



The Situational Mental File Account of the False Belief Tasks: A New Solution of the Paradox of False Belief Understanding

Albert Newen¹ · Julia Wolf¹

Published online: 19 March 2020

© The Author(s) 2020

Abstract

*How can we solve the paradox of false-belief understanding: if infants pass the implicit false belief task (FBT) by nonverbal behavioural responses why do they nonetheless typically fail the explicit FBT till they are 4 years old? Starting with the divide between situational and cognitive accounts of the development of false-belief understanding, we argue that we need to consider both situational and internal cognitive factors together and describe their interaction to adequately explain the development of children's Theory of Mind (ToM) ability. We then argue that a further challenge is raised for existing accounts by helping behaviour versions of the FBT. We argue that the common two-stage accounts are inadequate: we need to allow for *three central stages* in a continuous development. Furthermore, drawing on Perner et al.'s (*Cognition* 145: 77–88, 2015) and Perner and Leahy's (*Review of Philosophy and Psychology* 7 (2): 491–508, 2016) recent mental files account, we provide a new account of the development of these three stages of ToM ability by describing the changes of the structure and organisation of mental files including the systematic triggering role of types of situations. Thereby we aim to establish a *situational mental file (SMF) account* as a new and adequate solution to the paradox of false-belief understanding.*

Keywords False belief task · Paradox of false belief understanding · Mental files · Implicit false belief task · Helping behaviour

1 The Paradox of the False Belief Task

In the last decades the False Belief Task (FBT) has occupied a central position in the investigation of social understanding. It has become a litmus test in research into

Albert Newen and Julia Wolf contributed equally to the manuscript.

✉ Albert Newen
albert.newen@rub.de

¹ Institut für Philosophie II, Ruhr-Universität Bochum, Universitätsstr. 150, 44801 Bochum, Germany

children's Theory-of-Mind (ToM) ability, i.e. the ability to attribute beliefs to others and predict their actions accordingly.¹ The *explicit FBT*, which requires responding to a direct question with a verbal response to the task, has traditionally been considered as a reliable indicator that children acquire an understanding of false belief at around 4 years of age: in the original 'unexpected location' version of the task used by Wimmer and Perner (1983) children see a doll, Maxi, place her chocolate into a blue cupboard. Maxi then leaves and while she is gone her mother moves the chocolate from the blue cupboard to a green box. Maxi returns and the child is asked "Where will Maxi look for the chocolate?" Children typically only pass this task at the age of 4 (Baron-Cohen et al. 1985; Wimmer and Perner 1983). To give the correct answer, namely that Maxi will search in the blue cupboard, presupposes that the child can distinguish her own knowledge (i.e. that the chocolate is in the green box) from Maxi's false belief. This change in children's performance at age 4 was replicated across a variety of paradigms including the 'unexpected identity' test (Moses and Flavell 1990; Perner et al. 1987; Wellman 1990) or the 'unexpected contents' (Gopnik and Astington 1988) test. Thus, many researchers concluded that false belief understanding does not emerge until 4 years of age (Flavell 2004; Sodian 2005; see Wellman 2002 and Wellman et al. 2001 for a review). This picture was challenged when the *implicit FBT task*, requiring only non-linguistic responses, entered the scene.² The implicit FBT, first carried out by Clements and Perner (1994), became established with the influential study from Onishi and Baillargeon (2005) which provided evidence that 15 month old infants already had some sensitivity for other people's (false) beliefs. Since then three main types of implicit FBT have been employed (which we illustrate using the example of the Maxi unexpected location FBT): 1. violation of expectation FBT (where does Maxi look longer?), 2. anticipatory looking FBT (where does Maxi look first?), 3. helping behaviour FBT (how does the child help Maxi, does it account for his false belief?).

