of the BD-model in the current debate and opens up a serious discussion regarding the explanandum.

Scaffolding Interpretation

What is the interpreter doing when he makes sense of an agent's actions in terms of her reasons? According to Brandom's DS-model, he is attributing entitlement to the agent's practical commitment to act, by endorsing a practical material inference, featuring the agent's reason in the antecedent and her intention (and subsequent action) in the consequent. The DS-model describes the interpretation process in terms of the attribution of deontic statuses of commitment and entitlement, the specific import of which is determined by their place in the inferential web of background commitments and entitlements that make up the agent's deontic score or context.

In section three we argued that attribution of deontic statuses is not equivalent to attribution of mental states. In this section we want to elaborate on this point. Once again, consider our 'getting-up-early' example, in particular the answer B gives to A's question: 'Got an early shift this morning'. This is what we call a 'factive'

explanation, since it expresses an assumption or re-affirmation of common background knowledge between agent and interpreter. B reminds A of the fact that she has an early shift today. A immediately commits himself to this claim, thereby automatically granting entitlement to B with respect to the same commitment. A treats B as an authority on this matter in this instance: she is a speaker of truths, an indicator of facts. B's factive answer to A's why question has the effect of re-establishing a part of their common deontic background. On this common ground, A is able to endorse the relevant material inference (B has an early shift today, so she should get up early') and attribute entitlement to B's action. The relevant inference is agent-specific—it is B's practical commitment, not A's—yet makes no reference to the agent's mental states. Nor does the subsequent attribution of entitlement to the consequent: B's commitment to having got up early.

The same considerations apply to instances in which the reasons of the agent are shared by the interpreter. In such cases the relevant practical material inferences are not agent-specific in the sense described above: the commitments in the antecedent as well as the consequent pertain to both. Consider an example provided by Gordon (1992):

> 'You and a friend are hiking up a mountain trail, talking. Suddenly, in mid-sentence, your friend stops in his tracks, blurts out, "Go back!" then turns and walks quietly and quickly back down the trail. You are puzzled. You follow him, looking over your shoulder to search the environment for an explanation […] You look for salient features in the middle distance, particularly for menacing, frightening things further up the trail […] Then you spot it: above you, at the next switchback, something startling, menacing and frightening—a large bear, and it's a grizzly!' (p. 13–14)

In responding to the agent's warning (Go back!) and following her down the trail, the interpreter commits himself to going back and thereby instantaneously attributes entitlement to the agent's commitment. At this point, there is no question whether or not she is entitled to her practical commitment. The interpreter treats her as an authority, an indicator of shared reasons, whatever they turn out to be. Searching the environment for an explanation, the interpreter is looking for the reason he already committed himself to by following the agent's lead. By spotting the menacing, frightening grizzly, the interpreter hits on both their reason for action—a fact, something about a shared environment that prompted them to act. The relevant practical material inference should be expressed in the first person plural: 'there is something frightening and menacing up the trail, so *we* shall go back!'

If the bear had been walking up the trail below them, both of them would have responded by walking in the opposite direction; if the bear had been approaching from the left, they both would have gone to the right, etc. These counterfactuals point toward a shared *practical inferential pattern*. Brandom proposes to treat normative vocabulary such as 'should', 'supposed to' or 'ought to' as an expression of the fact that such patterns have import for more than one person in particular. Suppose the agent had spotted the bear at some earlier time. Pointing out the bear to the interpreter, still at a relatively safe distance, she says 'We really should go back!' Using this specific normative vocabulary, the agent makes it clear that going back is the right response upon spotting a bear. In her eyes, going back upon spotting a bear

is a practical pattern that supplies the standard, not for her in particular, not even for them in particular, but for anyone in their situation.[16]

This is not to say that by calling upon the facts to make sense of an agent (that she has an early shift this morning, that there is something frightening and menacing up the trail—a bear, it turns out, at the next switchback), the reasons attributed are free of emotional or conative import. The point is rather that such imports need not be peculiar to or have an effect on the agent *in particular.* There is something *frightening* and *menacing* up the trail, something that has certain emotional and desirability characteristics for *both* agent and interpreter. For most of us, probably.

