

Teleology and Mentalizing in the Explanation of Action

Uwe Peters

[1] Centre for Logic and Philosophy of Science, KU Leuven, Belgium

[2] Department of Economics, University College London, UK

[3] Department of Social and Educational Sciences, University of Applied Science Potsdam

[Penultimate draft. The paper is forthcoming in *Synthese*. Comments very welcome.]

Abstract

In empirically informed research on action explanation, philosophers and developmental psychologists have recently proposed a teleological account of the way in which we make sense of people's intentional behavior. It holds that we typically don't explain an agent's action by appealing to her mental states but by referring to the objective, publically accessible facts of the world that count in favor of performing the action so as to achieve a certain goal. Advocates of the teleological account claim that this strategy is our main way of understanding people's actions. I argue that common motivations mentioned to support the teleological account are insufficient to sustain its generalization from children to adults. Moreover, social psychological studies, combined with theoretical considerations, suggest that we do not explain actions mainly by invoking publically accessible, reason-giving facts alone but by ascribing mental states to the agent.

Introduction

Suppose that a subject *S* is on her way to the station to take a train to work. She arrives late at the station, looks at the platform, sees a train is about to leave, and suddenly starts running towards it. What *S* is doing is an intentional action. How do we explain it?

According to the standard account of action explanation in philosophy and psychology, we explain what *S* is doing by appealing to the mental states, more specifically, the propositional attitudes (e.g., beliefs, desires etc.) that underlie her behavior (Perner and Roessler 2010). We explain her running toward the train by holding that she (1) *believes* that the train is leaving, (2) *believes* that running towards it will help her catch the train, and (3) *wants* to catch it (Anscombe 1957; Davidson 1963; Goldman 1970; Malle et al. 2007; for discussion, see Alvarez 2016).

While this mentalistic account of what *S* is doing can be viewed as the received proposal on action explanation, many theorists have criticized it and proposed alternatives (Dancy 2008; Alvarez 2010; McDowell 2013). One alternative that has recently gained momentum among philosophers and developmental psychologists is the *teleological account* (Gergely and Csibra 2003; Roessler and Perner 2013; Perner and Esken 2015; Perner et al. 2018).

The account comes in different flavors.¹ According to the version that will be relevant here, we typically do not explain intentional action by reference to the agent's mental states but by

¹ There is, for instance, a *cybernetics* account of teleology, and an *Aristotelian* one (Perner and Esken 2015). Gergely and Csibra (2003) propose the former. They understand goal pursuit in the way it is done in cybernetics: a system is built so as to respond to different circumstances in such a way that it ends up in a particular state (the goal state). Gergely and Csibra don't appeal to evaluative facts as the source of goals but take the notion of goals as primitive. In contrast, on the Aristotelian view, which is currently favored by most theorists working on the teleological

appealing to the objective, non-mental facts that count in favor of performing the action (Scanlon 1998), including the objective desirability of the goal that underlies the action and the facts that are instrumental to achieving the goal.² On this view, we would explain *S*'s running toward the train by holding that (1) the train is leaving (an objective fact), (2) running towards it will help her catch it (an objective instrumental fact), and (3) she has the goal to catch the train, as getting one's train to work is generally desirable (an objective evaluative fact). We wouldn't explain *S*'s behavior by *mentalizing*, that is, by thinking about and ascribing to her propositional attitudes, but by assuming that what she is doing is directed at a state of affairs that is 'objectively' good, which is her goal (i.e., catching the train to work), and by appealing to the non-mental facts that enable her to achieve that goal. This explanatory strategy has been called *teleology* (Perner and Roessler 2010; Perner and Esken 2015; Perner et al. 2018).

My focus here will be on the recent debate on teleology³ in the interdisciplinary area linking philosophical and psychological research. In that area, most of the work on teleology is related to developmental psychological data on action explanations in 3- to 4-year-olds. These children have been found to be able to explain intentional action even when they still lack a full understanding of mental states such as beliefs. Advocates of the teleological account argue that this is because they use teleology (Gergely and Csibra 2003; Perner and Roessler 2010; Perner and Esken 2015).

What matters here is that proponents of the account tend to make claims that go beyond developmental psychology and cognition in children. They maintain that the "teleological account [is of] more than developmental interest", that teleology also plays a "central role [in] adult common-sense psychology" (Roessler and Perner 2013: 521), and that the account captures our "standard", "routine", and "main way of understanding people's behavior" (Perner and Esken 2015: 74, 69, 76), the "way in which we ordinarily understand others' intentional actions" (Perner, Priewasser, and Roessler 2018: 99).

Surprisingly, argumentative support for such general claims about the use of teleology in action explanation is scant. This is surprising because it doesn't follow that even if young children typically use a particular method for explaining actions, adults continue to do so. It is fair to say that it is currently still unresolved as to whether and, if so, to what extent adults use teleology to explain actions.

Why should *philosophers* care? One reason is that answers to these questions will help us demarcate the scope of the teleological account, which will in turn advance our understanding of the nature of action explanation, a key topic of philosophical inquiry (Davidson 1963; Alvarez 2016). Additionally, the teleological account suggests that the mentalistic view of action explanation is false. Since that view is widely accepted in philosophy of action (ibid), if

account (and the one I'll focus on here), goals are inherently value-laden, and their explanatory force is inseparable from their evaluative properties; for discussion, see Perner and Esken (2015: 76f).

² The mentalistic view and the teleological proposal can also be distinguished by reference to the notions of *justifying reasons* and *motivating reasons*. A *justifying reason* for *A*-ing is any consideration that counts in favor of *A*-ing. A *motivating reason* is a consideration out of which an agent acts (Alvarez 2016). The mentalistic view is that while justifying reasons may be facts, motivating reasons are mental states. In contrast, the teleological account proposes that motivating reasons are typically facts too: what motivates *S* to act is (in standard situations) the fact that *p*, not her believing that *p*, and to explain her action, it is sufficient to notice a non-mental fact about her action that, in combination with an evaluative fact, counts in favor of acting that way.

