

Why Desire Reasoning is Developmentally Prior to Belief Reasoning

Asbjørn Steglich-Petersen and John Michael

Penultimate draft.

Forthcoming in *Mind & Language*

Word count: 11.765

Abstract:

The predominant view in developmental psychology is that young children are able to reason with the concept of desire prior to being able to reason with the concept of belief. We propose an explanation of this phenomenon that focuses on the cognitive tasks that competence with the belief and desire concepts enable young children to perform. We show that cognitive tasks that are typically considered fundamental to our competence with the belief and desire concepts can be performed with the concept of desire in the absence of competence with the concept of belief, whereas the reverse is considerably less feasible.

Keywords: theory of mind; metacognition; cognitive development; false belief; desire.

We are grateful to Jane Heal, Simon Blackburn, Veli Mitova, and Jesse Prinz for helpful discussion, and to several anonymous referees for valuable feedback on previous drafts. For financial support, we thank the Gates Cambridge Trust and the Danish Council for Independent Research.

Address for Correspondence: Asbjørn Steglich-Petersen, Department of Philosophy and Intellectual History, Aarhus University, Jens Chr. Skous Vej 7, 8000 Aarhus C, Denmark.

Email: filasp@cas.au.dk

1. Introduction

Among the most important human cognitive capacities acquired early in life is the capacity to predict and explain the behavior of others by ascribing various mental attitudes to them. Such mental attitudes are usually divided into two main groups: (i) conative attitudes, sometimes also referred to as ‘pro-attitudes’, most importantly *desires*; and (ii) cognitive attitudes, most importantly *beliefs*. In our everyday folk-psychology, we understand behavior to a large part as motivated by combinations of beliefs and desires. For example, we might explain why someone opened a cookie jar in terms of a desire for cookies combined with a belief that there were cookies to be found in the jar.

In developmental research, the predominant view is that young children become fluent in reasoning with the concept of desire well before becoming fluent in reasoning with concept of belief. The empirical support for this claim comes from a substantial body of experiments on the ability of infants to employ these concepts in their understanding of others. In a classic study, Wellman and Woolley (1990) found that children as young as two years-old *predict* others’ behavior in terms of desires, and also ascribe positive or negative emotions according to whether an agent’s desire is fulfilled or unfulfilled. Specifically, the children predicted that an agent in a story (presented with cardboard cut-out characters) would continue searching for an object she desired (such as a dog or a rabbit) until finding the object, and would cease searching only if she found the object. Moreover, they correctly identified that each agent would be happy if she found precisely the object she was looking for (and not some other desirable object that a different agent was looking for), and otherwise sad. Indeed, they were able to do so even when they themselves had a different desire – e.g. if they themselves had expressed a preference for finding a dog, whereas the

character in the story was looking for a rabbit. This makes it unlikely that the children predicted the character's emotional reaction to finding the object by projecting their own emotion rather than by ascribing the emotion that was appropriate to the agent's desire. These findings build upon the results of several earlier studies (Stein and Levine 1989; Wellman and Bartsch, 1988; Yuill 1984), and are also consistent with several subsequent studies (Hadwin & Perner, 1991; Wellman & Banerjee, 1991; Rakoczy et al, 2007.). In one later study, Bartsch and Wellman (1995) also found that children at two years of age are able to *explain* actions by reference to desires. And perhaps most strikingly, Repacholi and Gopnik (1997) found that even 18 month-olds are able to understand that others may sometimes have preferences which they themselves would never have (i.e. Mommy prefers broccoli to crackers).

Importantly, all of these capacities are acquired well before children become capable of passing analogous tasks involving the concept of belief, such as the classic verbal false-belief test (Wimmer & Perner, 1983), and tasks involving belief-dependent emotions, such as surprise (Hadwin & Perner, 1991; Wellman & Banerjee, 1991). This developmental priority of desire reasoning is in fact well-illustrated by the aforementioned Wellman and Woolley (1990) study. In 'Experiment 2' of that study, there was a contrast between cases in which the character in a story expressed a different desire from the child but had no different beliefs ('Not-Own Desire') and cases in which the character expressed the same desire but had a different belief about the location of the object ('Not-Own Belief'). Although 85% of the (two year-old) children succeeded on all three trials of the Non-Own Desire task, only 45% of them succeeded on all three versions of the Not-Own Belief task. Because of findings such as these, the predominant view shared by most researchers in developmental psychology is that fluency in reasoning with the concept of desire emerges earlier

than fluency in belief reasoning. As Wellman and Woolley (1989) put it, 2-year-olds operate with a ‘simple desire psychology’, which ‘rests on an important distinction between internal intentional states and external reality’ (251). They argue that this simple desire psychology supports ‘differential prediction of continued action (because you did not yet get what you want) versus cessation of action (because you did) and also underpins prediction of emotional reactions of happiness versus unhappiness’ (251), but is limited insofar as it does not include a concept of belief and therefore does not enable children to predict that two agents with the same desire may perform different actions because of differing beliefs, or that an agent may perform an action that is unlikely to satisfy her desire because she has a false belief.

However, the view that the fluency in desire reasoning develops prior to the fluency in belief reasoning (henceforth ‘the received view’) has not gone unchallenged in the literature. In the following section, we will review two points of dissent that have been formulated, and argue that neither suffices to undermine the received view. We will then (sec. 3) present and assess three previous explanations that have been offered for the developmental priority of desire reasoning. We will be arguing that all three proposed explanations, though illuminating to a certain extent, are incomplete. We will then (sec. 4) present a novel explanation, which is complementary to the existing proposals discussed in section 3.

2. Challenges to the received view

2.1 Objective desirability

One important challenge to the received view is the proposal that although young children are proficient in ascribing *desire-like* attitudes before being able to ascribe beliefs, they do not yet have the more sophisticated notion of truly subjective and

perspectival desires, needed to understand, for instance, the possibility of different agents having conflicting desires; rather, young children at that stage operate with something like a notion of ‘objective desirability’ (see esp. Moore et al., 1995, Perner et al., 2005). Employing such a notion, young children may regard different events as objectively good or bad for different people in different situations, e.g. they may view it as objectively good that Mommy eat broccoli and objectively good that they themselves eat crackers. Correspondingly, they may expect people to be happy about objectively good events and sad about objectively bad events. What such a notion would *not* enable children to understand is that people may have conflicting desires about one and the same event, and accordingly respond with conflicting emotions to one and the same event. So, for example, although a notion of objective desirability would make it intelligible that Mommy chooses broccoli over crackers and is happy about eating broccoli, whereas someone else (e.g. the child herself) may choose crackers over broccoli and be happy about eating crackers, it would not make it possible to understand that Mommy may be unhappy about the cat eating crackers while someone else (e.g. the child herself) may be happy about that same event.

Support for this proposal comes from Moore et al. (2005). They devised a puzzle game in which children competed against a puppet, Fat Cat. In each round, either a red card or a blue card was drawn, and both players were then allowed to take a red or a blue puzzle piece, according to the color of the card, and to add it to their own puzzle. In the relevant test condition, the child needed a card of one color to complete her puzzle while Fat Cat needed a card of the other color. This scenario created a conflict because the next card could be *either* blue *or* red, and would thus satisfy either the child’s desire or Fat Cat’s desire, but not both. The child was then asked what color Fat Cat wanted the next card to be. The result was that three year-

olds performed poorly on the task, frequently ascribing to Fat Cat the desire that they themselves had. This finding suggests that children younger than 4 may have difficulty understanding genuinely subjective desire states, and thus provides support for the proposal that children younger than 4 years-old may rely on an objectified notion of desire rather than on a subjective concept of desire.