But suppose it is really a raccoon at the next switchback. As it happens, the agent is terrified of these creatures. Suppose further that there being a raccoon somewhere near him has no particular conative or emotional import for the interpreter, except perhaps for some general curiosity about wild animals. While responding to her warning at first, the interpreter immediately stops going down the trail after having spotted the raccoon in the direction the agent was looking. For as far as he is concerned there is nothing frightening and menacing up the trail, only a cute little raccoon. 'It's only a raccoon!' the interpreter cries out with an incredulous voice as he turns around and starts walking up the trail again. 'Make sure it's gone' the agent cries back, 'I hate those animals!' Characterizing the interpretation process here as involving the ascription to the *agent* of a desire, say, not to be near raccoons (and some specific emotional states) seems quite to the point.

We suggest that conative vocabulary be treated as a species of normative vocabulary that serves to articulate practical inferential patterns that are agent-specific in the counterfactual sense specified above. In saying that the agent doesn't *want* to be near raccoons, say, the interpreter would make it clear that there is a pattern of practical material inferences that *the agent* in particular endorses, but that *he himself* need not. If the raccoon had been approaching from below, *she* (but not he) would have responded by moving further up the trail, if the raccoon had been at her left hand side, *she* (but not he) would have responded by moving to the right, etc. Ascribing a desire to the agent in this scenario, the interpreter is *particularizing* or *individualizing* the reason of the agent.

In "Brandom's Deontic Scorekeeping Model" we argued that practical material inferences are always agent-specific in the sense that the soundness of the inference depends on the person or persons endorsing it. Here we should make a more fine-grained distinction in counterfactual terms between agent-specificity that is *inter*personally robust and agent-specificity that is only *intra*personally robust. The bear example and the getting-up early example are similar in one respect: the

---

[16] Brandom (1994, 243–253) distinguishes three kinds of what he terms 'normative' vocabulary: 'institutional' oughts (e.g. 'is supposed to'), 'unconditional' oughts ('ought to') and 'instrumental' or 'prudential' oughts (e.g. 'wants to'). See also his 2000 (p. 79–92). The first two make explicit inferential patterns that pertain to respectively a group of individuals (with some institutional status) and all individuals (irrespective of their institutional status). Brandom treats the unconditional oughts as specifically moral oughts. We think there are many more distinctions to be made in folk practice. In our bear-example, for instance, the 'should' in 'we really should go back' does not sound as a moral ought, nor as an institutional 'supposed to'. The important point for us is that the 'oughts' at issue here form a contrast class with the instrumental ought: interpretation in terms of the former does not require mindreading. In treating Brandom's 'instrumental' ought as individuating or particularizing the agent's reasons beyond what is interpersonally valid (see below), we go beyond Brandom's own analysis.

interpreter *will* share the agent's practical commitment *if the relevant counterfactual holds*. In the bear example, agent and interpreter find themselves in the same situation: there is a bear up the trail in front of them. In the getting-up-early example the interpreter does not find himself in the same situation as the agent: he does not have an early shift today. But if he had had an early shift this morning, he *should* have got up early. In both examples, the practical material inference that is endorsed is counterfactually robust in an interpersonal sense: given that certain circumstances pertain to both, they would both act in similar ways. In both examples there seems to be a general pattern at work: You're supposed to get up early when you have an early shift, you should retreat when you encounter a bear in the wild. These patterns provide the standard for assessing people's actions in the relevant situations. Since they don't provide the standard for anyone in particular, attribution of specifically *conative* states, such as a desire, is not necessary in these cases.

Attribution of such states becomes practically important when the agent-specificity of the relevant practical material inference is counterfactually robust only in an *intra*personal sense. By *ascribing* a particular desire or emotion instead of *describing* the relevant situation as undesirable or having a certain import, the interpreter underscores the merely *intra*personal validity of the practical inference endorsed. He thereby in effect individualizes or particularizes the agent's reason: it is a reason the agent responds to, but he himself (or anyone else in general) need not, under similar circumstances. Ascription of desires marks the differentiation between 'mere' motivating reasons and normative reasons. The important point is that in many instances of action interpretation, this differentiation is *absent*. When an interpreter makes sense of the agent by calling upon the fact that p, he interprets her as being *motivated* by the *normative* reason that p. It is precisely when sense-making in terms of interpersonal norms fails, that ascription of mere motivating reasons becomes practically important. Ascription of desires and the like enables the interpreter to mark the particularity of the inferential patterns endorsed by the agent and keep track of these patterns in subsequent acts of interpretation.