³ Theorizing about teleology has a long tradition in philosophy, going back to Aristotle (Shields 2015) and continuing on in the contemporary literature on philosophy of action (Wilson 1989; Mele 2010; McLaughlin 2012). I shall not delve into this purely philosophical work here, however.

the teleological alternative turns out to be correct (for children and adults) then a significant revision in the philosophy of action would be needed. Finally, the teleological account has been developed, defended, and critiqued by philosophers and not only by psychologists (e.g., Roessler and Perner 2013; Perner and Esken 2015; Burge 2018). The following discussion on the role of teleology in action explanation should thus be of interest for both philosophers and psychologists working on action explanation.

I shall argue for two claims. First, common motivations that are mentioned to support the teleological account are insufficient to sustain its generalization from children to adults. Second, social psychological studies, combined with theoretical considerations, suggest that as adults we do *not* explain actions mainly by invoking reason-giving, non-mental facts alone but by ascribing mental states to the agent. That is, we don't predominantly use teleology in our action explanations but mentalizing.

I begin in section 1 by mentioning the reasons that researchers have offered for holding that children are teleologists when making sense of people's behavior. In section 2, I turn to claims about the use of teleology in adults and distinguish different versions of the teleological account that one might adopt. Upon isolating the one relevant here, I introduce three points that have been proposed to defend it and highlight shortcomings with them. After that, I argue against the version of the teleological account at issue.

1. Teleology in children's action explanation

The teleological account of action explanation that I want to focus on emerged in part from the theorizing on children's performance in the classic⁴ false-belief task (FBT) (Wimmer and Perner 1983; Perner and Roessler 2010; Perner et al. 2018). To introduce the account, it will be useful to briefly outline the general structure of and key findings from the FBT.⁵

Schematically, in the task, children are asked to observe two individuals, *S* and *T*, in a room with two boxes, *B1* and *B2*. *S* has an object, *o*, puts it into *B1*, and leaves the room. While she is outside, *T* takes *o* out of *B1*, puts it into *B2*, and leaves the scene. *S* returns, and the children are asked where she will look for *o*. Studies found that 3-year-olds consistently say that *S* will go to *B2*, where *o* in fact is. It is not until they are 4 that neurotypical children pass the test and correctly say that *S* will look in *B1*, where she falsely believes *o* to be (Wellman et al. 2001).

Advocates of the teleological account argue that since 3-year-olds reliably predict that *S* will look for *o* in *B2* (rather than, say, under the kitchen table), these children assume that people's behavior is determined by factors counting in favor of acting. The idea is that the children think that since *S* wants to find *o*, looking in *B2* is what she should be doing. After all, that is where *o* is and where she will find it. Children assume that *S* has thus a good reason to go to *B2*. Or so advocates of the teleological account propose (Perner and Roessler 2010; Perner et al. 2018).

These researchers distinguish between two kinds of reasons that subjects might invoke: *objective reasons* and *subjective reasons* (ibid). Objective reasons are taken to be publically accessible facts of the world that count in favor of acting in a certain way. They include *instrumental facts*, which

⁴ The 'classic' FBT requires test subjects to produce verbal reports. More recent versions of the task are non-verbal and involve, for instance, looking time measures to assess mental-state understanding. In such FBTs, already 7-months-olds have been found to exhibit some understanding of false beliefs (see Schneider et al. 2015).

⁵ The term 'FBT' refers here and henceforth only to the classic version of the FBT (as opposed to tasks using, e.g., looking time measures).

pertain to the means to achieve a certain goal, and *evaluative facts*, which pertain to what is objectively important, desirable, necessary, or good for the agent, where the term ‘objectively’ is construed as ‘independently of one’s subjective viewpoint’ (e.g., catching one’s train to work is objectively good) (Roessler and Perner 2013: 37; Perner and Roessler 2010: 204). Advocates of the teleological account hold that in the FBT, *S* has an objective reason to go to *B2* because that is where *o* is and so for her to get. In contrast, “*subjective* reasons are the contents of mental states like beliefs and desires” (Perner and Esken 2015: 71). “To say that you have a subjective reason to [φ] is to say that, from your perspective, it looks as if you have an (objective) reason to do so” (Perner and Roessler 2010: 204). In the FBT, *S* has a subjective reason to go to *B1* because that is where *o* is from her own perspective (Perner and Roessler 2012: 521).⁶

Advocates of the teleological account propose that 3-year-olds’ normative attitude toward an agent’s behavior is based only on an understanding of objective reasons (Perner and Roessler 2010: 205). They grant that 3-year-olds also need to have an understanding of what *S* *wants* in order to be able to judge that she will go to *B2*. But they maintain that this doesn’t undermine the teleological account, for the notion of desire that these children have needn’t involve the assumption that wanting is a subjective, mental state. For instance, when 3-year-olds assume that *S* wants to find *o*, these children might take her wanting to do so to be “provided by the objective desirability” of finding *o*: it is generally desirable to find what one is looking for (ibid: 214). Alternatively, they might view *S* as wanting to find *o* in that she *needs* to do so, which is again specifiable without reference to mental states (ibid).

The teleological account does, however, seem to face significant problems with scenarios in which action is based on a *false* belief. This is because no non-mental fact then exists that could be a subject’s objective reason for acting. To deal with these cases, advocates of the teleological account hold that when, for instance, *S* proceeds to *B1* (which doesn’t contain *o*), children can explain *S*’s behavior by employing “teleology within [*S*’s] perspective”, interpreting her as acting on the basis of what *from her perspective* appears to be an objective reason (Roessler and Perner 2013: 46; Perner et al. 2018: 100). The transition from “pure teleology” (henceforth *PT*), which is the ability to explain behavior by appeal to objective reasons alone, to this “teleology-in-perspective” (henceforth *TiP*), which involves appealing to an agent’s subjective reason(s), is thought to happen at around age 4 (ibid).