Further support comes from a study by Lichtermann (1991), in which children were presented with stories about two characters traveling either in the same boat or in two different boats along a river. Upon coming to a junction, one of the characters wanted to go left whereas the other wanted to go right. In the condition in which they were in separate boats, their two desires were not incompatible although they were different (it was possible for one of the boats to go one way and the other to go the other way, satisfying both parties). However, in the condition in which the characters were in the same boat, their different desires were in fact incompatible, making it impossible to satisfy both parties. After the boat had gone one way, children were asked which character was happy, and then whether the other character was also happy. The result was that 3 and 4 year-old children had more difficulty with the second question in the incompatible condition (where the correct answer was 'No') than in the compatible condition (where the correct answer was 'Yes'). And in fact their performance in the compatible condition was highly correlated with their performance on a false belief task, thus supporting the view that the concept of desire as a truly subjective state does not arise prior to the concept of belief.

However, as Rakoczy et al. (2007) have pointed out, there are several problems with the methodologies of these studies. First of all, Moore et al.'s Fat Cat paradigm was rather complicated, and may have confused the children: in order to answer correctly, it was necessary to infer from the state of Fat Cat's puzzle what

color piece he would likely desire, and then on that basis infer what card Fat Cat would desire to be drawn. Moreover, since the children were themselves always involved in the game, they also had to inhibit their own desire for one card in order to identify Fat Cat's desire, thereby requiring substantial executive control. As for the Lichtermann study, it is important to note that the children were always first asked which character was happy, and only after this asked about the other character's emotional state. As Rakoczy et al. have suggested, it is possible that the outcome might have been different if they had first been asked who was sad. In fact, given that the characters were getting along well and traveling together in the same boat, the children might have expected that both would be happy, even though one of them had not had her original desire fulfilled.

In order to address these issues, Rakoczy et al. 2007 conducted a pair of studies, which incorporated modified versions of the methodologies used in the two earlier studies. In 'Study 1', they modified the Lichtermann paradigm such that the two characters implicitly expressed their desires (i.e. that the boat or boats should go left/right) rather than the experimenter explicitly informing the child of this. Moreover, in the test condition (i.e. when the two characters' desires were truly incompatible because they were in the same boat, which of course could only go one way), the two characters quarreled about which direction to take. Since young children frequently engage in quarrels and sometimes also witness others quarreling, this is likely to be a familiar way of experiencing a conflict between incompatible desires. This manipulation was intended to make it as salient to the children as possible that the two characters had conflicting, i.e. incompatible, desires. The result was that 3 year-old children were much better at ascribing incompatible desires and desire-dependent emotions than they were in false belief tasks. In Study 2, children

performed a simplified version of Moore et al.'s puzzle task. Instead of placing differently colored pieces in a puzzle according to the colors of selected cards, they placed stickers in a book according to where a marble landed that had dropped out of an apparatus. The most relevant difference was that instead of having to infer the other character's desire from the setup of the game (what color puzzle piece Fat Cat needed in order to complete his puzzle), the other character actually expressed their desire (what sticker they wanted to place in a book). Thus, children only had to infer what instrumental desire would lead to the fulfillment of the desire that the character had expressed (i.e. where the marble would have to land in order for the character to get the sticker that he wanted). And indeed children were much better at ascribing incompatible desires on his task than they were at a false belief task. Interestingly, though, they had difficulties in ascribing desire-dependent emotions even in this simplified paradigm. But, as Rakoczy et al. point out, this task requires an additional inference (i.e. inferring what emotion someone would have about an event which they had either desired to occur or not). So their difficulties are in fact consistent with the suggestion that task complexity was a major factor in the Moore et al. and Lichtermann paradigms.

In sum, Rakoczy and colleagues' findings largely undermine the evidence which had supported the objectified desirability account, and corroborate the received view that young children at 3 years of age *do* operate with a more sophisticated, subjective, concept of desire.

2.2 Early belief reasoning

While the objective desirability account aims to explain away the developmental priority of desire reasoning by arguing for a later development of the concept of

desire as a truly subjective state, a different challenge comes from researchers who take the opposite tack and argue that belief reasoning emerges much earlier than is implied by the received view.

This challenge is motivated by experimental evidence that infants begin exhibiting implicit understanding of false beliefs soon after their first birthdays (e.g., Onishi and Baillargeon 2005, Southgate et al. 2007, Baillargeon et al. 2010) – and perhaps already at seven months (Kovacs et al. 2010). The first evidence pointing in this direction came from a study conducted by Clements and Perner (1994), which revealed an intriguing inconsistency in young children’s responses in a false belief situation: 90% of the children between 35 months and four-and-a-half years old in this study looked first to the empty location (where the observed agent falsely believed the object to be located), and yet only 45% of them gave the correct verbal answer when asked where the agent with the false belief was likely to search. Building upon these findings, Onishi and Baillargeon (2005) reported the results of a study, which, they argued, demonstrated that 15 month-olds were capable of implicit understanding of an observed agent’s false belief. Specifically, they found that children looked longer at a scene in which an agent who should have had a false belief (because she had left the scene temporarily, and during her absence the object had been transferred to a different location) searched in the correct location for an object than at a scene in which the same agent searched for the object in the wrong location. According to the authors, the increased looking time indicated a violation of expectation (VOE), which revealed that the child had implicitly formed the expectation that the agent would search in the wrong location because she had a false belief. These findings have since been replicated by several different groups using similar paradigms: Surian and colleagues (2007), for example, observed the same pattern in a study involving 13

month-olds, and Kovacs et al. have even found similar evidence in a study with seven-month-olds.

These findings have also been extended in various respects. First of all, other studies have employed diverse methodologies. Thus, Southgate, Senju and Csibra (2007) recorded children's anticipatory looking (AL) – i.e. where the children first looked – when an agent returned to a scene to search for an object, which had been transferred to a new location. As in the aforementioned study by Clements and Perner, Southgate and colleagues found that young children (in this case 25 month-olds) indeed first looked toward the location where the agent was likely to search on the basis of a false belief. In a different study using a very different methodology, Buttelmann and colleagues (2009) developed an 'active helping' paradigm, in which 18 month-olds had to take into account an agent's false belief in order to determine his goal so that they could help him to attain that goal. Furthermore, Southgate, Chevallier & Csibra (2010) found 17 month-olds interpret referential communication in order to learn new words in a way that is sensitive to adults' false beliefs about the location of the object to which they are referring.