Ascription of *belief* is especially appropriate when the agent responds to something that is or might not be the case in the eyes of the interpreter. By ascribing a belief to the agent, the interpreter acknowledges that he himself does not undertake and in fact challenges entitlement to the doxastic commitment undertaken by the agent in the antecedent of the relevant practical material inference. Suppose the grizzly behind the bushes that featured in the agent's considerations was in fact a large bearded man (wearing a furry coat). Why did she go back? Scanning the environment the interpreter is still at a loss. He *asks* her. She responds that there is a bear behind the bushes. He corrects her, she realizes her mistake and they both have a good laugh about it. Now why did she go back? Because she *believed* there was a bear (over there). By ascribing a belief to the agent, the interpreter makes it explicit that he doesn't endorse the doxastic commitment that featured as the antecedent in the agent's practical inference leading to action. Here we have a situation in which it is practically significant for the interpreter to highlight the *representational* character of the agent's considerations. This makes sense in this scenario, precisely because the agent *misrepresented* the environment. On Brandom's account, belief ascriptions have this practical expressive function: crudely put, to introduce the representational dimension of the practice of giving and asking for reasons (1994, chapter 8). Giving

the commitments of the agent such representational status is another means of individualizing or particularizing the reasons acted upon. Like in the getting-up-early example, the relevant counterfactuals may be interpersonally robust: if the situations specified by the (false) beliefs in the antecedent of the inference had obtained (if there had been a bear), the interpreter might have undertaken the practical commitment himself (to back away). Unlike the getting-up-early example, however, the assessment of the agent's practical commitment in this example *requires* such (possibly) counterfactual considerations. The doxastic commitment in the antecedent is or may be contrary to fact as far as the interpreter is concerned. Ascribing a belief concerning the relevant commitment enables the interpreter to mark its non-factive nature and to keep track of it in subsequent rounds of reason discourse.

Brandom's DS-model sketches a picture of our folk psychological practice as proceeding from our engagements and interactions with each other in a shared practical world. It suggests an account of interpretation according to which sense-making is a cooperative enterprise that usually starts within a factive context, but which can proceed in non-factive, particularized forms that can be made explicit in terms of propositional attitude ascriptions. Action interpretation normally starts by calling upon what it was in the world to which the agent responded in performing her action. But it often happens that interpreters remain ignorant of what was responded to. Asking the agent for her reason and being given a factive answer will then often suffice. And sometimes the factive assumptions reflected in the agent's answer may be unwarranted in the eyes of the interpreter; the 'facts' provided by the agent may be of a peculiar kind to respond to; they may turn out not to be factive at all or remain subject of dispute or uncertainty. Obstacles of this kind are normally resolved *during* the game of giving and asking for reasons, not prior to it.

Ascribing mental states in folk psychological interpretation is an act of individualizing or particularizing of the reasons acted upon. This act of individualization reveals a reason that is particular or peculiar to the agent, in the sense that it is a reason to which the interpreter would not normally respond (for the interpreter, it lacks the particular affordance that it has for the agent), or a reason that concerns something that is or might not be the case (in the case of a (false) belief). But this is the exception that confirms the rule: default interpretation appeals to facts with shared imports.

Unsatisfactory explanations of actions in terms of such facts are exactly what *drives* reason discourse and allows for additional pieces of a deontic 'scaffolding' to be built: extensions of the agent's deontic scores from the common to the particular, on top of which the interpreter is now able to follow the agent's considerations. This is a more advanced and derivative form of action interpretation that normally comes into play when default factive interpretation breaks down. In our examples we have restricted this process of scaffolding to include only the ingredients provided by earlier phases of the same conversation. But in reality, these deontic scaffoldings are more diachronically extended, being constructed, maintained, torn down and repaired throughout our (personal) lives together, in terms of our beliefs, desires, hopes, fears, doubts, values, etc., depending on the intensity and the nature of our relationships together. Each conversation, each action, each overheard remark or characterization by third parties may add to the deontic structure in terms of which

we interpret others. But at the same time, each of these social happenings proceeds against the background of a common world we take each other to be responsive to.

At the end of this section, we would like to come back to a point briefly mentioned in "The Belief-Desire Model of Folk Psychology": that a belief-desire pair can be *implied* by the attribution of a reason, without its ascription being *actually* required for interpreting the action in terms of a reason. An important consequence of this observation is that it is always possible in principle and never wrong (just odd at times) for the interpreter to move up towards the representational and conative dimension of the interpreted action by ascribing a well-formed belief-desire pair to the agent. But from this it does not follow that ordinary interpretation in terms of reasons necessarily consists in the attribution of such belief-desire pairs to the agent. To the contrary, we have argued that it is more accurate to conceive of action interpretation as starting from a factive baseline.