Crucially, advocates of the teleological account tend to maintain that *TiP* doesn’t yet require mentalizing, that is, thinking about and ascribing propositional attitudes such as beliefs and desires. It is taken to rest only a more general understanding that “having informational access to the relevant facts is a prerequisite for action” (Perner and Esken 2015: 76), or an ability to engage in “counterfactual thinking” (e.g., ‘If *o* were still in its original location then there would be a good reason for *S* to go there to get it.’) (Roessler and Perner 2013: 37f). In fact, if advocates of the teleological account held that *TiP* does already involve mentalizing then their proposal would, when it comes to *TiP*, coincide with the standard mentalistic account of action explanation, which would be in tension with the overall anti-mentalistic spirit of the teleological approach. The proposal is thus different. It is that when employing *TiP*, subjects don’t yet conceptualize the agent’s subjective reasons as propositional attitudes but merely as

⁶ In holding that *S* has a subjective reason to go to *B1* because from her perspective that is where *o* is, i.e., where she believes it to be, advocates of the teleological account propose that non-factive mental states can be subjective reasons. Their notion of a subjective reason is thus different from that of certain externalist views about rationality, according to which what is rational to do is determined by what one *knows* rather than merely believes. Advocates of a factive account of subjective reasons might disagree with proponents of the teleological account. Since I don’t endorse the teleological account but only aim to offer an exposition of it here, I shall set the issue aside.

“perspectival” facts, as objective reasons from within the agent’s perspective (ibid; Perner et al. 2018: 107f)). To capture this *TiP*-specific point about the teleological account, I will use the term ‘subjective reason’ to refer only to perspectival facts, not to propositional attitudes conceptualized as such by the explainer. As will become clear in a moment, this non-mentalistic reading of *TiP* is well supported by claims that advocates of the teleological account make about teleology in adults’ action explanations.

2. From children to adults – Variants of the teleological account

So far I have focused mostly on the use of teleology in children, but, as noted, advocates of the teleological account maintain that their proposal equally holds for adults. Before introducing their specific view on the matter, it will be useful to distinguish different kinds of the teleological account applied to adults. Four proposals come to mind:

- (1) Adults explain action by ascribing propositional attitudes to agents and by relying *inter alia* on teleological principles linking objective or subjective reasons to action but they never explain action in terms of objective or subjective reasons alone.
- (2) Adults mainly explain action by ascribing propositional attitudes to agents, and rarely do so by invoking only objective or subjective reasons.
- (3) Adults often explain action by ascribing propositional attitudes to agents, and often do so by invoking only objective or subjective reasons.
- (4) Adults mainly explain action by invoking only objective or subjective reasons, and rarely do so by ascribing propositional attitudes to agents.

Proposal (1) is much in line with the received mentalistic view of action explanation. It only adds to it that adults might also use teleological principles to make sense of people’s behavior. Proposals (2) and (3) are modest too in that they grant that people frequently explain actions by mentalizing. I shall set (1), (2), and (3) aside for now. I want to focus on proposal (4), because it is the most ambitious type of teleological account (henceforth the sole referent of ‘teleological account’), and the one that is currently endorsed by most advocates of the view that teleology is involved in action explanation. Here is some textual evidence.

After mentioning an example in which an action is explained by reference to objective facts alone (i.e., ‘Bernhard the baker gets up at 3 am, because he has to have the bread ready by 7, and he can only deliver it by 7 if he gets up at 3.’), Perner and Esken (2015) write,

When describing the case of Bernhard the baker, we naturally provided reasons for explaining his unusual behavior of getting up at 3 in the morning. This made it a natural explanation, which makes us think that teleology is our standard way for making sense of what people are doing. The use of a theory of mental states as assumed in ‘mentalizing’, or imaginative identification as assumed in simulation theory, remain the techniques for special cases. (74)

The claim is that the view that we typically use mentalizing in action explanations “leads to an over-intellectualization of our everyday explanations of others’ actions”:

We do not see ourselves as being made to act by unobservable internal states. We see ourselves as acting for reasons, which are provided by publicly accessible circumstances in our surroundings. And so – our claim – we consider others to be and act like us. (Perner and Esken 2015: 74)

Similarly, Perner et al. (2018) note that a “teleologist is concerned primarily with worldly facts” (101-102). The idea is that even when *TiP* is at issue, “for this, the teleologist need not construe people as entertaining mental representations of events in the world. Mentalizing only becomes relevant for covering unusual cases” (Perner and Esken 2015: 76).

In short, the proposal is that teleology (*PT* or *TiP*) is distinct from (implicit or explicit) mentalizing in that it doesn’t involve viewing agents as acting on the basis of mental states, and it plays a “central role [in] adult common-sense psychology” (Roessler and Perner 2012: 521), constituting our “routine”, “main way of understanding people’s behavior” (Perner and Esken 2015: 76), “the way in which we ordinarily understand others’ intentional actions” (Perner et al. 2018: 99). This interesting and, if true, explanatorily powerful proposal will be my target from now on. Notice that while the responses of 3-year-olds in FBTs, which I mentioned in the preceding section, might provide good grounds to hold that these children use *PT*, they don’t yet show that *PT* or *TiP* is also involved in action explanation in *adults*. So what is the support that advocates of the teleological account offer for their generalization to adults?

3. Arguments for teleology in adults

To motivate the view that adults use teleology in their action explanations, advocates of the teleological account have mentioned different points. Interestingly, none of them pertains to adults’ use of *TiP* and subjective reasons. They relate only to *PT*, and adults’ reference to objective reasons in their explanations of actions. I shall briefly discuss what I take to be the three main points.

Argument 1. It is natural and familiar among adults to explain, for instance, *S*’s running toward her train to work by saying that the train is leaving, she needs to catch it, and running towards the train helps her achieve that. That is, it is common among adults to refer in their reports of action explanations to non-mental facts, which suggests that they use objective reasons in their action explanation (Perner and Esken 2015: 74).

Response. When adults mention objective facts in their reports on why *S* is acting in the way she is, this is arguably just shorthand. For there is reason to believe that implicit inferences about *S*’s mental states are involved in forming the action explanation. Consider the following three scenarios:

- (1) Why is *S* running toward the train?
 - (1a) Because that’s the way to catch the train.
 - (1b) Because that’s the way to catch the train, but she doesn’t believe it.

