



Traits, beliefs and dispositions in a pluralistic folk psychology

Harmen Ghijsen¹ 

Received: 10 May 2019 / Accepted: 24 September 2019 / Published online: 22 October 2019
© The Author(s) 2019

Abstract

According to pluralistic folk psychology (PFP) we make use of a variety of methods to predict and explain each other, only one of which makes use of attributing propositional attitudes. I discuss three related problems for this view: first, the prediction problem, according to which (some of) PFP's methods of prediction only work if they also assume a tacit attribution of propositional attitudes; second, the interaction problem, according to which PFP cannot explain how its different methods of prediction and explanation can interact; and third, the difference problem, according to which PFP cannot explain how all of its methods are truly different if it also assumes a dispositionalist account of belief. I argue that a promising solution to these problems should not overestimate the importance and ubiquity of propositional attitude attribution even if the difference between propositional attitude attribution and other types of attribution is a matter of degree rather than kind. Instead, a solution should be sought in a better appreciation of the breadth of folk psychological theorizing and the way in which this can be incorporated into model theory.

Keywords Theory of mind · Pluralistic folk psychology · Social cognition · Hierarchical predictive coding · Character traits · Model theory · Dispositionalism

1 Introduction

In the last couple of decades, the investigation of social cognition has largely focused on mindreading in terms of propositional attitudes. The relevant research question has

This work is part of the research programme “Should we trust our own eyes? Epistemic challenges from cognitive penetration and implicit bias” with Project No. 275-20-056, which is financed by the Netherlands Organisation for Scientific Research (NWO).

✉ Harmen Ghijsen
h.ghijsen@ftr.ru.nl

¹ Faculty of Philosophy, Theology and Religious Studies, Radboud University, Erasmusplein 1, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands

not so much been “How do we understand other people?” as it has been “How do we ascribe beliefs and desires to other people to understand what they are doing?”, thereby simply taking the indispensability of propositional attitudes for granted. This orthodox view of folk psychology has been challenged by a pluralistic view of folk psychology (call it “Pluralistic Folk Psychology” or PFP for short) according to which we make use of a variety of methods to predict and explain each other, only one of which makes use of ascribing propositional attitudes (Andrews 2008, 2012, 2017; Fiebich and Coltheart 2015). Crucial to PFP is the thought that we can perceive each other as full-blown *persons* with habits, character traits, emotions, goals, intentions and, yes, beliefs and desires as well. This echoes and develops a line of critique that is also present in theories of direct social perception, according to which the whole problem of inferring “hidden” mental states is based on a very impoverished and overly intellectualized view of social cognition (e.g., Ratcliffe 2007; Gallagher 2008; Zahavi 2011). By taking into account the multitude of ways in which we can understand each other, PFP hopes to address this worry and enrich our view of folk psychology.

However, the orthodox camp has pushed back against PFP’s broadening of folk psychology. Specifically, Westra (2018) has argued that PFP does not have the resources to explain, first, how all methods of prediction work without also ascribing propositional attitudes (call this the “Prediction Problem”), and, second, how its different types of prediction and explanation can interact without assuming some kind of broader unified theory (call this the “Interaction Problem”). In addition, we can outline an interestingly connected third problem for PFP having to do with its conceptualization of propositional attitudes: if we assume that propositional attitudes are types of dispositions (Schwitzgebel 2002; Andrews 2012, 2017), then this seems to make propositional attitude attribution ubiquitous after all (thereby exacerbating problem 1) and also seems to diminish the posited difference between some types of folk psychological theorizing, such as trait- and belief-based methods of prediction and explanation (thereby exacerbating problem 2). Call this last problem the “Difference Problem”. I will argue that all of these three problems are not insurmountable for PFP and that, in line with PFP, it would be a mistake to overestimate the importance and ubiquity of propositional attitude attribution even if the difference between propositional attitude attribution and other types of attributions is a matter of degree rather than kind.

The plan to show this is as follows. I will start by discussing PFP in more detail in Sect. 2. I will then focus on the Prediction Problem in Sect. 3, and argue that a tacit attribution of propositional attitudes would only make prediction more rather than less difficult. Instead, an appeal to social context can often explain how different types of prediction can get purchase on the folk psychological task at hand. The Interaction Problem is the focus of Sect. 4. Here I will discuss Westra’s (2017, 2018) hierarchical predictive coding framework as a more unified theory of folk psychological theorizing, and outline three worries related to the proposed priority of trait attribution and the proposed ubiquity of different types of mental-state and trait attribution. As an alternative solution to the Interaction Problem, I suggest looking towards model theory (Maibom 2003, 2007; Godfrey-Smith 2005; Spaulding 2018). In the final section (Sect. 5), I explore the Difference Problem. I argue that a dispositional notion of propositional attitudes will indeed make propositional attitude attribution ubiquitous

in a certain sense, but it will still not show that other types of attribution somehow *rely* on it. I also argue that PFP will have much to gain by stressing the different types of dispositions associated with different types of folk psychological categories, such as traits and beliefs, even if, in the end, the distinction between these categories is one of degree rather than kind.

2 Pluralistic folk psychology

The core idea of PFP is that people make use of a *variety* of methods to predict, explain and understand each other in their everyday interaction.¹ Among this list of methods, attribution of propositional attitudes such as beliefs and desires is just one candidate. In many—if not most—cases of everyday interaction people will have different ways of predicting or explaining each other that do not just collapse into attributing propositional attitudes. To give but a few abstract examples, we can use induction over past behavior to anticipate future behavior without having to go to the trouble of positing a precise belief/desire-pair that explains that behavior; we can project what we ourselves would do in a similar situation to predict what someone else will do without attributing a specific belief/desire-pair that motivates that behavior; and we can use information about what *should* be done to predict what a specific individual *will* do, again without attributing a specific belief/desire-pair that rationalizes the action. These methods of induction, projection from self, and norm-based prediction all appear to be, at least on the face of it, sufficiently different from each other (and from attributing propositional attitudes) to license the idea that we might bring to bear a genuine variety of methods in our attempt to make sense of people's behavior. In the mentioned cases we seem to be using different sources of information (about the habits of a specific person, about oneself, or about social norms) that might even be implemented by different mechanisms to predict or explain a certain course of action. And next to these three different folk psychological methods of prediction and explanation, there are far more that could be distinguished, such as those using goal-directed behavior, stereotypes, emotions, and traits.²

Two things should be noted immediately in relation to the different folk psychological methods posited by PFP. First, PFP does not hold that we are conscious of all of the different methods of prediction and explanation that we have at our disposal, nor does PFP hold that we consciously choose between these different methods when trying to make sense of behavior. In fact, often we might not even be conscious of our specific predictions and explanations: we might only become conscious of having made a specific prediction or explanation by feeling surprise when things turn out differently. In many cases one could thus also talk about anticipation and interpretation

¹ Different varieties of PFP have been proposed by Andrews (2008, 2012, 2017) and Fiebich and Coltheart (2015). Where Andrews mostly emphasizes that social cognition consists of different types of procedures by means of which we can predict and explain each other, Fiebich and Coltheart focus on spelling out some of the details of these procedures and provide a heuristic to determine which procedure is most likely to be used under which circumstances. I will here focus mostly on Andrews' higher-level description of PFP rather than Fiebich and Coltheart's hypotheses about the specifics of the different types of procedures.