Concerns have recently been raised about the replicability of implicit FBTs (Dörrenberg et al. 2018; Kammermeier and Paulus 2018; Kulke and Rakoczy 2018; Kulke et al. 2017, 2018). Nonetheless, there are also responses to these challenges (Baillargeon et al. 2018; Roby and Scott 2018; Rubio-Fernández 2018). In addition, there is also evidence of implicit false belief understanding in animals using the anticipatory looking (Kano et al. 2017) as well as the helping behaviour paradigm (Buttelmann et al. 2017). This indicates that despite the difficulty in replicating some of the exact results with infants it seems that we do not entirely lose the phenomenon as we also find it in non-verbal animals. More research to clarify this issue is needed, but for the purposes of this paper we will take for granted that while some versions of the

¹ The term 'Theory of Mind' is also sometimes used in a wider sense to denote the ability to attribute mental states to oneself and others and not merely beliefs. In this paper, however, we will exclusively be concerned with beliefs. Furthermore, attribution should not be understood too cognitively as being tied to language. Furthermore we do not distinguish between belief-like states and beliefs for the core part of the article. At the end we outline that beliefs are only implemented when a certain structure of mental files is realised (or can be realised by the cognitive system), namely bi-directional linking. But it is not the function of the paper to work out the borderline between beliefs and belief-like states.

² There have been different characterisations of the difference between implicit and explicit FBTs. Following De Bruin and Newen (2012) we have characterised the difference in terms of the child's response. A different possible way of drawing the distinction is in terms of whether there is a direct test question. Although these differ, both characterisations draw the distinction in the same place for most of the existing FBTs in the literature.

FBT might be considered ‘fragile paradigms’ (Rubio-Fernández 2018) the key phenomena can nonetheless be assumed. Given that we take seriously the evidence that children pass different versions of the implicit FBT much earlier than 4 years of age (overview see De Bruin and Newen 2014, 301), we are faced with what might be considered the developmental ‘paradox’ of false belief understanding (De Bruin and Newen 2012): if infants pass the implicit FBT with nonverbal behavioural responses like looking or helping behaviour why do they nonetheless typically fail to give the correct verbal response in the explicit FBT till they are 4 years old?

The challenge, therefore, is how a proposed theory can describe the complex developmental changes from the ability to pass different versions of implicit FBT (starting with early violation of expectation FBT at 15 months, Onishi and Baillargeon 2005) to the ability to pass the explicit FBT (4 years, Wellman et al. 2001). In this paper we will begin by providing an overview of the previously proposed solutions and emphasise the need for an account which considers both the role of internal cognitive development as well as situational influences and their interaction. We then highlight the demand to integrate the helping behaviour FBT (Buttelmann et al. 2009, 2015; 18 months) as a distinct stage in an account of these developmental changes. Having outlined these desiderata we make use of mental files as a useful tool for thinking about cognitive development (Perner et al. 2015) and suggest the *Situational Mental File (SMF) account* as a new and adequate answer to these challenges. In doing so we leave aside the closely related questions of when concept of belief emerges and whether it is the same concept of belief which underlies implicit and explicit false belief understanding (Apperly and Butterfill 2009; Rakoczy 2012). We will only touch upon these issues in the conclusion of the paper.

1.1 Cognitive and Situational Accounts of False Belief Understanding

How can we solve the paradox? In the literature we find two prominent opposing views. On the one hand, ‘nativists’ argue for an early ToM-ability based on an inborn module (or module-like structure) which allows infants to pass the implicit FBT (e.g. Baron-Cohen 1995; Baillargeon et al. 2010³; Carruthers 2013, 2016; Helming et al. 2016; Leslie et al. 2004; Westra 2016). Broadly, this is the view that infants have an early understanding of other people’s beliefs which allows them to pass the implicit FBT. The challenge for these views, therefore, lies in explaining why children nonetheless fail the explicit FBT. On the other hand, ‘empiricists’ argue that ToM-ability is not based on an inborn module but is based on later developing abilities and it is this development which is responsible for the shift in performance on the explicit FBT at four years old (Apperly and Butterfill 2009⁴; De Bruin and Newen 2012; Gopnik 1993; Gopnik and Wellman 1992, 2012; Heyes 2014, 2018; Perner 1991; Perner and

³ Strictly speaking, Baillargeon does not argue that ToM is innate, but rather holds a two systems account which is, however, fully developed already by the time infants pass the implicit FBT. Therefore, as Baillargeon does not assume a change in false belief understanding underlying the development we have classed her amongst the nativists.

⁴ Apperly and Butterfill’s two systems account might also be thought to hold an intermediate position between empiricism and nativism and laid out here. We have nonetheless classed them as empiricists as they too argue for a cognitive development in terms of children’s understanding of belief itself underlying the shift in performance at 4 years of age.

Ruffman 2005; Perner et al. 2015; Perner and Leahy 2016; Wellman 2014). The challenge for these views then lies in explaining why much younger infants are nonetheless able to pass the implicit FBT.