- (2) Why is *S* calling her boss?
 - (2a) To give notice.
 - (2b) To give notice, but she doesn’t intend to give notice.

- (3) Why is *S* working 7 days a week?
 - (3a) To make more money.
 - (3b) To make more money, but she doesn’t want to make more money.

While statements (1a), (2a), and (3a) are acceptable action explanations that cohere with and support the teleological account, (1b), (2b), and (3b) seem odd in ways that may bring to mind “Moore’s paradox”, which is captured in assertions of the form ‘*p*, but I don’t believe that *p*’

(Moore 1993: 207f). Because if *S* needs to catch the train but doesn't *believe* that running toward it is the way to do so, it becomes difficult to see what could lead her to act that way. Similarly, learning that *S* doesn't intend to give notice casts doubt on the view that she is calling her boss *to give notice*, just as much as learning that she doesn't want to make more money challenges the view that she is working 7 days *to make more money*.⁷ These statements might be interpreted in ways that make them tenable, but they do seem odd and violate expectations. This suggests that expressions of action explanations that appear to support a teleological account (i.e., (1a), (2a), and (3a)) are partly based on an implicit ascription of propositional attitudes. I shall refer to this strategy of revealing underlying implicit assumptions as the *intelligibility test* for action explanations. The test will become relevant again below. For now, the point is just that it suggests that citing expressions such as (1a), (2a), or (3a) isn't sufficient to show that adults explain behavior via *PT* or *TiP*.

Argument 2. In our own case, we typically determine and explain what we are doing by appealing to the non-mental facts that we assume will allow us to achieve our goals. For instance,

Suppose you suddenly remember that there is a staff meeting at 2 pm. Should you leave your office now to go to the meeting? It would not be sensible for you to try to settle this question by looking for evidence as to whether you *believe* it to be 2 pm. Whether you have a good reason to set off now depends on what the actual time is. (Perner and Roessler 2012: 521)

This, advocates of the teleological account argue, suggests that even as adults we use *PT* in action explanations (ibid).

Response. Notice first that the claim about the involvement of teleology in adults' action explanations is meant to hold for *both* one's own and other people's actions. Argument 2, however, focuses only on first-person action explanations. And even if we accept that we do explain what we ourselves are doing in terms of the objective reasons⁸ (i.e., the instrumental and evaluative facts) that allow us to achieve our goals and figure in our practical reasoning, it isn't obvious, and the point at issue doesn't show, that we also do so in the case of *other people's* actions.

Argument 3. When we are asked to offer practical advice to *S*, the natural way to proceed is to reflect on the facts of her situation. For instance, if *S* asked you where she should go to get *o*, it wouldn't be very helpful to say 'You should go to *B1*; after all, you believe that is where *o* is.' The more natural advice seems to be 'You should go to *B2*. You have reason to, that is where *o*

⁷ Unger (1978: 258f) has used a similar strategy (though focusing mostly on first-personal sentences) to argue that the oddness of Moorean assertions such as '*p*, but I'm not absolutely sure that *p*' or '*p*, but I don't know that *p*' indicates the presence of attitudes toward *p* other than belief (in the speaker), i.e., certainty or knowledge. I'm grateful to a reviewer of this journal for drawing my attention to Unger's work, and agree that it is worth thinking more about Moore-paradoxical sentences to explore the implications of such less noticed versions of them for our epistemological theorizing. The point also holds for the third-personal cases I'm considering here. Suppose, for instance, that in response to 'Why did Jane call her husband?' we hear 'Because she was late for dinner, but she didn't know it'. That response sounds problematic which suggests that in giving the reason that she was late we assume that Jane didn't only believe she was late but also knew it. This points to an interesting extension of my argument here: an analysis of the linguistic expressions of action explanations might suggest that actions require not only beliefs but also knowledge (Hyman 2015).

⁸ There is ground to reject the antecedent. Consider: 'Why are you running toward the train?' (i) 'Because it's leaving.' (ii) 'Because it's leaving, but I don't believe that it's leaving.' The oddness of (ii) seems to indicate an implicit self-ascription of a belief.

is.’ The structure of your advice suggests that people invoke objective reasons in their thinking about others’ actions (ibid).

Response. The point illustrates only the role that objective reasons might play in giving people advice. It is insufficient to support the view that in order to *explain* people’s actions, we tend to appeal to these reasons also. Giving people advice and explaining their actions are two different things. And for all that the mentioned consideration shows, we wouldn’t appeal to objective reasons in our explanations of others’ actions unless we (implicitly or explicitly) assumed that they believe the facts constituting these reasons.

In sum, more support is needed for the claim that adults explain action in terms of objective reasons. In fact, the teleological account is so far not only insufficiently supported, it is also directly challenged by empirical data. This is what I shall argue next.

4. A problem for the teleological account

If the teleological account were correct then people shouldn’t predominantly invoke an agent’s beliefs and desires in their action explanations. The verbal reports capturing these explanations should contain only an infrequent reference to beliefs and desires. One discipline in which reports of action explanations have been systematically gathered and analyzed is social psychology. To examine the tenability of the teleological account and explore the frequency of action explanations involving belief and desire ascriptions, it will thus be useful to consider social psychological data.

4.1 Social psychology, reasons, and action explanations

Bertram Malle, Joshua Knobe, and colleagues have done much experimental work on action explanations in adults (Malle 1999; Knobe and Malle 2002; Malle et al. 2005; Malle et al. 2007). To find out how people explain actions, they presented their test subjects with different kinds of behavior and then asked them for explanations of them. The behaviors were selected by the participants themselves, derived from their spontaneous interactions, or predetermined by the experimenters. It turned out that for intentional actions, people tended to use one of the following three modes of explanation. They mentioned, what Malle et al. call, (1) *reason explanations*, (2) *causal history of reason (CHR) explanations*, or (3) *enabling factor explanations*.