² See Andrews (2017) for an even fuller list of non-propositional predictive methods.

instead of prediction and explanation (cf. Spaulding 2018, p. 15). Second, PFP holds that there will be asymmetries between the methods used for prediction and explanation: some methods will be more appropriate for prediction, others will be more appropriate for explanation. For instance, I can predict how other people might feel or what they might do on the basis of what I would feel or do in a similar situation, but if I am already puzzled why a person acted in the way they did, reference to myself probably won't get me far. And although it might be useful to ascribe a specific set of beliefs and desires to a person to explain an apparent form of abnormal behavior (Why did he park all the way over there? He probably thought the parking lot here would be full), we will rarely go to all that trouble when we simply predict what the baker will do next upon asking for a particular loaf of bread.

Once one opens oneself up to the possibility that we might be using different folk psychological methods for the purposes of prediction and explanation, these different methods might show up more clearly in concrete cases of everyday mindreading. Think of going to the cinema, where one anticipates that one's fellow cinema-goers will not talk during the movie (prediction on the basis of social norms); or think of the office, where one anticipates one's colleague's immediately getting coffee on her arrival to work (prediction on the basis of past behavior). Or think of explaining a colleague's tendency to work long hours at the office by pointing to her perfectionism (explanation on the basis of character traits); or of explaining a friend's rapid talking by citing his being agitated (explanation on the basis of emotions). Although it's unclear to what extent all of these different folk psychological methods are fundamentally different, it definitely appears implausible that they can be reduced to just one (i.e., propositional attitude attribution).

All of this stands in stark contrast to what is claimed by Standard Theories of Folk Psychology (STFP), such as Theory Theory, Simulation Theory, or Theory-Simulation hybrids, which focus almost exclusively on the attribution of propositional attitudes such as beliefs and desires in order to conceptualize social cognition.³ Not only do such theories neglect the important role of different methods of prediction and explanation that are not based on attributing propositional attitudes, they also treat prediction and explanation as symmetrical even though there appears to be no strong empirical reason to assume that these abilities use the same cognitive mechanisms or even fulfill the same cognitive functions (Andrews 2012). The picture of social cognition sketched by STFP will thus be very skewed and impoverished when compared to the full range of social cognitive abilities that we have at our disposal according to PFP.

Importantly though, many defenders of STFP have felt that there is an obvious response to be made that is nicely expressed by Van Leeuwen (2013):

When I ascribe traits like selfishness or generosity, am I not also ascribing *desires*, like desires for the self or others to have in abundance? Correspondingly, if I ascribe rudeness on the basis of a stereotype, am I not implicitly ascribing mental states, like a bad attitude? Is not the stereotyped boy thought to *believe* jumping rope is for girls? Many cognitive scientists hold that such propositional attitude attributions are made by a fast, unconscious theory of mind system. The

³ For an overview of Theory Theory and Simulation Theory, see Ravenscroft (2016) and Barlassina and Gordon (2017).

same point applies to predictions from the self and from the situation [...] So a traditionalist about folk psychology, especially from the theory theory camp, would say that Andrews has only identified sub-categories of folk prediction that ultimately rely on implicit attribution of propositional attitudes.

If all of our different strategies for folk-psychological prediction and explanation ultimately rely on an implicit attribution of propositional attitudes, then PFP will not be importantly different from STFP. Note, though, that Van Leeuwen's case is not as strong as it may seem. What STFP need to show is not just that we often *also* ascribe propositional attitudes when we use different types of prediction and explanation (trait-based, situation-based, etc.), STFP need to show that those different types of prediction and explanation fundamentally *rely* on those propositional attitude ascriptions. It is exactly this claim that often seems doubtful, even in some of the cases mentioned by Van Leeuwen. For instance, if I have to explain why Alice didn't greet her colleagues when she came in this morning, I can easily do this by simply citing the fact that she is rude (based, say, on the stereotype that all high-profile professors are rude). Nevertheless, I might hesitate to ascribe to her any specific beliefs that might explain the same behavior, such as the belief that other people are inferior, or the belief that her time is too valuable to be spent on greeting everyone in the office. I might not know her that well, after all. But the fact that I don't know Alice's specific beliefs or desires need in no way prevent me from predicting and explaining what she will do based on her rudeness.

The same point applies to other types of prediction and explanation. Think back to the example of the colleague who starts her day at the office by getting a cup of coffee. Although it might seem likely that I will attribute to her some type of desire for coffee, it need not be this desire that is doing the work in either prediction or explanation. For one, it's unclear what the exact content of the desire would be. It need not be a general liking for coffee, as I might not expect her to get coffee during the rest of the day. It need also not be a desire for coffee in the morning, as I might not expect her to get coffee when outside of the office. In fact, I might even be surer of the fact that she will want to *get* the coffee than the fact that she will want to *drink* the coffee, given that I might not know what she does with the coffee once she has it—although I will, of course, expect that she will drink the cup of coffee rather than throw it down the drain. All of this is meant to show that the attribution of desire, insofar as there is such an attribution, will not add much to the habit-based prediction. It is thus not at all clear that the habit-based prediction fundamentally relies upon the ascription of the desire for coffee. And the case of explanation might even be clearer: if someone wanted to know why my colleague got a cup of coffee immediately after arriving at the office, it would not be helpful to cite the fact that she desired a cup of coffee. The mentioning of the habit seems to add information that is not present in the simple ascription of desire.

These considerations show that it appears unlikely that all of the types of folk-psychological prediction and explanation mentioned by PFP will fundamentally rely on propositional attitude attribution. The onus is on the defender of STFP to argue otherwise.