We have shown how the attribution of a desire can be viewed as specifying the intrapersonal nature of a practical inferential pattern and the attribution of a belief as specifying the doxastic commitments interpreters take to be idiosyncratic to the agent. Thus we suggested that it is the *function* of the ascription of beliefs and desires in social practice to individualize doxastic commitments and practical inferential patterns, respectively. This function is revealed in situ whenever doxastic and practical commitments are being challenged in reason discourse. Getting a hold on the function of attitude ascriptions, folk psychological interpreters in effect acquire a new strategy for playing the game of giving and asking for reasons. This is a strategy that, once acquired, can in principle be applied across the board: challenge entitlement to every doxastic and practical commitment undertaken by the agent in answering questions why. The result will be always be *complete individualization* of the agent's reasons. We suggest that the implication of motivating BD-pairs by the reasons attributed to the agent, mirrors the fact that proceeding in a game of giving and asking for reasons towards the *most advanced stage possible*, always yields complete individualization of those reasons. Again, this is something most competent interpreters know how to do *in principle*. Turning it into a practical requirement demands a rather far-fetched scenario. Consider the quite baffling action of someone swallowing an acorn (cf. Hutto 2008, p. 7). A fierce round of giving and asking for reasons follows. Suppose this is the result: the agent swallowed the acorn because he believed it to be an alien, wanted to go to a galaxy far away and thought that the alien would take him there once it had absorbed his life spirit. We have made sense of the action, if only in the 'somewhat anaemic sense in which every rationalization justifies: from the agent's point of view there was, when he acted, something to be said for the action.' (Davidson, ibid. p. 9). In insisting on further reasons we tried to bring the action back within reach of the normal, the shared practical reality. Instead, we have particularized the agent's reason completely; every piece of the scaffolding has lead us further away from our common ground.

In most situations, however, agents are able to bring their actions back within reach of the normal, thereby, as Bruner (1990, p. 47) puts it, forging 'links between the exceptional and the ordinary'. The capacity to individuate actions in terms of beliefs and desires may play an important role in this process. Competent folk psychological interpreters are able to step up a level in the game of giving and asking for reasons when 'factive' pieces of the deontic scaffolding no longer suffice.

Particularizing deontic contexts in conative or representational terms helps to highlight the individual differences between people. Keeping track of these differences will help to interpret them at later times. A good memory then obviates the need to go through similar rounds of giving and asking for reasons and enables the interpreter to make sense of the agent without verbal interaction.

Deontic Scorekeeping from a Developmental Perspective

We have used Brandom's DS-model in order to characterize an important *explanandum* of folk psychology: the interpretation of other people's performances in terms of their reasons for action. In this section, we will show that adopting the DS-model also has important consequences for the description of the ontogeny of action interpretation and allows for a reinterpretation of some of the developmental evidence.

On the BD model of action interpretation, children are not introduced into the space of reasons before they start to get a proper hold on the propositional attitude concepts of belief and desire in the process of rationalizing the agent's action. It is widely accepted that passing the false belief test (cf. Wimmer and Perner 1983; see also Baron-Cohen et al. 1985; Perner 1991) is a reliable indicator that infants have acquired the concept of belief. Passing this test has often been taken as the final developmental hurdle for the child's acquisition of a theory of mind. But this cannot be the end of the story according to a belief-desire model of action interpretation. For, as Hutto (2008, p. 26) correctly points out, having an understanding of belief (in certain experimental setups) does not, on this model, 'equate to ascribing to X a reason: that would require ascribing to X a complex state of mind, minimally consisting of a belief/desire pair with interlocking contents.' Children must learn how these propositional attitudes and their contents interlock to form proper reasons for action. And it is unlikely that 4-year-olds have reached this level of sophistication, considering, for example, that 'research that explores whether 5-year-olds can use simple false belief knowledge to make inferences about their own and other's perspective finds that they singularly fail to do so.' (Carpendale and Lewis 2004, p. 91). Understanding and ascribing reasons by attributing appropriately structured belief-desire pairs apparently takes some extra years.