There is a further important distinction, however, which cuts across the debate between ‘nativists’ and ‘empiricists’, namely between situational and cognitive factors. That is to say between those who explain the development from success in the implicit FBT to success in the explicit FBT primarily by an *intense development in the cognitive organisation* (cognitive accounts) and those who argue that this difference in performance is primarily a product of situational factors pertaining to the various tasks (e.g. whether the child’s focus on the other person is maintained or interrupted (Rubio-Fernández and Geurts 2013) or how many seekers there are (Lewis et al. 2012) - situational accounts). These situational accounts presuppose minimal cognitive conditions of adequate working memory and/or linguistic understanding as a background condition. This makes clear that the dialectic here is different to that in the debate between nativists and empiricists in that cognitive and situational accounts are not mutually exclusive. Rather they differ in their *focus*, focusing more on situational or internal cognitive factors respectively. There are both empiricist and nativist variants of cognitive and situational accounts. For an overview of these distinctions in the literature, see Table 1.

The traditional view amongst the nativists is that infants have an innate ToM module which enables early belief understanding, and the reason why children do not pass the explicit FBT till the age of 4 lies in the executive function and working memory demands the task places on the child (Baron-Cohen 1995; Baillargeon et al. 2010; Carruthers 2013, 2016; Leslie 1987; Leslie et al. 2004). That is to say they argue that it is a development in general cognitive skills which underlies the developmental shift in performance at 4 years of age. Recently, however, there has been a move towards integrating the role of situational factors within a nativist framework in the form of ‘pragmatic accounts’. These accounts form a subset of situational accounts which restrict the influence of situational factor to the understanding of the explicit question. In more detail, this is the view that although children have an early false belief understanding, there are additional external contextual factors in the specific setup of the explicit FBT which lead children to systematically misinterpret the explicit question posed and thereby prevents early success in the explicit FBT.

We find a similar distinction within the empiricist camp. As we saw above, the defining feature of the empiricists is the view that there is a development in children’s understanding of beliefs, usually in terms of a developing concept of belief (Apperly and Butterfill 2009; Perner and Ruffman 2005; Gopnik and Wellman 1992). Most accounts are purely cached in terms of an internal cognitive development or reorganization which leads to a full-blown representation of beliefs (e.g. Perner 1991; Perner et al. 2015; Perner and Leahy 2016; Apperly and Butterfill 2009; Perner and Ruffman 2005; De Bruin and Newen 2012). There are, however, also accounts such as Heyes’ (2014, 2018) which consider the role of situational factors in development. Similarly, Gopnik is committed to the idea that children’s ToM develops in the form of theory revision as a consequence of the evidence encountered. Therefore, it can be argued that there is also consideration of situational components in Gopnik’s account (Gopnik 1993, see also Gopnik and Wellman 1992) given the role of experience on cognitive development in the form of theory revision.

Table 1 Systematic overview of theoretical accounts of the development of false belief understanding

	Cognitivist account	Situational account	Cog/Sit combined account
Nativist	Theory of mind module	Baron-Cohen (1995)	Helming et al. (2016)
	Response account	Baillargeon et al. (2010)	Westra (2016)
	Triple mindreading	Carruthers (2013, 2016)	
	Theory of mind mechanism (ToMM)	Leslie et al. (2004)	
Neutral	Differential task demands view	Rubio-Fernandez (2013)	Situational mental files account
Empiricist	Two systems account	Apperly and Butterfill (2009), Butterfill and Apperly (2013)	Child scientist - theory revision
	Behaviour rules and meta-representation	Perner et al. (1987), Perner (1991)	Newen and Wolf (this paper)
	Mental files	Perner et al. (2015), Perner and Leahy (2016)	Gopnik (1993), Gopnik and Wellman (1992, 2012), Wellman (2014)
	Dual systems association account	De Bruin and Newen (2012)	