3 The prediction problem

Although it seems unlikely that all types of folk-psychological prediction and explanation can be reduced to propositional attitude attribution, defenders of STFP can raise closely related explanatory challenges for PFP: (1) how exactly do all the different folk-psychological methods lead to specific predictions without ascribing propositional attitudes, and (2) how do all of the different folk-psychological methods interact if they are not somehow part of the same theory? These challenges are a bit more sophisticated than the above worry: instead of *reducing* all different folk-psychological methods to propositional attitude ascription, they intend to show that all of folk psychology is still importantly related to propositional attitude ascription, and, in an important sense, more *unified* than PFP proposes. Westra (2018) can be interpreted as exactly raising these challenges for PFP in relation to trait-based prediction and explanation. I will first focus on the first challenge, labeled as the “Prediction Problem”, and then go on to the second, labeled as the “Interaction Problem”. Like Westra, I will focus mostly on trait attribution and propositional attitude attribution. My main point in discussing these problems is to show that they are unlikely to be satisfactorily resolved by giving a lot of weight to propositional attitude attributions, as proponents of STFP have done.

According to Westra (2018, p. 1225), PFP will have difficulty explaining how trait-based prediction works if it does not also explain how situations are individuated for the purposes of trait-based-prediction. After all, we don’t expect, say, generous people to be generous in *all* situations, we just expect them to be generous in situations that—somehow—sufficiently resemble other situations in relevant respects. PFP has to explain how we individuate situations at the right level, because trait-based prediction will not work without it. Individuate situations at too coarse a level, and we will inaccurately overgeneralize (even the very generous will not give large tips when they get awful service); but individuate situations at too fine a level, and we will not get any predictions going at all (people are usually not any less generous because the lighting is slightly different or it’s now Wednesday rather than Tuesday).

Importantly, Westra claims that once we add propositional attitude ascription to the framework, we will be able to individuate situations in a way that helps trait-based prediction:

What the pluralist proposal lacks is a principled basis for parsing situations for the purpose of behavioral prediction. But if we consider an agent’s beliefs and desires, the solution to this problem is obvious: the ‘situation’ will consist in those features of the local context that the agent believes are relevant to her goals. Moreover, this approach would facilitate predictions even in highly unfamiliar situations. This is because mental-state reasoning is a highly flexible, generative framework for predicting and interpreting behavior.

(Westra 2018, p. 1227)

Let’s apply Westra’s ideas to the example of leaving a tip at a restaurant. Given that people usually don’t believe that the precise lighting conditions in a restaurant are relevant to their goal of enjoying a good meal with good service, the precise lighting conditions should not count as a relevant feature of the ‘situation’ in which one might

exhibit the trait of generosity. Instead, the relevant features of the situation have to do with, e.g., the quality of the food, the politeness of the waiter, and the swiftness of the service, because those are the features of the situation that agents themselves believe to be relevant to their goals in going to the restaurant. These features should thus enable us to individuate the situation at the right level for the purposes of trait-based prediction, which, in this case, comes down to predicting whether the trait of generosity will exhibit itself.

However, Westra's proposal raises the question of how we are able to ascertain what the agent's beliefs about the relevant features of the local context are if we only have the relevant trait to work with. After all, *some* generous people might take lighting conditions to be an important factor in determining whether they will leave a large tip. Admittedly, *if* we already know that we are dealing with a generous person who has this atypical belief, *then* we will take this knowledge into account when individuating situations. But without this prior knowledge about the person, ascribing specific beliefs or desires to help individuate situations at the right level for the purposes of trait-based prediction just seems to make trait-based prediction even more intractable.

If propositional attitude ascription is of no help in answering Westra's Prediction Problem, then how do we individuate situations for the purposes of trait-based prediction? I suggest that we should focus, not on what the to be predicted agent *believes* to be relevant features of the situation in relation to her goals, but on what we ourselves expect to be relevant features of the situation given what we know about the agent's social context. To accurately predict whether the generous restaurant-goer will exhibit her generosity in leaving a large tip, we will have to take into account whether the service she got was merely adequate or really good. But if we're already familiar with the relevant social context of "going to the restaurant", then it seems that we thereby have the relevant information at our disposal to individuate the situation at exactly this level. Usually, the lighting conditions will not be relevant in determining whether generosity will be exhibited, but the quality of the food and service will be relevant. And we can do all of that without the ascription of specific propositional attitudes to the agent.

Note that the proponent of STFP could come back at this point and claim that we are still implicitly ascribing beliefs and desires to the agent in question, based on what people in general, or we ourselves, expect when going to a restaurant. But here again, the problem is that the ascription of such beliefs and desires won't add much to one's knowledge of the normative expectations connected with going to a restaurant. It would thus be unclear why we would go to the trouble of ascribing such specific propositional states to the individual agent in question.

All of this of course presupposes that we are already familiar with certain social contexts, such as that of going to the restaurant. This raises the question whether the attainment of this knowledge does not somehow still rely on the attribution of propositional attitudes. Again, I think the answer to this question should be negative. Attaining knowledge of social contexts and their norms seems to be a matter of partaking in them and being trained to partake in them, which is first and foremost a rule-governed practice: we are told that we, e.g., should not eat with our knives; should use our silverware from the outside and work our ways in; or should not add cheese to our spaghetti alle

vongole. The question whether anyone really *wants* to do this or *believes* that it is the proper way to proceed usually just does not come up.

4 The interaction problem

The attribution of propositional attitudes does not help to explain how we are able to individuate situations at the right level for the purposes of trait-based prediction, and only seems to make prediction even more difficult. Instead, I have suggested that our knowledge of social contexts will enable us to find out which aspects of situations are relevant in determining whether a trait will manifest itself. However, this suggestion brings us directly into the scope of the Interaction Problem for PFP: how do all different types of folk-psychological prediction and explanation (such as those somehow making use of social norms and traits) interact? I will again focus largely on the interaction between trait attribution and propositional-attitude attribution, as this is also mainly the focus of Westra's critique.

According to Westra (2018, pp. 1222–1223), there is plenty of evidence to support the idea that trait attribution and mental-state attribution are functionally intertwined. Although I don't want to dispute the point that such interaction can occur, I do want to note that most of the adduced evidence is not about propositional-attitude attribution *per se*—which is also why Westra rightly uses the broader notion of 'mental state attribution'.⁴ Even with this point made, the problem remains that PFP does not have a detailed story to tell about the *prima facie* plausible interaction that can occur between trait attribution and other types of attributions, among which propositional attitude attribution.