In addition to the variety of the measures that have now been used to find evidence for false belief reasoning in infants, recent studies have also revealed that children are sensitive to various ways in which agents acquire belief. In a study involving 15 month-olds, for example, Träuble et al. (2010) used an apparatus designed such that an agent could cause a ball to be transferred from one bucket to another by manipulating the apparatus without seeing it (with her back turned). The finding was that infants expect an agent not to have a false belief even though she did not see the object transfer because she was turned the other way. This demonstrates an impressive ability to reason flexibly about the effects that various kinds of evidence

(even non-perceptual evidence) will have on agents' beliefs. In other words, the infants must recognize a pattern insofar as they must interpret the adult agent's manipulation of the apparatus as being relevantly similar to (and thus constituting a pattern with) the agent's visual perception of the location of the ball. Similarly, Song, Onishi & Baillargeon (2008) found that 18 month-old infants' expectations are modulated if the experimenter *communicates* to the agent that the ball has been moved but not if she says merely that she likes the ball. The capacity to treat various sources of information about the location of an object as equivalent is important insofar as it suggests that the children are reasoning about the underlying mental representations (i.e. the beliefs) that arise as a result of various kinds of observable events.

Impressive as these recent findings are, there are at least four reasons not to think that they provide sufficient reason to abandon the received view. First of all, there are viable theoretical alternatives that do not interpret the extant findings as evidence that children ascribe beliefs to other agents in the 2nd year of life. For example, according to one influential interpretation advocated by Perner and Ruffman (2005; see also Ruffman 2014), the children in these studies succeed by using complex behavioral rules, such as 'agents retrieve objects at the locations where they last saw them'. On the one hand, this interpretation may appear to be increasingly unattractive as results are reported from ever more diverse paradigms (for an overview, see Baillargeon et al. 2010; Butterfill and Apperly 2013), since it may be necessary to postulate new behavioral rules ad hoc in order to account for each new finding. On the other hand, it may be possible to formulate a small set of behavioral rules that are sufficiently general to account for most or all of the findings, so this interpretation should be, and is, still regarded as a live option. A different deflationary

interpretation, proposed by Heyes (2014), is that the infants' looking time in the paradigms employed so far is a function of the degree to which the observed or expected low-level properties of the test stimuli – i.e. colors, shapes and movements – are novel with respect to events observed by the infants at earlier time-points during the experiment. If Heyes is right, false beliefs may actually be confounded with low-level novelty in most of the existing paradigms. A further, slightly less deflationary, option is to argue that children in the 2nd year of life do indeed succeed at the tasks in question by ascribing mental states to the agents in question – but not *beliefs*. Apperly and Butterfill (2009; see also Butterfill and Apperly 2013) for example, have developed an account according to which there are two systems for mindreading: an early-developing system which generates representations of non-propositional states, and a later-developing system capable of representing beliefs as propositionally structured mental representations. Thus, the early-developing system has the capacity to represent another agent as 'registering' an object as being where she last 'encountered' it, i.e. as being where it was when it last entered her perceptual field. The important point here is that, according to Apperly and Butterfill, registrations are distinct from beliefs: while beliefs are *representations* (e.g. of the locations of objects) registration is a *relation* among an agent, an object and a location. Hence, representations of beliefs are metarepresentational, while representations of registrations are not. In other words, in order for one agent to represent some other agent's registration of an object at a location, then, it is not necessary for the first agent to be capable of representing representations per se. On these grounds, Apperly and Butterfill argue that the early-developing system – which would account for the extant findings – does not rely upon a concept of belief.

Secondly, a recent longitudinal study conducted by Thoermer et al. (2012)

found evidence that favors the two aforementioned lean interpretations over a richer interpretation. Specifically, they found that 18 month-olds who passed an implicit FB task were more likely to pass an explicit verbal FB task about locations at four but were not more likely to pass an explicit verbal FB task about contents. Thus, the longitudinal continuity was *task-specific*. This is just what one would expect if one endorses a lean interpretation: if children construct behavioral rules by gradually acquiring familiarity with specific situations that they frequently encounter, then their performance should improve most quickly in situations that they frequently encounter and less quickly in situations that they encounter less frequently. In contrast, the findings speak against richer interpretations of the infant mindreading data: according to rich interpretations, the capacity that is tapped in implicit mindreading tasks in the 2nd year is the capacity to represent beliefs in general (i.e. it is the belief concept), so it is not clear why the continuity between the early-emerging capacity and later capacities should be specific to particular tasks¹.

Thirdly, in order to evaluate the relevance of this body of research for the question as to whether the desire reasoning is developmentally prior with respect to belief reasoning, it would be necessary to compare these findings with research on desire understanding that also employs implicit measures. For it is likely to be the case that, just as implicit reasoning about beliefs occurs earlier than explicit reasoning about beliefs, implicit desire reasoning also occurs earlier than explicit reasoning about desires. If so, the developmental priority of desire reasoning at the explicit level may simply be mirrored at the implicit level. In fact, the account that we will be offering below, as we shall see, provides a reason to predict that this is indeed the

¹ To be fair, it must be acknowledged that rich interpretations could accommodate this finding by identifying task-specific differences that account for the performance differences, so it is only weak evidence in favor of leaner interpretations.

case. Moreover, this conjecture is supported by some research using implicit measures (e.g. looking time), which has shown that infants are sensitive to other agents' goals by as early as 3 months (Somerville et al. 2005; for a review of similar research, see Woodward 2009).

Fourthly, and most importantly for our purposes, if a rich interpretation of the infant mindreading data turns out to be right, this still would not alter the fact that children are able to operate with the concept of desire on explicit tasks before they are able to do so with the concept of belief. Thus, a revised version of the received view would remain intact, namely that children are *more* proficient earlier at reasoning about desires than at reasoning about beliefs. In other words, the priority may not be as extreme as some may have previously thought, but it would nevertheless remain an important explanandum.

In the following, we will leave open the question of how best to interpret the infant mindreading literature, and will adopt the modest assumption that it is at least the case that children are more proficient earlier at reasoning about desires than at reasoning about beliefs. Our task, then, will be to consider what might explain this phenomenon.

3. Previous explanations

Despite the dominance of the view that desire reasoning develops prior to belief reasoning, relatively little has been done to *explain* this striking phenomenon. The explanations that have been offered so far in the literature can be divided into two main groups, in accordance with the two most prominent competing general theoretical approaches to the acquisition mechanisms underpinning the development of mindreading: one broadly nativist approach, which maintains that the capacity to

represent beliefs and desires is already present in infancy but is limited by performance factors and by the slower development of computational-cognitive competences extraneous to mindreading (main proponents include e.g. Alan Leslie and Tim German); and a broadly constructivist family of approaches, that instead stresses conceptual change, either in domain-specific theoretical knowledge (e.g. Alison Gopnik and Henry Wellman) or in domain-general representational capacities that are required in order to form representations of beliefs and desires as such (Josef Perner). In this section, we shall review these explanations in turn, and argue that they are insufficient on their own. In Section 4.3, which is devoted to the relationship between our proposal and the existing explanations, we also discuss the place of our explanation in this broader theoretical landscape.