On the DS-model, however, it seems very plausible that children already start participating in the game of giving and asking for reason before the age of 4. 3-year-olds are quite possibly already able to understand and appreciate reasons for actions in factive contexts, as in the original grizzly scenario depicted by Gordon. They also appear to be perfectly capable of asking why-questions regarding the performances of others and understand a limited array of factive answers given in return. While these children are not yet able to use the concepts of belief in order to distinguish between their own doxastic commitments and incompatible commitments of the agent they are interpreting and while they might not yet have the capacity to ascribe desires and thereby make explicit individual-specific inferential patterns that conflict with conclusions of their own practical reasoning, they already seem to be in the position to follow through certain factive considerations of others and discern the reasons they act upon.

At the same time, some developmental stepping stones must be in place in order for children to exercise the capacity for factive reason interpretation. To interpret other agents as acting on certain facts that feature in a world that is fundamentally shared, children have to meet two important requirements. In the first place, they have to be able to respond specifically and differentially to other human agents. There is much evidence that this capacity is already operative from the moment of birth. Imitation studies by Meltzoff and Moore, for example, have demonstrated that neonates are able to pick out a human face from the crowd of objects in its environment, with sufficient detail that will enable it to imitate the gesture it sees on that face (Meltzoff and Moore 1977, 1994). They suggest that the action of neonates and their perception of the action of others are coded in the same 'language', a cross-modal system that is directly attuned to the actions and gestures of other humans. This hypothesis from developmental science is compatible with the neuroscientific discovery of mirror neurons (e.g. Rizzolatti et al. 2001; Gallese 2001, 2005; Iacoboni 2005, 2008) and proposals in cognitive science about action coding (e.g. Prinz 2002). Findings such as these show that neonates and very young infants are already capable of individuating other agents and interact with them dyadically in several ways.

But the infants' ability to perceive other agents as differentiated from the rest of the world clearly does not amount to an understanding of things in the world as possible objects of (shared) attention and action. It is generally accepted that such notion only starts to emerge when infants start to interact with other agents in a triadic, world-involving way, thereby entering the truly social realm of 'secondary intersubjectivity' (Trevarthen and Hubley 1978). Experimental findings suggest that, by 12 months of age, infants 'understand that actors try to achieve goals, that they keep trying persistently after failed attempts and accidents and around obstacles, and that when they succeed they stop acting toward the goal'. (Tomasello et al. 2005. p. 679; see also Gergely and Csibra 2003). Moreover, they begin to use communicative gestures to direct adult attention and behavior to outside objects in which they are interested in themselves. The pointing gesture, for example, enables the infant to declare his interest in specific features of his environment, and the clear, conventional nature of the gesture ensures being easily understood by other people (Werner and Kaplan 1963). Infants not only flexibly and reliably look where adults are looking (gaze following), but also start to use caretaker's facial expressions of emotion as a kind of communication that is known as 'social referencing' (Campos and Stenberg 1981; Phillips et al. 1992).

Yet these capacities are not yet sufficient for regarding an agent as 'reason responsive'. The infant may be capable of discerning a limited array of means-ends relations with respect to certain rather 'proximal' goal-directed actions, and in this sense he or she may already have acquired a sensitivity to the appropriateness or inappropriateness of certain performances by others. But this does not add up to the appreciation of the normativity of reasons proper. What we have in mind here is the normativity exposed by practical inferential patterns such as heading for the table when dinner is ready or putting your shoes on when you go outside. Such inferential patterns reveal (possible) reasons for action in situ: that dinner is ready is a reason to (form the intention to) head for the table; going outside is a *reason* to put on your shoes. These reasonable inferential patterns are also taught to the infant in straightforward factive contexts, in *real time*. It is *when* the infant is going out, *when* dinner is ready that his caregiver helps him with his shoes or beckons him to the table.

In order to appreciate such complex inferential patterns the infant arguably requires a fair degree of linguistic competence. Infants probably won't display sensitivity to reasons before their second birthday, when their linguistic capacities finally start to take off. For the infant to get hold on reasons for action, he or she needs to have some understanding of *what can be said* for (or against) the agent's performances in the practice of giving and asking for reasons. It is by being introduced to and participating in this practice, we suggest, that the infant gets pulled up into the space of reasons, so to speak, and acquires the notion of a reason for action.