To overcome this Interaction Problem, Westra (2017, 2018) proposes that different types of attributions are integrated in a hierarchical framework of action-prediction, in line with recent other hierarchical predictive coding accounts of the mind (Hohwy 2013; Clark 2016) and, in particular, mindreading (Koster-Hale and Saxe 2013; de Bruin and Strijbos 2015). The rough idea of Westra's specific account is that expectations and hypotheses about more transient states of mind, such as particular goals and behaviors in particular circumstances, are constrained and shaped by expectations and hypotheses about more temporally stable states of mind, such as long-term goals and beliefs, which are, in turn, constrained and shaped by even more temporally stable states of mind, such as character traits. The idea is thus that hypotheses about character traits constitute a high level in the action-prediction hierarchy, hypotheses about mental states constitute a lower level, and hypotheses about particular goals and behaviors an even lower level.

Given that trait attribution has a high level in the hierarchy and helps to constrain behaviorally underdetermined propositional attitude attribution, Westra expects that

⁴ To be precise, Westra cites three lines of evidence; the first is about the way in which the ascription of ulterior motives influences the tendency to attribute the causes of behavior to underlying traits (correspondence bias) (Fein 1996; Ames et al. 2004; Krull et al. 2008; Reeder et al. 2004); the second is about the activation of the neural correlates of trait attribution and mental-state attribution (Cloutier et al. 2011; Ma et al. 2011); the third is about the way in which attribution of personality traits influences the attribution of intentionality and blame (Sripada 2012).

it will often be prioritized upon encountering someone new (2018, pp. 1232–1233). Although such rapidly formed impressions about traits may often be inaccurate, they can be updated in full Bayesian fashion when faced with prediction errors at lower levels (e.g., when predicted behavior does not manifest).⁵ To make this rather abstract picture more concrete, Westra provides us with the following illustration:

[S]uppose that you are observing Tom, whom you believe to be dishonest. A woman walks past, and accidentally drops her wallet in front of him. Tom looks toward the wallet, and then looks back at the woman. Because you know him to be dishonest, you assign a high probability to the hypothesis that Tom desires to steal the wallet. Given this desire-attribution, you might then expect that Tom will perform a series of actions: look around to see if anyone is watching, bend over discretely by the wallet as if tying his shoe, pick up the wallet and put it in his pocket. The prior trait attribution—dishonesty—thus serves as an over-hypothesis, raising the prior probability of mental-state hypotheses that are consistent with the trait in question—namely, self-interested desires [...]

(Westra 2018, p. 1231)

We can add to this example the thought that, if Tom picks up the wallet and immediately returns it to its owner, you might adjust your earlier attribution of him as being dishonest. Whether you would actually do so would depend on a range of factors, such as the amount of instances in which you've received information about Tom's dishonesty, the possibility of ascribing an ulterior motive to Tom, but also whether you expect to be interacting with Tom in the future (which might give you a good reason to form an especially *accurate* model of his personality).

The above illustrated hierarchical framework differs in at least two important respects from PFP. First, in contrast to PFP, which emphasizes that we use a variety of *different* folk psychological methods in understanding each other, Westra claims that (at least) several of those methods are importantly related. In the hierarchical framework under consideration, trait attribution shapes and constrains belief, desire and goal attribution, and the methods of prediction and explanation that make use of such attributions should thus not be considered as independent processes. Such a framework might best be described as a form of Theory Theory where character traits are added as a further underlying variable (cf. Westra 2018, p. 1224). Second, where PFP does not stress the priority of one of the posited different folk psychological methods (although it does tend to stress the overestimated importance of propositional attitude attribution), Westra claims that trait attribution is often prioritized at least in relation to propositional attitude and goal attribution because of its high level in the predictive hierarchy.

In relation to these specific differences with PFP, I want to mention three worries (in the order of least problematic to most problematic). First, as Andrews (2012, pp. 90–91) mentions, children do not explicitly use trait-based predictions and explanations until they are at a relatively late age—around 8 (Rholes and Ruble 1984; Rholes

⁵ Note that Westra rightly stresses that whether one's model is updated or not will depend on whether such an update would be action-relevant. Among other things, this will depend on whether one expects to interact with this person in the future. This can offer a nice explanation of why stereotypes continue to persist even if they're not very accurate (see also Westra 2017).

et al. 1990; Kalish 2002)—although they are capable of using traits as descriptions of behavior at an earlier age and have at least a rudimentary conception of traits already by age 5 (Liu et al. 2007; Heyman 2009; Gonzalez et al. 2010). This stands in contrast to their understanding of propositional attitudes such as beliefs and desires. Children explicitly start talking about desires around 1.5–2 years old (Bartsch and Wellman 1995), about beliefs around 3 years old (Shatz et al. 1983; Bartsch and Wellman 1995), and are able to pass a variety of explicit false-belief tasks around age 4 (Wellman et al. 2001; Oktay-Gür and Rakoczy 2017).⁶ What’s more, in a series of studies examining children’s trait-based explanations (versus explanations referring to some aspect of the situation), younger children (especially 4-year-olds) did not give many classical trait explanations (such as “she is brave”) but referred more to trait-like features (“she’s the bigger sister”, “she is older”) and, importantly, mental states such as beliefs and desires (“she wanted to splash”, “she thinks there is a shark in the water”) (Seiver et al. 2013; Meltzoff and Gopnik 2013). Now, if Westra is right in the functional architecture of our action-prediction hierarchy, then hypotheses about traits should be used to constrain and shape hypotheses about propositional attitudes. But given that children are able to understand and attribute beliefs and desires even before they have a full grasp of trait-concepts, it’s unclear to what extent that architecture is really necessary to ascribe propositional attitudes in the way we do.⁷ Nevertheless, it’s certainly true that children’s development of folk psychology does not end at 4, so it’s possible that the full hierarchy as described by Westra takes longer to develop. This would be in line with the idea that trait concepts capture more abstract behavioral patterns and take longer to be distilled from behavioral data. If this is correct, then one would expect there to be developmental differences in the way propositional attitudes are ascribed once trait attribution has become a part of the predictive hierarchy.