Within the nativist approach to acquisition mechanisms in the development of mindreading, the most prominent explanation of the order of concept acquisition is that ascribing beliefs is more difficult than ascribing desires due to increased executive demands, and that this accounts for the later onset of competence with the concept of belief (German & Leslie, 2001; Leslie 1994, 2000; Leslie et al., 2004; Moore et al., 1995; Russell, 1996). The reason for this is that while ascribing beliefs often involves inhibiting one's own cognitive perspective in representing others as having beliefs that conflict with one's own beliefs, ascribing desires does not necessarily involve such inhibition, since there doesn't need to be a similar conflict or incompatibility in desires, even if they differ from one's own desires. However, the aforementioned experimental findings reported by Wellman and Woolley (1990) and by Rakoczy et al. (2007) make this explanation doubtful. To recall, Wellman and Woolley found that two year-old children were able to predict an agent's actions on the basis of his desires and ascribe desire-dependent emotions to him even when their

own desires differed from those of the agent. Indeed, the authors deliberately contrasted this with a condition in which the agent had different beliefs rather than different desires, and the results show that children were much better at inhibiting their own desires than at inhibiting their own beliefs. Similarly, Rakoczy et al's findings showed that 3-year olds were quite proficient at ascribing incompatible desires to two persons, even though they mostly failed at analogous false-belief tasks of comparable complexity. Here, again, the tasks involving ascription of incompatible desires required inhibition of the exact same kind that should have accounted for the later onset of competence with the concept of belief. Thus, it seems doubtful that the decisive difference between reasoning about beliefs and reasoning about desires is that in the latter case one has to inhibit one's own perspective, even if it does play a contributing role.

Within the constructivist family of approaches to acquisition mechanisms in the development of mindreading, the developmental priority of desire reasoning has often been thought to pose a serious theoretical challenge. Indeed, domain-general constructivist accounts have previously been assumed to be *incompatible* with the developmental priority of desire reasoning. Specifically, versions of the theory-theory that explain the improvement in performance on explicit verbal false belief tasks around the age of 4 by postulating the emergence of a metarepresentational capacity (i.e. the ability to represent representations per se) have until now been embarrassed by the developmental priority of desire reasoning (Perner 1991a; 1991b). This is because the concepts of belief and desire appear to draw in equal measures upon the capacity to metarepresent, so if children reason with the concept of desire in the second year of life, then it seems that they must already be capable of metarepresentation by that time. But if this is so, then it could not be the case that the

capacity to metarepresent emerges around 4 and accounts for the sharp improvement in performance on explicit verbal false belief tasks among 4 year-olds. Domain-specific constructivist accounts like Gopnik's, on the other hand, are not specifically undermined by the developmental priority of desire-reasoning, since they are compatible with the possibility that children develop a theory of desire years before they develop a theory of belief. However, while they are compatible with this possibility, they do not offer any positive reason to predict it.

As possible remedies for these shortcomings of the constructivist/conceptual-change approach, Rakoczy et al. (2007) suggest two alternative explanations of the priority of desire reasoning by appeal to *pragmatic* and *experiential* factors, and by appeal to certain *logical features* of the concepts of belief and desire, which makes belief reasoning more demanding than desire reasoning.

On the first line of thought, Rakoczy et al. point out that children experience significantly more explicit reference to conative attitudes than to cognitive attitudes, for several reasons: explicitly stated psychological explanations are often elliptical in the sense that they take for granted a common world, and instead focus on the explanatory desires; young children often participate in cooperative activities, but what has to be coordinated in these are rarely beliefs, but more often desires and plans (e.g. Bratman, 1992); finally, when young children take part in conflicts, the conflicts typically concern what has to be done rather than what is the case (e.g. Dunn & Slomkowski, 1992; Schantz, 1987). We think that this appeal to pragmatic and experiential factors, though illuminating, cannot suffice to explain the developmental priority of desire reasoning. The first consideration, that young children are exposed to fewer explanations with explicit reference to belief, because psychological explanations are often elliptical, is correct, but this is something that should itself be

explained. Similarly, if the earlier emergence of desire reasoning is to be explained by suggesting that they frequently engage in cooperative activities in which desires and plans need to be coordinated rather than beliefs, and in conflicts about what is to be done rather than conflicts about what is the case, then these latter phenomena need to be explained. An important feature of the account of the priority of desire reasoning to be offered in the next section is that it explains why folk-psychological explanations are often elliptical and make fewer references to beliefs than to desires, and explains why cooperative activities require coordination of desires and plans more often than of beliefs. Another related reason not to be satisfied with the pragmatic-experiential explanation alone is that it doesn't explain the high degree of consistency in the order of concept acquisition exhibited by the data: if the developmental priority of desire reasoning is to be explained by the degree to which young children are exposed to psychological explanations with explicit reference to beliefs and desires respectively, then we might expect more variation in the observed developmental sequence, since it is reasonable to expect that there will be variations in the degree to which young children are exposed to explicit reference to belief and desire respectively.

The second complementary line of explanation proposed by Rakoczy et al. (2007) appeals to certain differences in the 'logical properties' of beliefs versus desires, which motivate the idea that the concept of desire should be 'easier to understand' and thus 'less psychologically taxing' than the concept of belief (see also Perner, 1991a, 1991b). The first important logical difference concerns the respective kinds of normative evaluation that beliefs and desires are subject to. While the normative aim or ideal for beliefs is that they are *true*, such that any belief with a false content ought to be revised, the normative aim or ideal for desires is that they are *fulfilled*, such that the world is 'revised' in accordance with the desire (this difference

is sometimes understood as a difference in ‘direction of fit’; see e.g. Anscombe 1957; Searle 1983; Steglich-Petersen 2006, 2009). According to Rakoczy et al., this means that beliefs are subject to instantaneous normative evaluation (*is it true?*), while desires are mostly future-related, and thus only to be evaluated for fulfillment at a later stage. But it is unclear why this should result in the concept of belief being more psychologically taxing than the concept of desire. Rakoczy et al. claim that when we ascribe beliefs, the question of truth (and thus normative correctness) arises at the time of ascription, and for that reason the default case is to ascribe true beliefs. In the case of desires, by contrast, at the time of ascription the question of fulfillment does not necessarily arise, and it is therefore not necessarily the default case that the desire will be fulfilled. But why should this make it more psychologically taxing to ascribe beliefs? One could just as easily turn the explanation on its head and argue that these considerations support the *opposite* conclusion, namely that the concept of desire is more psychologically taxing than that of belief. If the default when ascribing beliefs is to ascribe true beliefs, then ascribing beliefs does not normally involve representing the world to be different from how the person ascribing the beliefs represents it at the time of ascription. But if the default when ascribing desires is to ascribe desires that are not yet fulfilled, this must involve representing a state of affairs (i.e. the content of the desire) that is believed not yet to exist by the person ascribing it. In this respect, ascribing desires thus involves a higher degree of cognitive suppression than ascribing beliefs. This makes it unclear how the suggested differences in logical properties explain the order of concept acquisition.

It should be stressed that the explanations suggested by Rakoczy and colleagues are not presented as being jointly sufficient, but as contributing explanations, so proponents of these explanations may not object to the limitations we

have pointed out above. As we shall see in Section 4.3, the explanation we will be proposing in the following section may indeed complement some of the suggestions made by Rakoczy and colleagues.