On Malle et al.’s view, “[w]hen people cite a *reason explanation* for an action, they ascribe to the agent one or more beliefs or desires that (they presume) figured in the agent’s decision to so act” (2007: 493). Reason explanations are formed via “taking the perspective” of the agent so as to work out what led him/her to act (ibid). For instance, ‘Kim chose not to vote in the last election, because she believed that none of the candidates was trustworthy’ was categorized as a reason explanation. In contrast, *CHR explanations* are explanations that don’t involve belief-desire ascriptions, don’t pick out factors that figure in the agent’s practical inference motivating the action, and don’t involve perspective taking (ibid). They refer to causal aspects external to the agent’s judgment- and decision-making, and include an agent’s disposition, personality, upbringing, culture, or context. For instance, ‘Why didn’t John vote?’ – ‘He’s too lazy’; ‘Why would you ignore somebody’s arguments?’ – ‘I’m stubborn’; ‘Why did Anne greet her uncle politely?’ – ‘Because that’s how she was taught to’ are all examples of CHR explanations (Malle 1999: 31). Finally, *enabling factor explanations* do “not clarify why the agent intended to act but rather how it was possible that the intention turned into a successful action. Presupposing that the agent had reasons and an intention to act, they cite factors that enabled the action to be

performed as intended. For example, ‘She worked through the night because she had a lot of coffee’” (Malle 2011b: 308).

The key point here is that in different studies, Malle and colleagues found that “[r]eason explanations are the most frequently used mode [of explanation], comprising about three-quarters of all action explanations” (2011a: 82). For instance, from over 700 participants and over 6000 explanations, on average, across five studies, people offered 80% reason explanations of their own behavior, and 60% reason explanations of other people’s behavior (Malle 2005: 165). While these studies involved ordinary behaviors, in experiments involving puzzling behaviors, the same dominance of reason explanations was discovered (Korman and Malle 2016).

These findings are relevant for assessing the teleological account, because, as noted, reason explanations are in Malle et al.’s framework explanations in which people appeal to an agent’s beliefs and/or desires, that is, to ‘unobservable’ mental states. And advocates of the teleological account claim that in explaining actions, “[w]e do *not* see ourselves as being made to act by *unobservable internal states* [but as] acting for reasons, which are provided by publicly accessible circumstances in our surroundings. And [...] we *consider others to be and act like us* [emphasis added]”: “Mentalizing only becomes relevant for covering unusual cases” of action explanation (Perner and Esken 2015: 74, 76). The claim is that we are “concerned primarily with worldly facts” rather than propositional attitudes when we are making sense of people’s behavior (Perner et al. 2018: 102). Malle et al.’s data strongly suggest otherwise, indicating that most action explanations in fact involve viewing the agent as acting on the basis of propositional attitudes.

Notice, however, that Malle et al. distinguished between two different kinds of reason explanations, between explanations involving *marked reasons* and explanations involving *unmarked reasons*. To illustrate the distinction, in the studies mentioned, the explanation ‘Jane called her husband because she *thought* she was late for dinner’ was categorized as a *marked* belief reason explanation, because it involves an explicit reference to a mental state. In contrast, ‘Jane called her husband because she was late for dinner’ was treated as an *unmarked* belief reason explanation, for it involves no explicit reference to a belief but nonetheless captures a proposition that Jane believed and that figured in her practical reasoning (Malle et al. 2007: 494, 509). Importantly, Malle and colleagues found that *unmarked* reason explanations were provided more frequently than marked ones (Malle 2011b: 325). This is important because advocates of the teleological account might now hold that the data at issue in fact *support* their proposal. They might argue that unmarked reasons are effectively objective reasons, non-mental facts that count in favor of acting (e.g., the fact that Jane was late for dinner etc.).

However, now the intelligibility test introduced above becomes relevant again. For consider the following example,

- (I) Why did Jane call her husband?
- (Ia) Because she was late for dinner.
- (Ib) Because she was late for dinner, but she didn’t believe it.

(Ia) seems acceptable, but (Ib) is problematic. If Jane didn’t believe that she was late for dinner, it becomes difficult to explain why she would still call her husband because she was late for dinner. The point is the same as the one mentioned above *inter alia* with respect to the example

- (1) Why is *S* running toward the train?
 (1a) Because that's the way to catch the train.
 (1b) Because that's the way to catch the train, but she doesn't believe it.

Since learning that *S* does not believe that running toward the train is the way to catch the train creates doubts about why she is running toward the train, (1a) arguably rests on an implicit assumption that *S* *does* believe that running toward the train is the way to catch it. That is, the oddness of both (1b) and (1b) suggests that underlying expressions such as (1a) and (1a) are ascriptions of beliefs, because it is the negation of the view that *S* and Jane hold particular beliefs which gives rise to the puzzlement with (1b) and (1b). And indeed, similar⁹ considerations led Malle et al. (2007: 493) to categorize expressions such as (1a) and (1a) still as belief reason explanations (albeit unmarked ones) in the first place.

Notice that it isn't promising for advocates of the teleological account to now respond that the puzzlement elicited by expressions such as (1b) and (1b) can be explained without postulating that the explainer ascribes full-fledged propositional attitudes such as beliefs, rather than only subjective reasons (i.e., perspectival facts). For it is specifically the use of the term 'believe' in (1b) and (1b) that triggers the intelligibility problem with these statements. If the explainer implicitly ascribed mental states that aren't yet conceptualized as beliefs then (1b) and (1b) shouldn't appear odd.

With these points in mind, since "reason explanations are the most frequently used mode [of action explanation]" (Malle 2011a: 82), and since these explanations explicitly or implicitly appeal to the agent's propositional attitudes rather than objective or subjective reasons (i.e., non-mental or perspectival facts), the tenability of the teleological account (as I specified it above) becomes questionable. If the account were correct, we would expect to find the opposite. Subjects should mostly invoke non-mental/perspectival facts to make sense of people's behavior and not (explicitly or implicitly) ascribe propositional attitudes. But they don't.