Second, there is some evidence that trait-inferences are actually slower and less likely than inferences about intentionality, desires, or beliefs. In a range of studies, Malle and Holbrook (2012) presented participants with verbal descriptions and short video clips of different types of intentional behavior. Participants then received an inference probe (such as “THINKING” or “PERSONALITY”) and had to respond by pressing either a *Yes* key or a *No* key to answer the previously learned question belonging to the inference probe (e.g., “Did the behavior reveal what the main actor was THINKING in this situation” for “THINKING”, and “Did the behavior reveal a certain PERSONALITY characteristic the actor has” for “PERSONALITY”). The reaction times and proportions of *Yes* answers were then used to determine the speed and likelihood of each inference type. Across all the studies, there was a consistent ordering in terms of speed and likelihood, with intentionality and desire inferences

⁶ Note that some of the results about earlier implicit false-belief understanding (see Scott and Baillargeon 2017 for a review) are by now hotly contested (Kulke et al. 2018; Rubio-Fernández 2019). In any case, the results on explicit measures already provides support for a difference in developmental trajectory.

⁷ Note that Westra (2018, p. 1226) mentions the worry about children’s ability to use propositional attitudes before their ability to use traits in a slightly different context. He responds by noting that “there is now positive evidence that 3–4 years-old children respond in an adult-like manner to facial features associated with warmth and competence, despite the fact that they do not refer to such traits in their explanations of behavior (Cogsdill et al. 2014).” However, the ‘responding in an adult-like manner’ here comes down to selecting the same type of attribution (“mean”/“nice”; “strong”/“not strong”; “smart/not smart”) as adults to the same type of face, which is rather weak evidence for having a strong grasp of traits.

fastest and most likely, followed by belief inferences, and finally, trait inferences as slowest and least likely (nicely in line with the above developmental data). Note, though, that when the stimuli were specifically tailored to elicit trait-inferences, they no longer differed significantly from the other inferences in terms of speed or likelihood.

Overall, these results appear to be in conflict with the action-prediction hierarchy sketched by Westra. Recall that according to Westra, trait attribution should be prioritized, given its high level in the action-prediction hierarchy and its role of constraining propositional attitude attribution. Instead of trait inferences being fastest, though, subjects appear to be slowest in making trait-inferences—at least for non-trait-tailored stimuli. Of course one might claim that the non-trait-tailored stimuli made it too difficult to extract trait information, but then the same would be true for many daily encounters between people. And this goes against positing an action-prediction hierarchy in which trait-inferences are prioritized. However, one can certainly question whether the conceptual hierarchy between trait and propositional attitude attribution—where traits are at a higher, more abstract level than beliefs and desires—also entails a similar temporal hierarchy. Compare it to the case of object recognition: the category an object belongs to is certainly more abstract than the color or shape an object has, but that does not entail that object recognition should be faster than color or shape recognition. So by letting go of the idea that more abstract levels in the hierarchy will always be prioritized, one can maintain a version of the sketched action-prediction hierarchy that is more in line with the empirical evidence.

The third worry for Westra's specific action-prediction hierarchy lies at the heart of PFP's complaints against STFP: in a lot of cases of prediction and explanation, we don't appear to rely on propositional attitude attribution at all. Think back to the case where you predict that a colleague will immediately get a cup of coffee upon arriving at the office: we need not, and don't seem to, attribute a belief or desire to make this inference. Instead, knowledge of the colleague's habit is sufficient. The method we use thus does not seem to depend on propositional attitude or even trait attribution, but instead appears to consist out of a simpler method of induction over past behavior. This gives us some positive evidence for PFP's idea that there are truly *different* folk psychological methods for prediction and explanation of everyday behavior, even if Westra is correct in the action-prediction hierarchy with regard to the interconnectedness of attributing traits, propositional attitudes and particular goals and intentions.

But even this latter connection between trait and propositional attitude attribution is not as strong as Westra makes it out to be. Suppose someone cuts in line in front of you at the supermarket. You might explain this behavior by positing that the person is rude, without attributing any further specific beliefs and desires. This makes it seem as if, contrary to what Westra supposes, we are not constantly generating hypotheses about persons at the different levels of traits, propositional attitudes and goals. Instead, it seems possible that these attributions work independently on some occasions. This suggestion is bolstered by the fact that not all types of inferences (trait, belief, desire and intentionality) were equally likely in the studies by Malle and Holbrook (2012), and the fact that belief attributions have been found to occur not automatically, but in response to task demands (Apperly et al. 2006; Back and Apperly 2010). The kind of

action-prediction hierarchy that we would thus need would be one where some of the different levels can provide outputs independently of the other levels.

These three worries make it difficult to accept the precise action-prediction hierarchy as promoted by Westra, where all levels of attribution are essentially interconnected and priority is given to trait attribution, although they do not (and perhaps need not) provide knock-down arguments against a slightly different hierarchical framework. But even without a fully convincing alternative from the perspective of STFP, PFP is not relieved of its explanatory burden. It still needs to provide an account of how the supposedly different types of methods of folk psychological prediction and explanation are sometimes able to interact. One alternative suggestion that could help to answer this challenge comes from Heidi Maibom's (2003, 2007) and Godfrey-Smith's (2005) account of Model Theory.

According to Maibom (2007, p. 558), our social knowledge consists in knowledge of at least three different types of models, or better yet, three different types of families of models. We can understand actions in terms of simple perceptual and goal states (the behavioral model), we can understand actions in terms of the result of interacting mental states (the folk psychological model), and we can understand actions in terms of the role of the actor in a broader social structure (the social model). Although we can deploy multiple models to gain a fuller understanding of someone's behavior, it is often unnecessary to do so in our everyday encounters with other people.

Importantly, the idea of having knowledge of models is not that we have rich knowledge of universal generalizations that connect, say, beliefs and desires to certain behaviors. Rather, the idea is that we are familiar with a variety of different models, familiar with constructing variations of those models by adding or subtracting elements, and that we have a practical *skill* in applying those models to specific persons or situations depending on our precise aims. For instance, we can use a "strong desire" model that consists out of an agent with a single strong desire, apply it to the case of someone wanting to be a doctor, and predict that this person will, in time, become a doctor. We could also use a "conflicting strong desires" model, consisting out of an agent with several conflicting desires, apply it to someone who wants to become a doctor but also wants to become a musician, and predict that this person will choose to fulfill only one of her desires. But we might as well add skills (good at learning, bad at playing an instrument), and character traits (open-minded, spontaneous, creative) to make an even more nuanced prediction. And, given that we can apply models in more than one way, we can use the same models to explain behavior or categorize behavior in certain ways (as one might do when explaining how someone came to become a doctor).