The DS-model suggests that the younger child's capacity to interpret actions in terms of reasons is severely restricted in the sense that it is applied *successfully* only in rather straightforward factive contexts. The acquisition of the psychological concepts of belief and desire enables the child to vastly expand and improve his abilities to interpret the agent's reasons by opening up new ways of getting at those reasons. This not only requires a further development of linguistic skills, but also actively taking part in rounds of giving and asking for reasons.

Consider the emergence of false belief understanding, for instance. Several longitudinal studies reveal a strong correlation between the understanding of false belief and linguistic competence (Dunn et al. 1991; Astington and Jenkins 1999; de Villiers and de Villiers 2000; Watson et al. 2002; Farrar and Maag 2002). Other evidence comes from training studies, which usually begin with children who show little or no understanding of false belief, and then systematically expose them to some kind of training involving language for several days. Afterwards, the children are given one or more post-tests of false belief understanding (cf. Appleton and Reddy 1996; Swettenham 1996; Slaughter and Gopnik 1996; Slaughter 1998; McGregor et al. 1998).

Although these studies obviously show that language plays a role in false belief understanding, there are several hypotheses about the specific *nature* of this role. Moore et al. (1990), for example, point out children's mastery of the *semantics* of mental state terms such as think, know and believe, emerge in roughly the same age period as their mastery of the various false belief tasks. Hale and Tager-Flusberg (2003) and Lohmann and Tomasello (2003) have shown in two training studies that learning the *syntactic* structures required for using folk psychological concepts has a significant effect on children's performance on false belief tasks.

Most interesting for the purpose of this article and our introduction of the DS model, however, is Harris' (1996, 1999) proposal that the *process of linguistic interchange itself*, that is, the active participation in reason discourse, has a key influence on the understanding of false belief. The basic idea is that the notion of belief (as a propositional attitude) only becomes comprehensible in the context of alternative possible beliefs about a situation, including one that is true (implying that others may be false). Harris argued that it is in the to and fro of reason discourse that the child comes to understand that other people might know things he or she does not know, that they do not know things they ought to know, and that they do have different perspectives on things. In a similar vein, Tomassello (1999) and (Siegal 1999) refer to those forms of reason discourse in which misunderstandings and requests for clarification function as particularly clear signals of people's different perspectives and understandings of situations. Support for a 'discourse-based' view of folk psychological understanding comes from training studies that involve rich discourse interaction (Appleton and Reddy 1996) and correlation-studies in which deaf children,

who were given the opportunity to engage in richer discourse interactions with others, were also more successful on false belief tasks (Peterson and Siegal 1999, 2000).

Although these findings provide us with valuable insight in the development of deontic scaffolding, they shouldn't distract us from the main message of the DS-model: that children may already be able to make sense of others in terms of reasons before they acquire the concepts of belief and desire and apply them in the service of more sophisticated forms of deontic scorekeeping. The scaffolding of interpretation that takes place on a daily basis then finds its counterpart on the developmental level: both start from a factive baseline and proceed to more advanced forms of interpretation with the support of others.

References

Appleton, M., & Reddy, V. (1996). Teaching three year-olds to pass false belief tests: a conversational approach. *Social Development, 5*, 275–291.

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relationship between language and theory-of-mind development. *Developmental Psychology, 35*, 1311–1320.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*, 37–46.

Borg, E. (2007). If mirror neurons are the answer, what was the question? *Journal of Consciousness Studies, 14*, 5–19.

Braddon-Mitchell, D., & Jackson, F. (2007). *Philosophy of mind and cognition*. Oxford: Blackwell.

Brandom, R. B. (1994). *Making it explicit*. Cambridge: Harvard University Press.

Brandom, R. B. (2000). *Articulating reasons: An introduction to inferentialism*. Cambridge: Harvard University Press.

Bruner, J. (1990). *Acts of meaning*. Cambridge: Harvard University Press.

Bruner, J., & Kalmar, D. A. (1998). Narrative and metanarrative in the construction of self. In M. Ferrari & R. J. Sternberg (Eds.), *Self-awareness: Its nature and development* (pp. 308–331). New York: Guilford.

Campos, J., & Stenberg, C. (1981). Perception, appraisal, and emotion: The onset of social referencing. In M. E. Lamb & L. R. Sherrod (Eds.), *Infant social cognition*. Hillsdale: Erlbaum.