For the purposes of this paper, we want to remain neutral concerning the nativist/empiricist debate.⁵ Although we will be making use of the mental files framework from Perner et al. (2015) and Perner and Leahy (2016) who do not advocate a nativist position, it is not clear whether the development of the ability to link mental files has to be a domain specific development in terms of a specifically developing understanding of belief, or whether this is also something which can be cashed out in terms of the development of domain general processes (e.g. working memory, executive function, or even a ‘decoupling mechanism’ as advocated by Leslie 1987). Instead, we want to focus on the distinction between cognitive and situational factors which, as we have argued, runs orthogonal to the nativist/empiricist divide. *Concerning this debate will argue that both factors – cognitive and situational - play a crucial role, i.e. we need to account for a change in the internal cognitive organisation as well as of the triggering and supporting role of specific situational features and their interaction to adequately describe the development of the false-belief understanding.* Only accounting cognitive and situational aspects might appear trivial. If pressed most accounts detailed above would accept a claim like this. Some even do so explicitly (e.g. Helming et al. 2016; Westra 2016). These accounts, however, still have a *focus* on one factor and fail to adequately integrate the dominant factor with the other factor. The critical element of our account, therefore, is that we not only argue that both cognitive and situational factors play a role in development, but that we illustrate which aspects are relevant and clarify how they interrelate. It is this which has not been done in the literature. In this regard, our account may be thought to be similar to that of Gopnik, who also considers a combination of cognitive and situational factors. As we will make clear in the next section, however, we will be making use of the mental files framework in order to put forward a more detailed alternative to Gopnik’s account of theory revision, which is also not subject to the strong Theory Theory commitments.

1.2 The Relevance of Situational Factors; Limits and the New Direction

There has been increased interest in the role of situational factors on development over the past years, in particular in the form of ‘pragmatic accounts’ (Helming et al. 2016; Westra 2016; Westra and Carruthers 2017). This is partly due to the open challenge to provide an explanation of the paradox of false belief that does justice to the competence displayed by children across a variety of implicit FBTs. Westra (2016), for example, has recently put forward the idea that children’s failure in the explicit FBT is due to misunderstanding the question they are being asked (see also Westra and Carruthers 2017): As belief talk is infrequent, children systematically misinterpret the question as being about where the chocolate actually is as opposed to the other’s belief. Providing support for these types of accounts, there has been increasing evidence that situational factors such as the way in which the question is posed to children in the explicit FBT have an influence on children’s performance (Hansen 2010; Rubio-Fernández and Geurts 2013). We will be considering this evidence in more detail below, see section 4.1). The idea that children may fail the FBT because they are systematically confused by the question they are being asked is not a new one (see for example Siegal and

⁵ This is one of the reasons why we consider the role of situational factors more broadly, rather than limiting ourselves to the pragmatic effect of situational factors changing children’s interpretation of the question.

Beattie 1991). But, while we agree that situational factors do play a role in the change in children's ToM performance, we do not think that this can give a full account of their development. Firstly, there are by now many variations of the explicit FBT in which the questions posed range widely from questions concerning behaviour (e.g. "where will Maxi look for her chocolate") to questions explicitly about the other person's beliefs (e.g. "where does Maxi believe the chocolate is?"). This, however, has only a limited effect on children's performance (Psouni et al. 2018; Wellman et al. 2001). Secondly, in those studies where changing the question posed did have an effect (Hansen 2010; Rubio-Fernández and Geurts 2013), this effect was primarily achieved in 3 years olds, that is to say in children who were on the cusp of being able to pass the explicit FBT already. This seems to indicate that while situational factors such as the type of question being posed can have an effect, this is limited to children who are quite close to passing the explicit FBT already. In other words, performance can be somewhat improved due to situational factors, but this cannot provide a full explanation of the paradox of the FBT. Lastly, 3 year old children who still fail normal versions of the FBT are already pretty well linguistically developed and use belief terms in conversation (Bartsch and Wellman 1995). This makes it implausible to argue that children fail the explicit FBT primarily because they systematically misinterpret the question they are being asked. While this plays a role, further cognitive factors are needed to provide a full account of the development.

The role of situational factors can extend beyond influencing the child's understanding of the question posed. For example, Helming et al. (2016) argue that young children struggle in dealing with different perspectives. While they are able to compute the other person's belief, when the experimenter asks the test question the child is forced to adopt the perspective of the experimenter in order to understand the question and answers this based on the shared knowledge with the experimenter as opposed to the false belief of the other agent. The explicit FBT therefore poses an extra perspective problem in virtue of the set-up which, jointly with the cooperation bias, explains children's systematic errors. On this account, however, we need some story of how children are able to overcome this problem, and this seems to require some reference to cognitive development which allows children to overcome the perspective problem they face. Therefore, this account too depends on both situational and cognitive factors. While Helming et al. (2016) explicitly do allow for the role of a cognitive factor in terms of a processing load account, they do not provide a detailed account of how the situational factors and cognitive development interrelate.