A view along these lines would immediately allow for interaction between different folk psychological attributions, such as trait, mental-state and propositional-attitude attribution, without implying that one of these types of attribution (i.e., propositional attitude attribution) is more fundamental or ubiquitous than the others. Indeed, Shannon Spaulding (2018, p. 71) has explicitly argued that Model Theory's ability to explain the diversity in modes of mindreading, inputs to mindreading, goals that mindreading serves and products that mindreading delivers is one of the major advantages of Model Theory over competing Theory Theories and Simulation Theories of social cognition. In this respect Model Theory would also clearly adhere to the core idea of PFP that

propositional attitude ascription is just one method among many. Spaulding (2018, p. 72) further takes Model Theory to be perfectly compatible with predictive coding accounts of mindreading such as those presented by Westra, but, as we've seen above, such accounts would have to be validated by finding different patterns of propositional attitude attributions over time (worry 1); uphold the idea that the temporal hierarchy of attributions need not reflect the conceptual hierarchy of the different levels (worry 2); and accept that the different levels of the hierarchy can work both independently and in tandem (worry 3).

One worry that is left about Model Theory specifically is that it still appears to reduce folk psychology to only one set of skills: that of model-building and model-application. One might think, in line with a general worry about versions of Theory Theory, that this suggestion is too intellectualized to work for all relevant cases, such as those where one simply reads off an agent's intention directly by looking at her action (Maibom's behavioral model). To the extent that a proponent of PFP wants to maintain that the variety of folk psychological methods is implemented by a variety of different cognitive *mechanisms*, Maibom's, Godfrey-Smith's and Spaulding's notion of folk psychology as skilfully applying different families of models might not be the best fit. Nevertheless, it could still constitute a promising starting point for explaining the interaction between trait and mental-state attribution, and shows that there is at least a version of PFP that is already able to overcome the Interaction Problem.

5 The difference problem

So far I've argued that the Prediction Problem and Interaction Problem are unlikely to be resolved satisfactorily by assuming that we are constantly in the business of attributing propositional attitudes to other people (in line with STFP), and that, in fact, there is a way in which these challenges can be answered from the perspective of PFP. In this final section, I want to sketch another open (but seemingly surmountable) explanatory challenge for PFP that is related to the specific *nature* of traits and propositional attitudes, and which has a bearing on its answers to the Prediction and Interaction Problem.

Andrews (2012, p. 30; 2017, p. 120) is open to the idea that we can understand beliefs, at least in part, as specific sorts of dispositions along the lines of Schwitzgebel (e.g., 2002, p. 253): “to believe that P is [...] nothing more than to match to an appropriate degree and in appropriate respects the dispositional stereotype for believing that P”. Crucial here is that the proposed dispositional stereotype does not just consist out of behavioral dispositions (such as taking your umbrella with you when you believe that it is raining), but also out of phenomenal dispositions (such as feeling surprise when you find out that it isn't raining) and cognitive dispositions (such as inferring that it might be a better idea to take the car rather than the bike).⁸ However, if we do view beliefs as dispositions with behavioral, phenomenal and cognitive components, rather

⁸ Note that I am here focusing on beliefs, although I assume that a similar dispositional story could be given for the other propositional attitudes.

than as representations with propositional contents, then this has several interesting consequences for PFP.

First, as Andrews also notes, with beliefs construed as dispositions, belief attribution can immediately become less cognitively demanding. Instead of the need for meta-representations, we can now claim that attributing a belief comes down to attributing a disposition to behave, experience and cognize in a specific sort of way. In this light-weight sense of ‘belief’, it would be far more plausible to think that we are constantly ascribing beliefs to other people in our everyday encounters. After all, we expect people to behave in certain patterned ways, and those expectations of patterned behaviors seem sufficient for counting as attributing beliefs to other agents in the light-weight dispositional sense.⁹ Does that mean that proponents of STFP have been right all along in their claim that propositional attitude attribution is ubiquitous and also at the heart of different folk psychological methods of prediction; in other words, does the Prediction Problem return in full force?

To answer this question, let’s start by noting that most proponents of STFP do *not* think of belief attribution in this light-weight sense, but rather in the meta-representational sense. It is thus of no help to those proponents that belief attribution is ubiquitous in the dispositional sense of ‘belief’. More importantly, though, is that even if we accept the ubiquity of propositional attitude attribution in the light-weight sense, then this still does not show that other types of attribution somehow *rely* on it. Although attributing a belief now fully comes down to anticipating certain patterns of behavior, there’s no reason to think that our anticipating other patterns of behavior (say, the anticipation that someone will act rudely in the light-weight sense of attributing traits) will always depend on the anticipation of the former (belief-like) behavioral patterns. In some cases, the “trait pattern” might simply be more recognizable than the “belief pattern”.

These last remarks are closely related to the second consequence for PFP’s accepting beliefs as a specific sort of disposition. Given that traits are usually seen as the paradigm dispositional mental entities, PFP would have to provide an explanation of why trait-based prediction and explanation could still count as a different folk psychological method from belief-based prediction and explanation, once we also accept that beliefs are a specific sort of disposition. Doesn’t this lend more support to the idea that both beliefs and traits belong to a broader unified hierarchical theory just as Westra (2018) proposed in answer to the Interaction Problem? Moreover, if both types of attributions are ultimately about ascribing dispositions to other people, wouldn’t PFP face the same challenge as Westra in explaining the earlier mentioned

⁹ If ascribing beliefs is now merely a matter of anticipating certain patterns of behavior, then does that mean that some types of non-human animals are also capable of ascribing beliefs? Although it certainly removes a certain sort of principled worry about the possibility of belief ascription in non-human animals (i.e., the cognitively sophisticated skill of meta-representation), it does not follow that a dispositionalist is committed to granting that some non-human animals are capable of ascribing beliefs. The dispositional stereotype associated with believing that P might itself be too complex for non-human animals to grasp, especially when considering the kind of cognitive dispositions that might be part of that stereotype. That being said, if one looks at the mindreading skills that apes have (see Andrews 2017, for an overview), and takes into account the variety of beliefs that we can ascribe to other people (also see a few paragraphs down), it seems plausible that they are capable of ascribing at least some simple types of beliefs (e.g., the belief that there’s food over there) on a dispositionalist account.

empirical data, that emphasize the different developmental and temporal trajectories for belief and trait attribution, and the different speeds and likelihoods with which belief and trait inferences are made?