4.2 Actor-observer asymmetries and teleology

There is another way of assessing the tenability of the teleological account. It focuses on a claim that advocates of the account make on the relationship between explanations of one's own action and explanations of other people's actions. The claim is that the basic strategy that we use to explain our own action is much the same as the one that we employ to explain others' behavior. Advocates of the teleological account maintain that we assume that others act just like us for reasons that are provided by publically accessible circumstances (Perner and Esken 2015: 74):

Teleology starts from the close resemblance between the reasoning involved in understanding others' actions and one's own practical reasoning involved in deciding

⁹ Malle et al. (2007: 493) write that a "reason explanation becomes meaningless if the agent's awareness of the reason is denied, as in 'Anne invited Ben for dinner because he had helped her paint her room (even though she was not aware that he had helped her paint her room).' Thus, when people offer reason explanations, they make an assumption of subjectivity – they assume that the agent was aware of her reasons and acted on those subjectively held reasons (whether or not they reflect objective facts)." Malle et al.'s "subjectivity test" isn't yet sufficient to show that implicit *belief* ascriptions are involved in unmarked belief reason explanations, however, because it pertains only to an implicit ascription of awareness of reasons.

what to do. [...] The teleologist [...] does not imagine being in the other's situation. However, she needs to engage in the same evaluation of this situation and what needs doing to improve the situation as the other does, that is, the practical reasoning has to be essentially shared. [...] [Teleology] captures that we consider other people to act in essential ways for the same reason as we do [...]. (Perner et al. 2018: 99, 103, 111)

If, as these passages suggest, we explain people's action by reference to the same kind of objective reasons (i.e., instrumental and evaluative facts) that figure in our own practical reflection on what to do and in our explanations of our own acting, then there shouldn't be significant differences in the kind of explanations that we produce for our own actions and those that we produce for other people's action. Yet, data from another set of studies conducted by Malle, Knobe, and colleagues challenges this prediction too. Malle et al. (2007) conducted six experiments exploring whether people explain their own action (henceforth *actor perspective*) in the same way as they explain other people's actions (henceforth *observer perspective*). Subjects were presented with actions that either they themselves or someone else performed, and then they were asked to explain why they, or the other person, acted the way they did. Three strong and reliable actor-observer asymmetries emerged, what Malle et al. call, a (1) *reason asymmetry*, a (2) *belief asymmetry*, and a (3) *marker asymmetry*:

Reason asymmetry. Actors offered significantly more reason explanations than observers. Observers provided significantly more CHR explanations than actors.

Belief asymmetry. Within participants' reason explanations, actors used more belief reasons and fewer desire reasons than observers did.

Marker asymmetry. Actors left the mental states in their reason explanations more often unmarked than observers did (Malle et al. 2007: 495f, 507).

Prima facie, these asymmetries are inconsistent with the teleological account. After all, according to the account, agents' reasons for acting are not only "publically accessible", but also "essentially" the same in the case of one's own acting, and the case of other people's acting (Perner et al. 2018: 103). So if the account were correct, one would expect to find actor-observer *symmetries* rather than asymmetries. The point is perhaps most evident when it comes to *PT*, in particular. For by assumption, *PT* involves only a reference to objective facts and no perspective taking. If *PT* were predominantly involved in both actor and observer action explanations then there should be no significant actor-observer asymmetry. Since there are three such asymmetries, we have ground to reject the view that adults mainly use *PT* in their action explanations.

The point is less clear when it comes to *TiP* though because *TiP* does involve perspective taking (Roessler and Perner 2013: 40). With *TiP* in mind, the following response on behalf of the advocate of the teleological account might be proposed. When people use teleology to explain their *own* action, they don't need to engage in perspective taking but they still need to do so when they use *TiP* to explain *another* agent's actions. For they will then need to determine what the relevant evaluative and instrumental facts are from within that agent's perspective (ibid). Using teleology in action explanation as an *observer* is thus likely to be more difficult than using it in action explanation as an *actor*. The actor-observer asymmetries hence don't pose a problem for the teleological account because *TiP* belongs to teleology as well. Or so it might be argued.

However, if teleology in general were our main, routine, and preferred strategy to explain agents' action, one whose use is only reduced when insufficient background knowledge for perspective taking is available then we would expect that when observers gain more

information about agents (e.g., about their feelings, intentions, and personal history), this should facilitate their perspective taking and *TiP*. And *TiP*, *qua* (by assumption) preferred, routine strategy, should become used more frequently, leading to a decrease in actor-observer asymmetries.

Malle et al. (2007) conducted a follow-up study pertaining to the issue. In the study, they compared two types of observer explanations of an agent's behavior, one where the agent was a stranger, and one where the observer knew the agent closely (e.g., a friend). The prediction was that the actor-observer asymmetries would disappear when observers explain close agents' actions rather than that of strangers because they would have more intimate knowledge in these cases that they could use to put themselves into the position of the agents. Interestingly, the prediction was not confirmed. For instance, intimate knowledge didn't lead observers to offer more reason explanations and it didn't lead them to reduce their CHR explanations. This casts doubt on the view that teleology is our routine and preferred method for explaining other people's action: if that view were right, providing observers with resources that facilitate *TiP* should lead them to engage more often in perspective taking in their action explanation, leading to an increase in reason explanations, which didn't happen.

But if it isn't a lack of intimate knowledge that causes the actor-observer asymmetries, what else could be responsible? Based on previous studies in which subjects increased their reason explanations when they were motivated to make the actor look rational (Malle et al. 2000), Malle et al. (2007) thought that motivation might play a key role in driving the asymmetries. They thus examined the role of "impression management" in action explanations, that is, the role of observers' "attempts to influence [their] audience's impression" of the agent by, for instance, making her look positive with the action explanation (2007: 495). Distant observers were told that when they were answering the Why-questions about an agent's behavior, they should try to create a good impression of the agent on the experimenter who was asking the question. It turned out that in observers with the explicit motivation to create a positive impression of the agent, the reason asymmetry, in particular, now began to disappear. When observers "were motivated to portray the actor in a positive light[, they produced] almost as many reason explanations as actors themselves did, but without decreasing their causal history explanations" (ibid: 507). These findings suggest that observers *can* explain agents' actions in the same way as actors do and do so when they are motivated in the right way, but frequently aren't.