4. A new explanation

The explanation that we will suggest in the following is instead based on considerations concerning the cognitive function of belief and desire reasoning, i.e. the cognitive tasks that the ability to reason with these concepts enable us to perform. As mentioned above, the standard view is that the main function of reasoning with these concepts that of *predicting* and *explaining* the behavior of creatures that are motivated by beliefs and desires. As we shall argue, however, belief and desire reasoning exhibit a crucial difference in just this regard: while it is typically feasible to predict and explain behavior by ascribing desires only, while abstaining from ascribing beliefs, the reverse operation is *not* typically feasible; it is not typically feasible to predict and explain behavior by ascribing beliefs only. If this is the case, a simple yet powerful explanation of the developmental priority suggests itself. If the main function of belief reasoning *cannot* typically be fulfilled in the absence of ability to reason with the concept of desire, while the main function of desire reasoning *can* often be fulfilled in the absence of ability to reason with the concept of belief, it is no wonder that desire reasoning develops prior to belief reasoning. There would quite simply be no or little cognitive use for belief reasoning before the ability for desire reasoning is acquired; competence with belief reasoning would provide little cognitive advantage to its holder, thus explaining the later onset of that competence.

Our argument will proceed in two steps. First we will show that it is possible, indeed commonplace, to interpret behavior in terms of desires only. We then show that the analogous task of interpreting behavior by ascribing beliefs only is typically much less feasible, if not altogether impossible. Along the way, we point to some general features of the concepts of belief and desire, which explain this difference.

4.1 Interpreting behavior by ascribing desires only

The first and easiest step is to show that it is possible to predict and explain behavior by invoking the concept of desire only. To see how this is possible, it will be useful to operate with an example of a simple interpretation task. Suppose that some agent to be interpreted, *Sally*, desires cookies, and believes that there are cookies to be found in a closed cookie jar. Suppose that these mental attitudes cause Sally to open the cookie jar. Now imagine another agent, *Bob*, who observes Sally's behavior and tries to interpret it. A 'full' folk-psychological interpretation of Sally's behavior on Bob's behalf would consist of Bob having the following beliefs, where the brackets indicate the contents of Bob's beliefs:²

(a) Bob believes that [Sally desires cookies]

Bob believes that [Sally believes that there are cookies in the cookie jar]

What is important to note about this interpretation, is that it requires that Bob is able to represent both beliefs and desires, since the propositional contents of Bob's beliefs

² Strictly speaking, an interpretation of Sally's behavior by Bob would also include Bob believing that the bracketed propositions jointly explain/predict Sally opening the cookie jar. Since an additional belief of this kind plays no role in the following, we ignore it throughout this paper.

indicated by the brackets include both of these concepts. So what we have called a ‘full’ folk-psychological interpretation involves competence with both concepts.

But it seems that Bob could also interpret Sally’s behavior in a way that does not invoke the concept of belief. This is possible because Bob can rely on his *own* beliefs about the cookie jar in making his interpretation. Such a ‘limited’ interpretation of Sally by Bob would consist in Bob having the following beliefs:

(b) Bob believes that [Sally desires cookies]

Bob believes that [There are cookies in the cookie jar]

Here, Bob ascribes a desire for cookies to Sally, and predicts that Sally will open the cookie jar in order to get cookies, based on *his own* belief that there are cookies in the cookie jar. What is important to note about this interpretation is that although it involves Bob *having* certain beliefs, neither of these beliefs *represent* beliefs as part of their propositional contents, and the interpretation does therefore not require Bob to reason with the concept of belief.

This kind of interpretation is limited in certain important ways. For it to be successful, it is necessary that the interpreter and the interpreted *share* the beliefs that are relevant to the interpretation. If Bob doesn’t believe that there are cookies in the cookie jar, he will not be able to predict or explain Sally’s opening the jar, despite his belief that Sally desires cookies. The interpretation will also fail if Bob believes that there are cookies in the jar, but Sally doesn’t. So whenever there is a difference between the interpreter and the interpreted in regard to the relevant beliefs, it is necessary for the interpreter to ascribe beliefs to the interpreted (beliefs that the

interpreter will himself regard as false), and the strategy of simply relying on one's own beliefs will not be available.

However, it is reasonable to suppose that the relevant beliefs are shared often enough to make this a reliable interpretation strategy, especially when it comes to the kind of simple interpretation tasks infants most often engage in. Such interpretation tasks will often be directed towards behavior motivated by beliefs about the immediate spatial surroundings of the infant and the agent being interpreted. And – crucially – *these kinds of beliefs* are rarely mistaken. In other words, while beliefs about distant or absent objects and events are prone to error, the kinds of beliefs that motivate most of our everyday actions, and especially most of the everyday actions that young children are interested in, are beliefs about concrete objects that are available for inspection in our immediate spatial surroundings. Although the strategy of relying on one's own beliefs about the objects in the immediate spatial surroundings and only ascribing desires will sometimes fail, it is thus reasonable to suppose that it will succeed often enough to be useful for infants to rely on it, and that having competence with desire reasoning provides a cognitive advantage even in the absence of belief reasoning.

In fact, it seems that the strategy of relying on one's own beliefs in interpreting others is rather commonplace, both for adults and young children. In our everyday interpretations of each other, we very often rely on desire ascriptions only, and simply ignore beliefs on the tacit assumption that beliefs about the relevant states of affairs are shared ('he flipped the switch because he wanted light', 'she went to the grocery store because she needed milk', etc.); a study carried out by Wellman & Woolley (1990) suggests that the same is the case for young children.

4.2 Interpreting behavior by ascribing beliefs only?

So far, it has only been shown that it is often feasible to interpret behavior by ascribing desires only, while relying on one's own relevant beliefs. But in order to explain the developmental priority of desire reasoning, it must also be shown that the reverse operation is *not* typically feasible, or at least that it is typically substantially *less* feasible, to interpret behavior by ascribing beliefs only.

How should we think of an interpretation that ascribes beliefs, but instead of ascribing desires relies on the interpreter's own desires in making the interpretation? Going back to the example used above, what we are looking for is a desire of Bob that could make the following interpretation of Sally's behavior successful, keeping in mind that the neither of Bob's attitudes can represent desires as part of their propositional contents:

- (c) Bob believes that [Sally believes that there are cookies in the cookie jar]
Bob desires that [... ? ...]

What possible content of Bob's desire could make this interpretation successful? The simplest possible candidate for a desire of Bob that could help interpret Sally's behavior, is a desire for *himself* to have cookies, resulting in the following interpretation:

- (d) Bob believes that [Sally believes that there are cookies in the cookie jar]
Bob desires that [Bob gets cookies]

In this explanation, Bob interprets Sally's behavior by ascribing a belief to Sally that there are cookies in the cookie jar, and then relying on his own desire for cookies in explaining her behavior. But there is an obvious problem with this interpretation, namely that the desire is *personal* in the sense of including Bob himself as a constituent, and it is therefore not strictly speaking shared with Sally. Sally, after all, desires that *she* get cookies, not that Bob does. So even if they both desire cookies, the contents of their desires are not the same – they each desire for *themselves* to get cookies. And needless to say, it is unclear how a desire of Bob for *himself* to have cookies could help explain why Sally would engage in behavior that would help *her* get cookies.