My suggestion is that, to answer this question, PFP will have to spell out in more detail what the relevant differences are between the sort of dispositions that traits consist of, and the sort of dispositions that beliefs consist of. Although this will probably mean that the difference between several folk psychological methods will be one that is a matter of degree rather than kind, it could still enable PFP to explain why there is an intuitive difference between the trait-based method and the belief-based method, and, preferably, why there are the mentioned empirical differences in their use. As to what those relevant differences in dispositions precisely are, there are at least two related properties that spring to mind: first, their temporal stability, and, second, their context-specificity.

With regard to the first property, temporal stability, it is clear that traits are often seen as stable and difficult to change properties of individuals. This stands in contrast to beliefs, as we actually expect beliefs to change in response to new evidence. Thus, if we encounter the same individual after a longer period of time, we will expect more similarity in character traits than in beliefs. With regard to the second property, context-specificity, the idea is that we can use traits across multiple contexts to predict or explain behavior in a way that we are unable to do with beliefs. For instance, when we know that a person is helpful, we can predict what will happen in a large number of different circumstances, whether it is at the office, at home, with friends, or any other occasion (whether the predictions will always come out right is a different matter). In contrast, when we know that, say, Sally believes that the marble is in the box rather than in the basket, this doesn't help us determine what she will do (or why she did it) in any other context than the one currently at play. This is so even when we look at beliefs that are more general in nature than the mentioned perceptual beliefs (which are arguably about *this* box and *that* basket), such as the belief that spaghetti alle vongole is delicious. Only in very specific contexts will you be able to predict or explain behavior by attributing that belief (for one thing, spaghetti alle vongole should be present on the menu). However, by the same token it might often be easier to attribute beliefs than traits in a specific context: given that traits should appear stably across contexts, a specific context might not give you sufficient information to attribute a trait (given what I see now, is she really helpful *in general*?) even though that same context does give you sufficient information to attribute a belief (well, at least she believes she should help *this* person). This is one factor that could also help to explain the empirical fact that trait-inferences appear relatively late in development and are in general more slow and less likely than belief-inferences: before we can attribute traits, we have to be able to recognize a pattern in different behavioral manifestations (just think of the different ways in which “being helpful” can manifest behaviorally) that is stable across multiple contexts (see also Rholes et al. 1990, p. 371). Note that this developmental trajectory is compatible with our over-attributing traits on the basis of a mere instance of behavior (i.e., correspondence bias) once we have grasped how they work.

The stronger context-specificity of belief attribution appears related to the specificity of the attributed content. The fact that Anne believes that it's fun to play tricks

on Sally will only tell you something about Anne's behavior in contexts where Sally is present (broadly speaking), but is of no immediate help when Anne isn't interacting with Sally. The upside of this is that, in contexts where a specific belief is relevant, the prediction or explanation can often be more specific and accurate than one made with the help of traits. For instance, with the help of the former belief we can predict that Anne will play a trick on Sally even if Anne isn't really the type of person to play tricks on other people. So the specificity of belief attribution can work both as a strength (when applicable, we can use it to make precise predictions and give precise explanations) and a weakness (only applicable in specific contexts). But note that we can make belief contents as specific as we like, which also makes it possible to attribute far less specific beliefs. For instance, we could also attribute to Anne the general belief that it's fun to play tricks on people. This will immediately make the distinction between belief and trait attribution less clear: there doesn't seem to be much difference between a trickster and a person who believes that it's fun to play tricks, at least when we understand 'belief' in the dispositional sense again. I think this is exactly right: even though, in general, attributed beliefs tend to be more context-specific and less stable than character traits, there is an enormous variability along these dimensions among different kinds of beliefs. It would thus also be an interesting empirical question to what extent the attributions of quite stable, context-independent beliefs (such as the belief that it's fun to play tricks on people, or the belief that one should be kind to other people) resemble the developmental data of trait attributions.

With this in mind, let's return to the discussion between PFP and Westra's hierarchical framework. If beliefs and traits only differ in degree rather than kind, then one might think that a hierarchical framework is exactly suited to explain our abilities of attributing them: stable, context-independent traits are just what one gets once one goes one level of abstraction up from less stable, more context-specific beliefs. In that case one could say that although PFP is correct in characterizing a *variety* of methods from a folk psychological perspective, Westra's more *unifying* account is a better fit with the underlying implementation for those different methods—as long as we don't suppose that this implementation really make use of meta-representational entities. Although I am sympathetic to this proposal, I think it might underestimate the variety of folk psychological methods we have at our disposal. After all, we can group trait- and belief-based prediction together as person-based prediction, but we can also predict behavior based on social norms, situations, or the self to name but a few alternatives. The burden of proof remains with those who claim that all of these diverse folk psychological methods have something important in common.

6 Conclusion

I have argued that the Prediction Problem and Interaction Problem are not insurmountable challenges for PFP, and have suggested that we should focus on social contexts and, perhaps, flexible model-building to find satisfying answers. Attributing propositional attitudes to individuate situations at the right level will only make trait-based prediction more intractable, and adding (hierarchical) propositional attitude attributions to all versions of folk psychological prediction and explanation seems to

get the empirical relations between trait-inference and mental-state inference wrong. However, in response to the Difference Problem I have also argued that, once we conceptualize beliefs as certain types of dispositions, (1) belief attributions will be more ubiquitous and less problematic, and (2) the difference between trait-based prediction and explanation and belief-based prediction and explanation will become a matter of degree rather than kind. It will then be a further challenge for PFP to account for the empirical differences between trait-inferences and mental-state inferences, and I have suggested that the properties of temporal stability and context-specificity could serve as useful dimensions along which those differences can be described. Even if this means that some folk psychological methods are closer connected than previously thought, there is no reason to suppose that one can be fully reduced to another, and it remains an open question to what extent a unifying story can be told for the whole variety of folk psychology.