In light of this, while we think that situational factors play an important role in explaining the paradox of the FBT, they do not provide a full explanation of the paradox. Instead the role of situational factors has to be seen in conjunction with a systematic cognitive development. We agree that, given a certain level of cognitive development, situational factors can improve performance. Nonetheless, in order to fully explain the paradox of the FBT, we require a specific internal cognitive reorganisation which we will describe using mental files. Thus, we are arguing for an account in terms of both situational and cognitive factors and their interaction. This, as a first step, might not be so controversial. As noted above, Helming et al. (2016) acknowledge that there is likely to be some cognitive development in children's executive function which allows children to overcome the additional situational demands of the explicit FBT. Similarly, our claim is not that the accounts which we have

described as ‘pure’ cognitivist are incompatible with the pragmatist picture. Rather, what we want to stress is that these accounts, while compatible with each other have usually focused only on one component to explain the development of false belief understanding (with the exception of Gopnik and Wellman) and this seems to be a systematic deficit. We argue that cognitive and situational factors interrelate in development and therefore need to be considered in conjunction in order to provide an account of development. This is similar to an idea found in Gopnik and Wellman’s (1992, 2012) work. They account for the role of situational factors in development by considering the role of the child’s systematic learning experiences in shaping theory development. We, however, do not want to take on such a strong commitment to theory formation in young children. Therefore our aim in this paper is to work out a mental file account as a less demanding and cognitively more convincing alternative to the strong Theory-Theory proposal defended by Gopnik. Thus, we aim to demonstrate that making use of the mental files framework provides us with a more detailed and adequate way of spelling out the development of the performances. Furthermore, we want to be much more specific in our account of how cognitive and situational factors interact in development. We will outline how the cognitive organisation develops using the framework of mental files and how this cognitive organisation is triggered by situational features for both its development and its application. These comprise the core aspects of the new SMF account.

To unfold the new account we proceed in two steps: (i) we argue that we need a three-stage account to adequately describe ToM development (section 2) and (ii) we introduce the mental file account (section 3) and use it to describe this three stage development (section 4 and 5).

Given these lines of debate, our contribution can be summarized as follows: firstly, we aim to show that the rich data of the FBTs can only be explained: 1. by accepting a crucial role of both internal cognitive reorganisation of tools to deal with the FBT (cognitive factors) and situational contextual features (situational factors) triggering this development. 2. by explaining the cognitive development in terms of a three-stage theory. Secondly, we argue that these requirements can be met by describing the changes of the cognitive structure and organisation in terms of mental files and taking into account the role of situational factors in triggering the activation of mental files. The result we are aiming for is a new SMF framework which is supposed to provide a new explanation of the paradox of the FBT which accounts for the main empirical findings from developmental studies.

2 The Problem of Helping Behaviour

While the distinction between the performance on implicit and explicit FBT is undoubtedly an important one, there is a danger of it overshadowing further developmental stages which lie between implicit and explicit false belief understanding. In the literature which explicitly tries to account for this development, we basically find only suggestions of different versions of two-step developments in the cognitive organisation. These two steps have three different characterisations: 1. A development of a unitary system which has two subsystems which unfold one after the other (Baillargeon et al. 2010), 2. a development from associative representations which enable behaviour

reading to linguistic representations of the same situation which enable an explicit false-belief understanding (Perner and Ruffman 2005; Ruffman 2014) and 3. the two systems account of Apperly and Butterfill (2009). Since the latter is the most widely accepted two-step account we will focus on this view. This account relies on the two-systems distinction as described by Kahneman (2003, 2011). According to Apperly and Butterfill's (2009) popular two systems account, there is one early developing, quick, "cognitively efficient but limited and inflexible system" (system 1) (Apperly and Butterfill 2009, 966) which allows children to pass the implicit FBT. A second, slower "highly flexible but cognitively inefficient" mindreading system (system 2) (Apperly and Butterfill 2009, 966) is needed in order to pass the explicit FBT. This system only develops later – specifically around the age of 4 – and it is due to this that children do not pass the explicit FBT earlier while system 1 is sufficient for success on the implicit FBT.