Having said that, these particular data don't yet show that actors and observers use different methods for explaining actions. Indeed, for all they suggest, teleology might still be predominantly involved in both cases. For instance, overall, averaged across five studies, Malle et al. found that observers offered 60% reason explanations, and 40% CHR explanations (Malle 2005: 165). So observers, just as actors, formed more reason explanations than CHR explanations overall. Thus, while the actor-observer asymmetries do contradict the view that *PT* is our main method for explaining actions, they don't yet undermine the view that *TiP* fits that description.

But I argued in the preceding section that reason explanations are *not* teleological ones but based on (implicit/explicit) mentalizing. If so, then the fact that reason explanations are the most frequent ones among *both* actors and observers does speak strongly against the view that *TiP* is our main strategy for explaining people's actions; it adds further detail to the initial argument.

4.3 CHR explanations and teleology

Since observers offer a significant number of CHR explanations (about 40%; Malle 2005: 165) alongside reasons explanations when making sense of intentional action, advocates of the teleological account might explore the following response to the discussion so far. They might hold that even if reason explanations aren't teleological ones, CHR explanations still are. And since these explanations play a significant role particularly in observers, teleology remains an important, frequently used method for explaining what people do.

But are CHR explanations teleological explanations of the kind advocates of the teleological account are focusing on?¹⁰ For that to be the case, CHR factors would need to be objective reasons (i.e., instrumental or evaluative facts) or subjective reasons (i.e., perspectival facts). There is ground to believe that they are neither.¹¹

Consider two examples of CHR explanations. 'Why didn't John vote?' – 'He's too lazy'; 'Why would you ignore somebody's arguments?' – 'I'm stubborn'. It seems clear that the CH factors involved (i.e., laziness, stubbornness) aren't *subjective* reasons. For advocates of the teleological account maintain that to "say that you have a subjective reason to $[\varphi]$ is to say that, from your perspective it looks as if you have an (objective) reason to do so" (Perner and Roessler 2010: 204). And in the cases at issue, laziness and stubbornness don't enter the agent's own perspective but are "external" to it (Malle et al. 2007: 494).

Might they be *objective* reasons, that is, instrumental or evaluative facts? In the examples, laziness and stubbornness evidently aren't instrumental facts. They aren't a means to an end. Turning to evaluative facts, these facts pertain to what is important, desirable, necessary, or good for the agent (Perner and Roessler 2010: 204; Roessler and Perner 2013: 37), and determine the agent's goal, attracting the agent towards a particular end state. Now, propositions such as *S is lazy* and *S is stubborn* refer to *negative* values (laziness, stubbornness). Since laziness or stubbornness are not 'good for' *S* (neither *per se*, nor from her perspective), they aren't evaluative facts of the kind figuring in teleological explanations as objective reasons. For according to the teleological account, teleological explanations assume that agents act on the basis of goals which are thought to be "good for" them: the agent's "goal is something 'good' in some minimal sense of being attractive, desirable, needed, worthwhile having, etc." (Roessler and Perner 2013; Perner and Esken 2015: 71). Hence, when observers mention an agent's laziness or stubbornness as the facts that account for why the agent is acting, it is unlikely that they are using teleology rather than, for instance, principles such as 'lazy people don't like to vote', 'stubborn people don't like to grant others' arguments' etc.

Of course, the category of CHR factors includes not only laziness or stubbornness. It covers all sorts of dispositions, personality, upbringing, culture, and context. Since that is so, there might also be positive evaluative facts that fit the bill of objective reasons. But since CHR factors don't provide a basis for holding that teleology is involved in action explanations appealing to *negative* features, and we have no independent ground to suspect that things are different in the positive case, it seems explanatorily more parsimonious to hold that teleology isn't involved when CH

¹⁰ I shall set aside the question as to whether CHR explanations might be reason explanations, and whether CH factors might be reasons according to some other view than that of advocates of the teleological account. My focus here is only on the teleological account.

¹¹ There might also be notions of objective reasons and subjective reasons according to which CH factors *are* objective and/or subjective reasons. I here want to focus only on the notions introduced above and proposed by advocates of the teleological account.

factors are positive evaluative facts either. In any case, since at best only a restricted subset of CHR explanations qualifies as explanations referring to evaluative facts that might count as objective reasons, it is fair to say that in most cases, these explanations won't be teleological ones. Thus, even if we only focus on the frequency of observers' CHR explanations, it is unlikely that teleology is our main and routine method for making sense of other people's intentional behavior.

5. Conclusion

Philosophers and psychologists have recently proposed a teleological account of action explanation according to which we explain people's actions predominantly by invoking only publically accessible, non-mental facts. The tenability of the account is philosophically interesting, because the account challenges the received view of action explanation in philosophy of action, which holds that intentional actions are explained by mentalizing. Focusing on action explanation in adults, I argued in support of the received view. I did so by using social psychological data combined with theoretical considerations related to the intelligibility test. They suggest that most often when we explain intentional behavior, we don't do so in terms of non-mental or perspectival facts alone but by (implicitly or explicitly) ascribing propositional attitudes to agents. Social psychological research also suggests that there are actor-observer asymmetries that contradict the view that *PT*, in particular, is our main method for explaining other people's actions. None of this is to deny that teleological thinking plays a role in action explanation in children (especially in those younger than 4) and adults. Perhaps, in our action explanations, as children and adults, we often draw on considerations pertaining to evaluative and instrumental facts that count in favor of acting and assume that agents act in ways they should, given their goals. This might be an important part of the mentalizing involved in making behavior intelligible.

Acknowledgements

I'm grateful to two reviewers of the journal for helpful comments. Many thanks also to Ramiro Glauer and Prof. Frauke Hildebrandt. This paper was financially supported by a grant given to Prof. Albert Newen for a project of the D-A-CH-group 'Structure and Development of Understanding Reasons and Beliefs' (Deutsche Forschungsgemeinschaft; NE 576/14-1).