Acknowledgements Thanks to my colleagues at the CCCL (Center for Cognition, Culture and Language) at Radboud University and two anonymous reviewers for all of their helpful comments and suggestions.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Ames, D. R., Flynn, F. J., & Weber, E. U. (2004). It's the thought that counts: On perceiving how helpers decide to lend a hand. *Personality and Social Psychology Bulletin*, *30*(4), 461–474.
- Andrews, K. (2008). It's in your nature: A pluralistic folk psychology. *Synthese*, *165*(1), 13–29.
- Andrews, K. (2012). *Do apes read minds? Toward a new folk psychology*. Cambridge: MIT Press.
- Andrews, K. (2017). Pluralistic folk psychology in humans and other apes. In J. Kiverstein (Ed.), *The Routledge handbook of philosophy of the social mind* (pp. 117–138). London: Routledge.
- Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, *17*(10), 841–844.
- Back, E., & Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition*, *115*(1), 54–70.
- Barlassina, L., & Gordon, R. M. (2017). Folk psychology as mental simulation. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2017 ed.). Stanford: Metaphysics Research Lab, Stanford University.
- Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. Oxford: Oxford University Press.
- Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford: Oxford University Press.
- Cloutier, J., Gabrieli, J. D., O'young, D., & Ambady, N. (2011). An fmri study of violations of social expectations: When people are not who we expect them to be. *NeuroImage*, *57*(2), 583–588.
- Cogsdill, E. J., Todorov, A. T., Spelke, E. S., & Banaji, M. R. (2014). Inferring character from faces: A developmental study. *Psychological Science*, *25*(5), 1132–1139.
- de Bruin, L., & Strijbos, D. (2015). Direct social perception, mindreading and bayesian predictive coding. *Consciousness and Cognition*, *36*, 565–570.
- Fein, S. (1996). Effects of suspicion on attributional thinking and the correspondence bias. *Journal of Personality and Social Psychology*, *70*(6), 1164.
- Fiebach, A., & Coltheart, M. (2015). Various ways to understand other minds: Towards a pluralistic approach to the explanation of social understanding. *Mind and Language*, *30*(3), 235–258.
- Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, *17*(2), 535–543.

- Godfrey-Smith, P. (2005). Folk psychology as a model. *Philosopher's Imprint*, 5(6), 1–16.
- Gonzalez, C. M., Zosuls, K. M., & Ruble, D. N. (2010). Traits as dimensions or categories? Developmental change in the understanding of trait terms. *Developmental Psychology*, 46(5), 1078.
- Heyman, G. D. (2009). Children's reasoning about traits. *Advances in Child Development and Behavior*, 37, 105–143.
- Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press.
- Kalish, C. W. (2002). Children's predictions of consistency in people's actions. *Cognition*, 84(3), 237–265.
- Koster-Hale, J., & Saxe, R. (2013). Theory of mind: A neural prediction problem. *Neuron*, 79(5), 836–848.
- Krull, D. S., Seger, C. R., & Silvera, D. H. (2008). Smile when you say that: Effects of willingness on dispositional inferences. *Journal of Experimental Social Psychology*, 44(3), 735–742.
- Kulke, L., Reiß, M., Krist, H., & Rakoczy, H. (2018). How robust are anticipatory looking measures of theory of mind? Replication attempts across the life span. *Cognitive Development*, 46, 97–111 (**Understanding theory of mind in infancy and toddlerhood**).
- Liu, D., Gelman, S. A., & Wellman, H. M. (2007). Components of young children's trait understanding: Behavior-to-trait inferences and trait-to-behavior predictions. *Child Development*, 78(5), 1543–1558.
- Ma, N., Vandekerckhove, M., Baetens, K., Van Overwalle, F., Seurinck, R., & Fias, W. (2011). Inconsistencies in spontaneous and intentional trait inferences. *Social Cognitive and Affective Neuroscience*, 7(8), 937–950.
- Maibom, H. L. (2003). The mindreader and the scientist. *Mind and Language*, 18(3), 296–315.
- Maibom, H. L. (2007). Social systems. *Philosophical Psychology*, 20(5), 557–578.
- Malle, B. F., & Holbrook, J. (2012). Is there a hierarchy of social inferences? The likelihood and speed of inferring intentionality, mind, and personality. *Journal of Personality and Social Psychology*, 102(4), 661.
- Meltzoff, A. N., & Gopnik, A. (2013). Learning about the mind from evidence: Children's development of intuitive theories of perception and personality. In S. Baron-Cohen, H. Tager-Flusberg, & M. Lombardo (Eds.), *Understanding other minds: Perspectives from developmental social neuroscience* (pp. 19–34). Oxford: Oxford University Press.
- Oktay-Gür, N., & Rakoczy, H. (2017). Children's difficulty with true belief tasks: Competence deficit or performance problem? *Cognition*, 166, 28–41.
- Ratcliffe, M. (2007). *Rethinking commonsense psychology: A critique of folk psychology, theory of mind and simulation*. New York: Palgrave Macmillan.
- Ravenscroft, I. (2016). Folk psychology as a theory. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2016 ed.). Stanford: Metaphysics Research Lab, Stanford University.
- Reeder, G. D., Vonk, R., Ronk, M. J., Ham, J., & Lawrence, M. (2004). Dispositional attribution: Multiple inferences about motive-related traits. *Journal of Personality and Social Psychology*, 86(4), 530.
- Rholes, W. S., Newman, L. S., & Ruble, D. N. (1990). Understanding self and other: Developmental and motivational aspects of perceiving persons in terms of invariant dispositions. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 369–407). New York: The Guilford Press.
- Rholes, W. S., & Ruble, D. N. (1984). Children's understanding of dispositional characteristics of others. *Child Development*, 55, 550–560.
- Rubio-Fernández, P. (2019). Publication standards in infancy research: Three ways to make violation-of-expectation studies more reliable. *Infant Behavior and Development*, 54, 177–188.
- Schwitzgebel, E. (2002). A phenomenal, dispositional account of belief. *Noûs*, 36(2), 249–275.
- Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences*, 21(4), 237–249.
- Seiver, E., Gopnik, A., & Goodman, N. D. (2013). Did she jump because she was the big sister or because the trampoline was safe? Causal inference and the development of social attribution. *Child Development*, 84(2), 443–454.
- Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition*, 14(3), 301–321.
- Spaulding, S. (2018). *How we understand others: Philosophy and social cognition*. New York: Routledge.
- Sripada, C. S. (2012). Mental state attributions and the side-effect effect. *Journal of Experimental Social Psychology*, 48(1), 232–238.
- Van Leeuwen, N. (2013). Review of do apes read minds? Toward a new folk psychology. *Notre Dame Philosophical Reviews*. <http://ndpr.nd.edu/news/39001-do-apes-read-minds-towards-a-new-folk-psychology/>. Last visited March 20th 2019.

- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72(3), 655–684.
- Westra, E. (2017). Stereotypes, theory of mind, and the action-prediction hierarchy. *Synthese*, 196, 1–26.
- Westra, E. (2018). Character and theory of mind: An integrative approach. *Philosophical Studies*, 175(5), 1217–1241.
- Zahavi, D. (2011). Empathy and direct social perception: A phenomenological proposal. *Review of Philosophy and Psychology*, 2(3), 541–558.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.