This suggests the next obvious candidate for a desire of Bob that could be relied upon in interpreting Sally's behavior, namely a desire that *Sally* gets cookies. This would result in the following interpretation by Bob of Sally's behavior:

- (e) Bob believes that [Sally believes that there are cookies in the cookie jar]
Bob desires that [Sally gets cookies]

In contrast to the first candidate desire we considered, this desire is a desire for a relevant state of affairs, namely the very state of affairs which Sally herself desires and is motivated to bring about by opening the cookie jar. The desire can thus be considered to be shared by Bob and Sally: they both desire that Sally get cookies. This interpretation, however, suffers from a serious limitation: it requires that Bob share a desire with Sally, this time a desire for Sally to get cookies, but it is highly doubtful that Bob will actually share the relevant desire in very many cases – especially now that it is a desire for *someone else* to have cookies. Infants, or adults

for that matter, simply cannot be assumed to share desires for someone else to achieve something in enough cases to make the strategy of relying on one's own desires in this way reliable. Of course, we could imagine other candidate desires that are less specific in target, such as a generalized desire that *everyone* have cookies. But it seems that any such desires will be of limited use in interpretation for the reasons already mentioned: although it would be a desire for a relevant state of affairs (Sally desiring that *everyone* have cookies would, among other things, motivate Sally to provide cookies for *herself*), it is highly doubtful that the interpreter will actually have the relevant desire in sufficiently many cases to make the strategy reliable.

A more general problem with the strategy of relying on one's own desires as in models (d) and (e), is that it is unclear how an interpretation could be made up in part by a desire of the interpreter. It was mentioned above that an interpretation of Sally by Bob consists of Bob having a number of relevant *beliefs*, which may or may not include the concept of belief as part of their propositional contents. This is due to the simple fact that when one adopts an explanation, or makes a prediction, one *believes* the propositions that explain, or serve as a basis for prediction of, the relevant behavior. This is simply what it means to hold or adopt an explanation or prediction to be true. By contrast, it is not clear how a desire of mine by itself could be the state of mind in virtue of which I adopt an explanation, or make a prediction. This may seem like question-begging in the context of trying to decide whether an interpreter's own desires could form part of an interpretation, but it does seem difficult to understand how this could be so. This does not by any means show that the desires of the interpreter could play no role in an interpretation. What it shows is that the interpreter would need to form beliefs about those desires in order for them to play such a role, such that the interpreter's beliefs about his own desires form part of the

interpreter's interpretation; but if that is needed, the interpreter must be able to represent desires, otherwise no beliefs about those desires could be had. It is thus unclear how an interpreter could rely on his own desires in his interpretation without being able to represent and reason with desire. Since this is not the case when it comes to relying on one's own beliefs, this lends further support to the claim that there is an important kind of interpretation that can be performed relying on desire reasoning, without relying on belief reasoning, but not vice versa, which explains why belief reasoning develops later than desire reasoning.

A clue for a more promising way of articulating how one could interpret others' behavior by ascribing beliefs but without ascribing desires is based upon the objective desirability account which we reviewed above (in sec. 2.1). Recall that the objective desirability account, if correct, would explain how children can predict others' behavior in a range of cases and ascribe desire-dependent emotions without a concept of desire or a concept of belief. Thus, the account would not explain why belief reasoning would be useful without desire reasoning, nor is it intended to do so. However, a modified version of the account could potentially provide such an explanation. On this modified version, one may use a concept of belief and a notion of objective desirability in interpreting behavior. There are several ways one could understand this strategy. According to the most straightforward version, the assessment of some relevant object or state of affairs as objectively desirable figures as a *belief* having that assessment as its propositional content. Applying this to our example, the explanatory schema would look something like this:

(f) Bob believes that [Sally believes that there are cookies in the jar]

Bob believes that [Cookies are objectively desirable]

The assessment of something as ‘objectively desirable’ is of course highly general, and will typically be expressed in some more particular and context-appropriate way. For example, in the above case, the relevant belief might be more realistically thought of as a belief that ‘Cookies are yummy’, or something similar. But there is another and perhaps more plausible way of understanding the ‘objective desirability’ interpretation. According to this, the assessment of some relevant object or state of affairs as objectively desirable consists in the agent *valuing* that object or state of affairs, and tacitly assuming that this valuing attitude is universal, where the state of valuing something is not a propositional attitude but simply a relation between the agent and the relevant object or state of affairs. The advantage of this version over the previous one is that there can be no doubt that the relevant attitude can be had without employing any version of the concept of desire. Applying this to the example, we arrive the following schema:

- (g) Bob believes that [Sally believes that there are cookies in the jar]
Bob values cookies

Both of these proposals avoid several of the problems afflicting the previous proposals (d) and (e). Neither of the proposals involves *personal* desires concerning either the interpreter or the interpreted, and neither of the proposals demands that the interpreter have general desires on behalf of *everyone*. It avoids this by instead ascribing a property to the relevant object or state of affairs (desirability), or by consisting in a relation to it (valuing it). Furthermore, in neither case is the interpretation supposed to consist partly in a desire, in a way that would give rise to

the puzzle pointed out above about the nature of holding an interpretation or an interpretation to be true. Granted, in (g), the interpretation consists partly in Bob valuing something and tacitly assuming this attitude to be shared. But it is perhaps easy enough to see how such a tacit assumption can sometimes structure interpretations. Indeed, it is easy to think of interpretations that are elliptical in leaving out explicit reference to goals or values in cases where the goals and values structuring the relevant situation are sufficiently clear and general (example: ‘Why did she take the penalty shot to the right? Because she saw the goalkeeper go to the left!’).

However, while interpretation strategies such as those modeled by (f) and (g) may be possible to rely on in some cases, it should be clear that they will be feasible in substantially fewer cases than the analogous strategy relying on shared beliefs exemplified by (b), especially for infants and young children. The reason for this is that while it is plausible to suppose that the majority of the *beliefs* needed to make the kinds of interpretations infants and young children engage in are shared (this was argued in Section 4.1), it is much less plausible to suppose that the relevant assessments of objects and states of affairs as *objectively valuable* will be shared sufficiently frequently. Just think of the many actions of adults in pursuit of objects and states of affairs that the infant or child do not find valuable or worth pursuing, but that nevertheless would be useful for the child or infant to interpret. But the *beliefs* informing those actions, i.e. beliefs about the spatial location of relevant objects, etc., *will* often be shared. It may be objected that some *highly general* value assessments may be shared in many cases, e.g. the valuing of tasty food, having a good time, absence of pain, etc. But unless there’s also a shared opinion about what particular objects or activities *constitute* those values, such general assessments cannot be relied

upon in interpreting particular actions. For example, while infant and adult may share the assessment of good food as valuable, they are likely to differ in what they regard as good food, thus requiring understanding of an assessment that differs from one's own. It is worth pointing out that even if an infant or child interpreter occasionally does share a relevant desire or value assessment with the interpreted agent, this does not suffice to make the strategy of relying on one's own desires or value assessments viable. In order for the strategy to be viable, it must be sufficiently reliable, i.e. successful in a sufficiently high proportion of instances, since predictions and explanations of behavior based on it would otherwise not turn out correct in a sufficiently high proportion of cases to reinforce the reliance on own desires or value assessments. So it is not enough for it to only work occasionally.³

The search for a relevant shared desire or value assessment that could be relied on in interpreting behavior has thus led us to an impasse. Going back to the example, if Bob relies on a desire for himself to have cookies, it cannot explain Sally's quest for *herself* to have cookies. If Bob desires that *Sally* have cookies, he will share a relevant desire with Sally, but it is highly unlikely that desires for *someone else* to achieve something occur frequently enough (especially among infants) to make the strategy of relying on such 'altruistic' desires sufficiently reliable for reinforcement and thus learning. The same will be the case if Bob instead relies on an objective value assessment of cookies as yummy. Although the assessment of cookies as yummy may be shared, the majority of actions pursue objects that children or infants will *not* assess as objectively valuable.