References

- Alvarez, M. (2010). *Kinds of Reasons: An Essay in the Philosophy of Action*. Oxford University Press.
- Alvarez, M. (2016). Reasons for Action: Justification, Motivation, Explanation. *The Stanford Encyclopedia of Philosophy*. E.N. Zalta (ed.), URL: <https://plato.stanford.edu/entries/reasons-just-vs-expl/>
- Anscombe, G. E. M. (1957). *Intention*. Oxford, England: Basil Blackwell.
- Bartsch, K., & Wellman, H. (1989). Young children's attribution of action to beliefs and desires. *Child Development*, 60, 946–964.
- Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York, NY: Oxford University Press.
- Burge, T. (2018). Do infants and nonhuman animals attribute mental states? *Psychological Review*, 125(3): 409–434.
- Dancy, J. (2008). On How to Act – Disjunctively. In *Disjunctivism: Perception, Action, Knowledge*, A. Haddock and F. MacPherson (eds.), Oxford: Oxford University Press, 262–279.
- Davidson, D. (1963). Actions, reasons and causes. *Journal of Philosophy*, 60: 685–700.
- Fiske, S., and Taylor, S. (2013). *Social cognition*. London: Sage.
- Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences*, 7, 287–292.
- Goldman, A. (1970). *A theory of human action*. Englewood Cliffs, NJ: Prentice Hall.
- Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Morristown, NJ: General Learning Press.
- Knobe, J., & Malle, B. F. (2002). Self and other in the explanation of behavior: 30 years later. *Psychologica Belgica*, 42, 113–130.
- Korman, J. and Malle, B. (2016). Grasping for Traits or Reasons? How People Grapple with Puzzling Social Behaviors. *Personality and Social Psychology Bulletin* 42 (11): 1451–65.
- Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: State of the art. *British Journal of Developmental Psychology*, 30(1), 1–13.
- Malle, B. F. (1999). How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review*, 3, 21–43.

- Malle, B. F., Knobe, J., O’Laughlin, M. J., Pearce, G. E., & Nelson, S. E. (2000). Conceptual structure and social functions of behavior explanations: Beyond person–situation attributions. *Journal of Personality and Social Psychology*, 79, 309–326.
- Malle, B. F. (2005). Self-other asymmetries in behavior explanations: Myth and reality. In *The self in social judgment*, Edited by: Alicke, M. D., Dunning, D. A. and Krueger, J. I. 109–130. New York: Psychology Press.
- Malle, B. F., Knobe, J., & Nelson, S. E. (2007). Actor– observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology*, 93, 491–514. doi: 10.1037/0022-3514.93.4.491
- Malle, B. F. (2011a). Attribution theories: How people make sense of behavior. In D. Chadee (Ed.), *Theories in social psychology* (pp. 72-95): Wiley-Blackwell.
- Malle, B. F. (2011b). Time to give up the dogmas of attribution: An alternative theory of behavior explanation. In J. M. Olson and M. P. Zanna, *Advances of Experimental Social Psychology* (Vol. 44, pp. 297-352). Burlington: Academic Press.
- McClure, J. (2002). Goal-based explanations of actions and outcomes. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 12, pp. 201–235). Chichester, England: Wiley.
- McDowell, J. (2013). Acting in the Light of a Fact. In *Thinking About Reasons: Themes from the Philosophy of Jonathan Dancy*, D. Bakhurst, B. Hooker, and M.O. Little (eds.), Oxford: Oxford University Press, 13–28.
- McLaughlin, B. (2012). Why Rationalization Is Not a Species of Causal Explanation. In J. D’Oro (ed.), *Reasons and Causes: Causalism and anti-Causalism in the Philosophy of Action*, London: Palgrave MacMillan.
- Mele, A. (2010). Teleological explanations of actions: Anticausalism versus causalism. In Aguilar, J. H. & Buckareff, A. A. (eds.). *Causing human actions: New perspectives on the causal theory of action*. Cambridge: The MIT Press, 183–198.
- Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31, 838–850.
- Moore, G. E. (1993). Moore’s Paradox. In *G. E. Moore: Selected Writings*. Edited by Thomas Baldwin, 207-212. London: Routledge.
- Perner, J., & Roessler, J. (2010). Teleology and causal reasoning in children’s theory of mind. In J. Aguilar, A. Buckareff (Eds.), *Causing human action: New perspectives on the causal theory of action* (pp. 199–228). Cambridge, MA: MIT Press.
- Perner, J. & Roessler, J. (2012). From infants’ to children’s appreciation of Belief. *Trends in Cognitive Sciences* 16 (10), 519–525.
- Perner, J. & Esken, F. (2015). Evolution of human cooperation in Homo Heidelbergensis: Teleology versus mentalism. *Developmental Review* 38, 69-88.

- Perner, J., Priewasser, B., & Johannes Roessler, J. (2018). The practical other: teleology and its development. *Interdisciplinary Science Reviews*, 43:2, 99–114.
- Phillips, A. T., & Wellman, H. M. (2005). Infants' understanding of object-directed action. *Cognition*, 98, 137–155.
- Roessler, J. & Perner, J. (2013). Teleology: Belief as perspective. In S. Baron-Cohen, H. Tager-Flusberg, & M. Lombardo (Eds.), *Understanding other minds* (3rd edition; Chapter 3, 35–50), Oxford University Press.
- Scanlon, T. (1998). *What We Owe to Each Other*. Cambridge: Harvard University Press.
- Schneider, D., Slaughter, V., & Dux, P. (2015). What do we know about implicit false-belief tracking? *Psychonomic Bulletin & Review*, 22(1), 1–12.
- Shields, C. (2015). Aristoteles. *The Stanford Encyclopedia of Philosophy*. E.N. Zalta (ed.), URL: <https://plato.stanford.edu/entries/aristotle/>
- Unger, P. (1978). *Ignorance: A Case for Scepticism*. Clarendon Press: Oxford University
- Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind-development: The truth about false belief. *Child Development*, 72, 665–684.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128.
- Wilson, G. (1989). *The Intentionality of Human Action*. Stanford, CA: Stanford University Press.