Carpendale, J. L. M., & Lewis, C. (2004). Constructing and understanding the mind: the development of children's social understanding within social interaction. *Behavioral and Brain Sciences, 79*, 79–151.

Carruthers, P. (2006). *The architecture of mind*. Oxford: Oxford University Press.

Currie, G., & Sterelny, K. (2000). How to think about the modularity of mind-reading. *Philosophical Quarterly, 50*, 145–160.

Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy, 60*, 685–700. Reprinted in Davidson 2001.

Davidson, D. (1970). Mental events. In L. Foster & J. W. Swanson (Eds.), *Experience and theory*. Amherst: University of Massachusetts Press. Reprinted in Davidson 2001.

Davidson, D. (2001). *Essays on actions and events*. New York: Oxford University Press.

De Villiers, J., & De Villiers, P. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 191–228). Hove: Psychology.

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Younblade, L. (1991). Young children's understanding of other people's feelings and beliefs: individual differences and their antecendents. *Child development, 62*, 1352–1366.

Farrar, M. J., & Maag, L. (2002). Early language development and the emergence of a theory of mind. *First language, 22*, 197–213.

Fodor, J. (1992). A theory of the child's theory of mind. *Cognition, 44*, 283–296.

Frith, U., & Happé, F. (1999). Theory of mind and self consciousness: what is it like to be autistic? *Mind and Language, 14*, 1–22.

Gallagher, S. (2001). The practice of mind: theory, simulation, or interaction? *Journal of Consciousness Studies, 8*, 83–107.

Gallagher, S. (2007). Simulation trouble. *Social Neuroscience, 2*, 353–365.

Gallagher, S., & Hutto, D. (2008). Understanding others through primary interaction and narrative practice. In J. Zlatev, T. Racine, C. Sinha, & E. Itkonen (Eds.), *The shared mind: Perspectives on intersubjectivity* (pp. 17–38). Amsterdam: Benjamins.

Gallagher, S., & Zahavi, D. (2008). *The phenomenological mind: An introduction to philosophy of mind and cognitive science*. London: Routledge.

Gallese, V. (2001). The 'shared manifold' hypothesis: from mirror neurons to empathy. *Journal of Consciousness Studies, 8*, 33–50.

Gallese, V. (2005). 'Being like me': Self-other identity, mirror neurons and empathy. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From cognitive neuroscience to social science: Mechanisms of imitation and imitation in animals* (pp. 101–118). Cambridge: MIT.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: the naive theory of rational action. *Trends in Cognitive Sciences, 7*, 287–292.

Goldman, A. (1989). Interpretation psychologized. *Mind and Language, 4*, 161–185.

Goldman, A. (2006). *Simulating minds: The philosophy, psychology and neuroscience of mindreading*. New York: Oxford University Press.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge: MIT.

Gordon, R. M. (1986). Folk psychology as simulation. *Mind and Language, 1*, 158–171.

Gordon, R. M. (1992). The simulation theory: objections and misconceptions. *Mind and Language, 7*, 11–34.

Gordon, R. M. (2000). Simulation and the explanation of action. In H. H. Kögler & K. R. Stueber (Eds.), *Empathy and agency: The problem of understanding in the social sciences* (pp. 62–82). Boulder: Westview.

Gordon, R. M. (2001). Simulation and reason explanation: the radical view. *Philosophical Topics, 29*, 175–192.

Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: a training study. *Developmental Science, 61*, 346–359.

Harris, P. (1996). Desires, beliefs and language. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 200–220). Cambridge: Cambridge University Press.

Harris, P. (1999). Acquiring the art of conversation: Children's developing conception of their conversational partner. In M. Bennett (Ed.), *Developmental psychology: Achievements and prospects*. Philadelphia: Psychology.

Heal, J. (1986). Replication and functionalism. In J. Butterfield (Ed.), *Language, mind, and logic* (pp. 135–150). Cambridge: Cambridge University Press.

Heal, J. (1998). Co-cognition and off-line simulation: two ways of understanding the simulation approach. *Mind & Language, 14*, 477–98.

Herschbach, M. (2008). Folk psychological and phenomenological accounts of social perception. *Philosophical Explorations, 11*(3), 223–235.

Hutto, D. D. (1999). A cause for concern: reasons, causes and explanations. *Philosophy and Phenomenological Research, 59*, 381–401.

Hutto, D. D. (2008). *Folk psychological narratives: The sociocultural basis of understanding reasons*. Cambridge: MIT.