As we will explain below, however, a problem for this view is posed by the findings from the *active helping behaviour paradigm*, as this seems to be an intermediary stage between success in the implicit FBT and success in the explicit FBT (Buttelmann et al. 2009; Buttelmann et al. 2015). To be clear, our claim is not that it is only the active helping paradigm which poses a problem for the view that there is a strict two stage dichotomy between early implicit and later explicit false belief understanding. Rather, the active helping behaviour paradigm is a particularly good example of the kind of active, "action-based usage of theory of mind skills" (Knudsen and Liszowski 2012, 673) which emerges at around 18 months. It is characterised by an increased ability to contrast another's perspective with one's own. This, we suggest should be distinguished from early implicit false belief understanding and later explicit false belief understanding. While our main focus will be on the active helping behaviour paradigm, we think the same argument can be made using other forms of active helping such as Southgate et al. (2010)⁶ or the communicative pointing in order to warn the experimenter seen in the studies by Knudsen and Liszowski (2012).⁷

The active helping behaviour paradigm makes use of the finding that children show helping behaviour from 14 months (Warneken and Tomasello 2007), for example retrieving a toy for the experimenter which is out of his reach. In the original study by Buttelmann et al. (2009), one experimenter places his toy into a pink box. Like in the classical false belief studies the experimenter then leaves and while he is gone the toy is moved from the pink box to the yellow one. Rather than asking where he will look for her toy the experimenter returns and tries to open the pink box. The child is then told to 'help' the experimenter. Just based on this encouragement, 18 month olds tend to help the experimenter by opening the yellow box to retrieve the toy.⁸ In the control condition, however, where the experimenter was present when the location of

⁶ See also Carpenter et al. (2002) for similar evidence from 3 year olds.

⁷ A further paradigm in which a similar ability to reconcile perspectives is that of Moll et al. (2017), who showed that 2.5 year old children express suspense when watching another person approach reality with false beliefs. This paradigm is especially interesting because it suggests that this intermediate stage is not only possible in the context of action. It should be noted, however, that these children were older than those in the previously mentioned action based studies, so it may still be that initially progress to this stage is achieved through action.

⁸ Children's responses were also counted if they merely 'chose' the box by touching it. This reduces the worry that children might fail the helping behaviour task because the mechanism of opening the box is too demanding for them.

the toy was switched, if the experimenter tried to open the pink (now empty) box then the child helped the experimenter by opening the pink box. The key idea of this experiment is that it requires the child to determine the experimenter's goal *based on the beliefs the experimenter has*. It is because the experimenter has a 'false belief' that the toy is in the pink box that his action is interpreted as an attempt to retrieve the toy. On the other hand, in the control condition where the experimenter knows that the toy has been moved, the very same action is instead seen just as an attempt to open the pink box. The control condition is crucial for the experiment: without it all that is shown is that the children retrieve the toy. They might do so without any regard to the experimenter's goals merely because the toy is the most interesting object in the scene. What is notable is that they do not retrieve the toy in the control condition. It is this difference between the children's interpretation of the experimenter's behaviour that lends support to the interpretation children with 18 months are helping by accounting for the belief (true or false) of the experimenter. Buttelmann et al. (2009) also carried out the same task with 16 month olds but found that evidence of false belief understanding amongst this age group was inconclusive.

Although the helping behaviour paradigm is often considered to be a further variation of an implicit FBT, the task actually differs from the above mentioned looking behaviour based implicit FBTs in a number of important ways. Firstly, the helping behaviour task is an elicited response task, unlike the looking behaviour based task which is a spontaneous response task (Priewasser et al. 2018). Secondly, the helping behaviour paradigm requires the child to access two perspectives simultaneously. This involves being able to coordinate both their own perspective of the situation with that of the other agent: the child first needs to consider the agent's perspective on the situation in order to realise that they want to retrieve their toy. In addition, they then need to rely on their own, differing perspective on the situation in order to determine where the toy actually is so that they can help by retrieving it. This is different to the looking behaviour implicit FBTs which children could pass based only on the perspective of the other person. All that is needed to pass these tasks is the understanding that Maxi expects the chocolate to be in the blue box. Relating this information to where the chocolate actually is is not required in this task (Tomasello 2018). In this sense, the helping behaviour paradigm seems to require skills which are more similar to those which are required in the explicit FBT, namely activating perspectival information and relating this to the actual scenario, and making use of this information to guide an action. Nonetheless, children are able to pass the helping behaviour task much earlier (onset 18 months, Buttelmann et al. 2009) than the explicit FBT (onset 4 years, Wellman et al. 2001). It is also interesting that performance in an active helping behaviour paradigm does not correlate with performance on a looking behaviour based task, while it does correlate with explicit FBT performance (Grosse Wiesmann et al. 2016, supplementary materials).⁹