³ One possible instance of shared desires between young children and adults occurs when desires are *transferred* from adults to children, i.e. when an adult's desire causes a child to form the same desire, sometimes referred to as 'mimetic' desire (for a classic general treatment, see e.g. Girard 1965). The occasional effectiveness of such a mechanism, however, does not plausibly undermine the main thesis of insufficient frequency. We thank an anonymous reviewer for bringing this to our attention.

4.3 Revisiting existing explanations

How does our proposal relate to the existing explanations discussed in Section 3? And where does it fit with respect to the two broad theoretical approaches to acquisition mechanisms in the development of mindreading, that the existing explanations form part of? It is first of all worth emphasizing that the proposed explanation need not be in competition with the existing ones discussed above. Although we have raised a number of questions about those explanations, and argued that they are incomplete as they stand, all of the explanations considered in Section 3 are in principle compatible with ours, and may be seen as each playing a part in a broader explanation. Perhaps more remarkably, this means that our explanation could be adopted by both sides of the divide between the theoretical approaches to which the existing explanations belong, although perhaps somewhat more naturally by the constructivist approach than by the nativist one.

Consider first the explanation provided by the nativist approach, focusing on underlying computational/cognitive capacities. This explanation rested on the claim that ascribing beliefs that differ from one's own is more taxing than ascribing desires that differ from one's own because only the former requires inhibition of one's own perspective. The primary objection to this proposal was that 3 year-olds are more proficient at the desire-reasoning than belief-reasoning even when inhibition of the same kind is required in both cases. Nevertheless, if one wanted to defend this explanation, one way to do so would be to argue that even though reasoning about desires may also frequently require inhibition of one's own perspective in just the same way as reasoning about beliefs that differ from one's own, children develop the capacity to inhibit their own desires earlier than their own beliefs. Our proposal could be drawn upon to complement this explanation by providing a reason why children

develop the capacity to inhibit their own desires before their own beliefs – namely that doing so is more useful to them in that it enables them to understand a comparatively broader range of behavior, than a capacity for inhibiting their own beliefs would.

Likewise for the explanation appealing to pragmatic and experiential factors, championed by Rakoczy et al. (2007) as part of the constructivist approach focusing on conceptual development. That explanation was based on the observation that children may have more practice reasoning about desires because they frequently engage in cooperative activities in which desires and plans need to be coordinated rather than beliefs, and in conflicts about what is to be done rather than conflicts about what is the case. Our primary concern with this explanation was that it fails to explain these latter phenomena. Indeed, these latter phenomena appear to presuppose what they are supposed to explain insofar as children need to reason about desires in order to be able to engage in coordination of different desires or to quarrel over them in the first place. However, our proposal may contribute to explaining why such negotiations over desires arise more frequently than negotiations over beliefs, namely by pointing out that people's desires differ more frequently than their beliefs (at least those beliefs that are relevant to understanding their behavior in most cases), which is also what makes it useful for children to become proficient at reasoning about desires prior to their becoming proficient at reasoning about beliefs. Once this underlying reason is pointed out, the pragmatic-experiential explanation may then go some way toward accounting for how young children's ability with the desire concept develops and becomes reinforced by experience.

More generally, and perhaps also more surprisingly, our proposal may complement constructivist accounts of theory of mind development, which have

previously been assumed to be in tension with the developmental priority of the concept of desire. This is especially pertinent to versions of the theory theory that explain the improvement in performance on explicit verbal false belief tasks at around 4 by postulating the emergence of a metarepresentational capacity have been thought to be undermined by the developmental priority of the concept of desire (Perner 1991a; 1991b). However, our proposal provides a reason to expect that desire reasoning would emerge prior to belief reasoning even if they do rest upon the same (meta)representational capacity – i.e. because desire reasoning is much more useful in the absence of belief reasoning than vice versa. Thus, proponents of the metarepresentational account could avail themselves of our proposal in order to explain the developmental priority of desire reasoning, which otherwise threatens to undermine their account. Our proposal could also contribute to an understanding of the learning mechanisms involved in becoming proficient in desire- and belief-reasoning, that would fit well within the overall constructivist framework, by providing an insight into why success, and thus reinforcement in learning, comes earlier in the case of desire-reasoning than in belief-reasoning. If so, then this could provide proponents of both domain-general (Perner) and domain-specific (Gopnik) versions of the constructivist approach with a principled reason to predict the developmental priority of desire-reasoning.

Although we cannot here develop these possible relationships between the existing explanations and our proposal, we hope at least to have shown that there is good reason to believe that they can complement each other in important ways.

4.4 Directions for future research

Before concluding, we would like to highlight two new directions that our proposal opens up for future empirical research, one developmental and one that is non-developmental. First, our proposal provides grounds for expecting that the developmental priority of explicit desire-reasoning is mirrored at the implicit level. In other words, just as recent research on belief-reasoning in infancy and early childhood (which we surveyed in section 2.2) has provided evidence that implicit belief-reasoning may be in place long before children succeed at explicit verbal false belief tasks, our proposal suggests that implicit desire-reasoning emerge much earlier than explicit desire-reasoning as well, and indeed earlier than implicit belief-reasoning. If so, then for any paradigm that has been used to probe implicit belief-reasoning in infancy, we would predict that infants should succeed in a version of the paradigm that is suitably modified to probe implicit *desire*-reasoning at an even earlier age (assuming well-matched task demands).

Secondly, our proposal also motivates the hypothesis that desire-reasoning is more fundamental than belief-reasoning in adults. If this is right, then adults should be more likely to neglect beliefs than to neglect desires if, for example, subjected to time pressure or cognitive load on a task requiring both belief- and desire-reasoning. In such a scenario, we would therefore predict that they would make more mistakes in belief-reasoning than in desire-reasoning. Similarly, we would predict that desire-reasoning is more likely than belief-reasoning to be performed spontaneously or automatically in any paradigm in which neither one is relevant. Interestingly, there is some evidence from recent research suggesting that adults tend to track others' beliefs spontaneously in various paradigms in which it is actually task-irrelevant, but that they do not do so automatically, and various factors can prevent them from doing so

(Apperly et al. 2006; Surtees et al. 2012; for a review, see Apperly 2011, chap. 5). In order to test our prediction, it may be possible to adapt some of the paradigms employed in this research in order to contrast the degree to which belief-reasoning and desire-reasoning are performed spontaneously or automatically.

5. Conclusion

In the foregoing, we have reviewed two lines of argument against the received view that children tend to be more proficient earlier at reasoning with the concept of desire than with the concept of belief, and argued that both of those lines of argument fail to undermine the consensus that there is a developmental priority of desire reasoning that demands explanation. We have also reviewed existing explanations of this phenomenon, and found them to be flawed or incomplete on their own. We then showed that a simple and powerful explanation is available, which instead focuses on the cognitive tasks that competence with ascribing and reasoning with the concepts of belief and desire enable us to perform. We showed that the most important of these, namely that of interpreting the behavior of others, can be performed using the concept of desire only, but that it is not typically feasible using the concept of belief only. This means that there would be little advantage in being competent with belief reasoning before having acquired competence with desire reasoning, thus contributing to explaining the observed developmental priority. We have also shown that the proposed explanation has some surprising consequences for existing theoretical accounts, and that it generates novel hypotheses to be explored in future research. Finally, we suggested that it may be possible to integrate the proposed explanation with some of the existing explanations, revealing them to be more successful as components that function within a larger explanation than as stand-alone explanations.