Hutto, D. D. (forthcoming). Presumptuous naturalism: a cautionary tale. *American Philosophical Quarterly*.

Iacoboni, M. (2005). Understanding others: Imitation, language, empathy. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From cognitive neuroscience to social science: Mechanisms of imitation and imitation in animals* (pp. 77–99). Cambridge: MIT.

Iacoboni, M. (2008). *Mirroring people: The new science of how we connect with others*. New York, NY: Farrar, Straus & Giroux.

Leslie, A. M., German, Y., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology, 50*, 45–85.

Lohmann, H., & Tomasello, M. (2003). The role of language in the development in false belief understanding: a training study. *Child Development, 74*, 1130–1144.

Maibom, H. (2003). The mindreader and the scientist. *Mind & Language, 18*, 296–315.

McGregor, E., Whiten, A., & Blackburn, P. (1998). Teaching theory of mind by highlighting intention and illustration thoughts: a comparison of their effectiveness with 3-year-olds and autistic individuals. *British Journal of Developmental Psychology, 16*, 281–300.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*, 75–78.

Meltzoff, A. N., & Moore, M. K. (1994). Imitation, memory and the representation of persons. *Infant Behavior and Development, 17*, 83–99.

Moore, C., Pure, K., & Furrow, D. (1990). Children's understanding of the modal expression of speakers certainty and uncertainty and its relation to the development of represenational theory of mind. *Child Development, 61*, 722–730.

Nichols, S., & Stich, S. (2003). *Mindreading. An integrated account of pretence, self-awareness, and understanding of other minds*. Oxford: Clarendon.

Perner, J. (1991). *Understanding the representational mind*. Cambridge: Bradford Books/MIT.

Peterson, C. C., & Siegal, M. (1999). Representing inner worlds: theories of mind in austistic, deaf and normal hearing children. *Psychological Science, 10*, 126–129.

Peterson, C. C., & Siegal, M. (2000). Insights into theory of mind from deafness and autism. *Mind & Language, 15*, 123–145.

Phillips, W., Baron-Cohen, S., & Rutter, M. (1992). The role of eye-contact in the detection of goals: evidence from normal toddlers, and children with autism or mental handicap. *Development and Psychopathology, 4*, 375–383.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*, 515–526.

Prinz, W. (2002). Experimental approaches to imitation. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind: Development, evolution, and brain bases* (pp. 143–162). Cambridge: Cambridge University Press.

Ratcliffe, M. (2007). *Rethinking commonsense psychology: A critique of folk psychology, theory of mind and simulation*. Basingstoke: Macmillan.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131–141.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience, 2*, 661–670.

Scholl, B. J., & Leslie, A. M. (1999). Modularity, development and 'theory of mind'. *Mind & Language, 14*, 131–153.

Sellars, W. (1953). Inference and meaning. *Mind, 62*, 313–338.

Siegal, M. (1999). Language and thought: the fundamental significance of conversational awareness for cognitive development. *Developmental Science, 2*, 1–34.

Slors, M. V. P. (2007). Intentional systems theory, mental causation and empathic resonance. *Erkenntnis, 67*, 321–336.

Slaughter, V. (1998). Children's understanding of pictatorial and mental representations. *Child development, 69*, 321–332.

Slaughter, V., & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: training children to understand belief. *Child development, 67*, 2767–2788.

Smith, M. (1987). The humean theory of motivation. *Mind, 96*, 36–61.

Spaulding, S. (2010). Embodied cognition and mindreading. *Mind & Language, 25*(1), 119–140.

Swettenham, J. (1996). Can children with autism be taught to understand false belief using computers. *Journal of Child Psychology and Psychiatry, 37*, 157–165.

Tomassello, M. (1999). *The cultural origins of human cognition*. Cambridge: Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences, 28*, 675–691.

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech* (pp. 321–72). Cambridge: Cambridge University Press.

Trevarthen, C., & Hubley, P. (1978). Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year. In A. Lock (Ed.), *Action, gesture and symbol: The emergence of language* (pp. 183–229). London: Academic.

Watson, A., Painter, J., & Bornstein, M. (2002). Longitudinal relations between 2-year-olds' language and 4-year-olds' theory of mind. *Journal of Cognition and Development, 2*, 449–457.

Werner, H., & Kaplan, B. (1963). *Symbol formation*. New York: Wiley.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103–128.