References

- Anscombe, G.E.M. 1957: *Intention*. Oxford: Basil Blackwell.
- Apperly, I. A. 2011: *Mindreaders: The cognitive basis of theory of mind*. New York: Psychology Press.
- Apperly, I. A., & Butterfill, S. A. 2009: Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116, 953–970.
- Apperly, I. A., Riggs, K.J., Simpson, A., Chiavarino, C., & Sampson, D. 2006: Is belief-reasoning automatic? *Psychological Science*, 17, 841-844.
- Baillargeon, R., Scott, R. M., & He, Z. 2010: False-belief understanding in infants. *Trends in Cognitive Sciences*, 14, 110–118.
- Bartsch, K., & Wellman, H.M. 1995: *Children talk about the mind*. New York, NY: Oxford University Press.
- Bratman, M. 1992: Shared cooperative activity. *The Philosophical Review*, 101, 327-41.
- Butterfill, S. A., & Apperly, I. A. 2013: How to Construct a Minimal Theory of Mind. *Mind & Language*, 28, 606–637.
- Dunn, J., & Slomkowski, C. 1992: Conflict and development of social understanding. In C.U. Shantz, & W. Hartup (eds), *Conflict in child and adolescent development*. Cambridge: Cambridge University Press.
- German, T.P. & Leslie, A.M. 2001: Children's inferences from 'knowing' to 'pretending' and 'believing'. *British Journal of Developmental Psychology*, 19 (Pt 1), 59-83.
- Girard, R. 1965: *Deceit, Desire, and the Novel*. Washington DC: Johns Hopkins University Press.
- Hadwin, J., & Perner, J. 1991: Pleased and surprised: Children's cognitive theory of

- emotion. *British Journal of Developmental Psychology*, 9, 215-34.
- Heyes, C. 2014: False belief in infancy: a fresh look. *Developmental Science*, DOI: 10.1111/desc.12148
- Kovács, A., Teglas, E., & Endress, A. 2010: The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, 330, 1830–1834.
- Leslie, A.M. 1994: Pretending and believing: Issues in the theory of ToMM. *Cognition*, 50, 211-38.
- Leslie, A.M. 2000: Theory of mind as a mechanism of selective attention. In M.S. Gazzaniga (ed.), *The New Cognitive Neurosciences* (2nd ed., pp. 1235-47). Cambridge, MA: MIT Press.
- Leslie, A.M., Friedman, O., German, T.P. 2004: Core mechanisms in 'theory of mind.' *Trends in Cognitive Sciences*, 8, 528-33.
- Lichtermann, L. 1991: *Young children's understanding of desires. Third year project report*, unpublished manuscript. As cited by Rakoczy et al., 2007.
- Marsh, H., Legerstee, M., Stavropoulos, J. and Nienhuis, T. 2010: Six-and 9-month-old infants discriminate between goals despite similar action patterns, *Infancy*, 15, 94-106.
- Moore, C., Jarrold, C., Russell, J., Lumb, A., Sapp, F., & MacCallum, F. 1995: Conflicting desire and the child's theory of mind. *Cognitive Development*, 10, 467-82.
- Onishi, K. H., & Baillargeon, R. 2005: Do 15-Month-Old Infants Understand False Beliefs? *Science*, 308, 255-258.
- Perner, J. 1991a: *Understanding the representational mind*. Cambridge, MA: MIT Press.
- Perner, J. 1991b: On representing that: The asymmetry between belief and desire in

- children's theory of mind. In D. Frye, & C. Moore (Eds.), *Children's theories of mind: Mental states and social understanding* (pp. 139-56). Hillsdale, NJ: Lawrence Erlbaum.
- Perner, J., & Ruffman, T. 2005: Infants' insight into the mind: How deep? *Science*, 308, 214–216.
- Perner, J., Zauner, P., & Sprung, M. 2005: What does “that” have to do with point of view? Conflicting desires and “want” in German. In J.W. Astington, & J. Baird (eds), *Why language matters for theory of mind* (pp. 220-44). New York, NY: Oxford University Press.
- Rakoczy, H., Warneken, F., & Tomasello, M. 2007: “This way!”, “No! That way!” - 3-year olds know that two people can have mutually incompatible desires. *Cognitive Development*, 22, 47-68.
- Ruffman, T. 2014: To belief or not to belief: Children's theory of mind. *Developmental Review*, <http://dx.doi.org/10.1016/j.dr.2014.04.001>.
- Sommerville, J.A., Woodward, A.L., & Needham, A. 2005: Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, 96, B1–B11.
- Song, H., Onishi, K. H., Baillargeon, R., & Fisher, C. 2008: Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition*, 109, 295– 315.
- Southgate, V., Senju, A., & Csibra, G. 2007: Action anticipation through attribution of false belief by two-year-olds. *Psychological Science*, 18, 587–592.
- Southgate, V., Chavalier, C., & Csibra, G. 2010: Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, 13, 907–912.
- Surian, L., Caldi, S., & Sperber, D. 2007: Attribution of beliefs by 13-month-old

- infants. *Psychological Science*, 18, 580–586.
- Surtees, A., Butterfill, S., Apperly, I.A., 2012: Direct and indirect measures of Level-2 perspective taking in children and adults. *British Journal of Developmental Psychology*, 30, 75-86.
- Russell, J. 1996: *Agency: Its role in mental development*. Hove: Erlbaum (UK) Taylor and Francis Ltd.
- Searle, J.R. 1983: *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.
- Shantz, C.U. 1987: Conflicts between children. *Child Development* 58, 283-305.
- Steglich-Petersen, A. 2006: No norm needed: On the aim of belief. *The Philosophical Quarterly*, 56, 499-516.
- Steglich-Petersen, A. 2009: Weighing the aim of belief. *Philosophical Studies*, 145, 395-405.
- Thoermer, C., Vuori, M., Perst, H. and Kristen, S. 2012: From an implicit to an explicit understanding of false belief from infancy to preschool age. *British Journal of Developmental Psychology*, 30, 172-187
- Traüble, B., Marinovic, V., & Pauen, S. 2010: Early theory of mind competencies—Do infants understand others’ beliefs? *Infancy*, 15, 434–444.
- Wellman, H.M., & Bartsch, K. 1988: Young children’s reasoning about beliefs. *Cognition*, 30, 239-77.
- Wellman, H.M., & Woolley, J.D. 1990: From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, 35, 245-75.
- Wellman, H.M., & Banerjee, M. 1991: Mind and emotion: Children’s understanding of the emotional consequences of beliefs and desires. *British Journal of Developmental Psychology*, 9, 119-24.

Wimmer, H., & Perner, J. 1983: Beliefs about beliefs: Representation and
constraining function of wrong beliefs in young children's understanding of
deception. *Cognition*, 3, 103-28.

Woodward, A. 2009: Infants' grasp of others' intentions. *Current Directions in
Psychological Research*, 18, 53-57.