The helping behaviour paradigm therefore raises the following problem for any two systems account: on the one hand, if system 1 is used to explain both implicit looking behaviour and the helping behaviour while system 2 explains performance in the explicit FBT, then we lack an explanation for the difference in performance between

⁹ It must be acknowledged, however, that this study failed to replicate the early success in the helping behaviour study: they found successful performance only in 4 year olds.

looking behaviour (as early as 7 months Kovács et al. 2010) and helping behaviour (observed from 18 months onwards). If, on the other hand, system 1 deals only with implicit looking behaviour and system 2 deals with helping behaviour then we lack an explanation for the difference between helping behaviour (onset 18 months) and the explicit FBT (onset 4 years). Those in favour of collapsing the former distinction may argue that there is no strong evidence that children cannot pass the helping behaviour task before the age of 18 months. Indeed, Buttelmann et al. (2009) considered the evidence from the 16 month old infants to be inconclusive in contrast to the results of the 18 months olds. Furthermore, younger children may be prevented from passing the helping behaviour task due to other task specific factors as, for example, their general helping behaviour and motor abilities may still be limited. However, the actual behaviour required of the child in Buttelmann's paradigm was quite limited as even indicating the correct box was considered to be a success, and there is some evidence to suggest that helping behaviour has developed to a sufficient degree in children of this age (Warneken and Tomasello 2007). It is, of course, possible that new evidence may show that younger children are capable of passing the helping behaviour task. Up to now, however, there does not appear to be any strong evidence that they can and problems in replicating the Buttelmann findings – which have so far only been replicated in 3 year olds (Priewasser et al. 2018)¹⁰ – provide some reason to be cautious of thinking that such evidence may be found.

In the light of this, we therefore think that the most prudent solution given the current evidence is to posit that false belief understanding should be seen in terms of a continuous development, in which three important stages can be distinguished:

1. Early sensitivity to belief (resp. belief-like states) (e.g. as shown in the looking behaviour based tasks)
2. More sophisticated usage of early sensitivity to belief (resp. belief-like states) (e.g. as required for the helping behaviour task)
3. Explicit ToM involving belief attribution

While we keep the distinction between implicit and explicit false belief tasks, we argue that we should distinguish between two different kinds of implicit ToM.¹¹ We will further cash out this distinction in terms of different kinds of linking between mental files in section 4.

While we think that this evidence from the helping behaviour paradigm is highly significant and allows us to draw important conclusions concerning the development of children's false belief understanding, it must nonetheless be acknowledged that, as with most of the implicit FBTs, the interpretation (as well as the replicability, see section 1)

¹⁰ In a direct replication study by an independent lab. Buttelmann et al.'s (2014, 2015) studies which investigated false belief about identity might be considered a conceptual replication, also finding success in 18 month olds, but did not test younger children.

¹¹ A similar distinction between two different kinds of implicit false belief understanding is also drawn by Vierkant (2012, 142), who points out the difference between a child that "has transformed his behavioural repertoire" and one that "has the right looking patterns [but] will at most display very moderate differences in her behaviour". Neither child has explicit false belief understanding as, if asked to verbally report where Maxi will look for the chocolate they still give the wrong answer, but the kind of implicit false belief understanding displayed by the children in the looking behaviour and the helping behaviour tasks is very different.

has been highly debated. For example, it might be objected that this task is crucially different from other FBTs because children do not have to resist the ‘pull of the real’. That is to say that in order to retrieve the toy the child can make use of where the toy actually is. In this case, however, children must resist the pull of the real in the control condition where they do not go to retrieve the toy. That this poses a genuine problem for the children can be seen given that the true belief condition of Buttelmann et al’s paradigm has proven especially difficult to replicate (Kulke and Rakoczy 2018; Prieuwater et al. 2018). Nonetheless, Buttelmann et al’s findings seem to suggest that children can resist this pull of the real under some conditions. Secondly, while the child does not need to resist the pull of the real with regard to the action, they do need to do so in order to determine the experimenter’s goal: the experimenter is looking in the empty box, not the one where the toy is, nonetheless children need to interpret this as searching for the toy in order to succeed in the task.

A further worry is that this task seems to require children only to attribute ignorance to the experimenter as opposed to false belief itself: he searches in the wrong box because he *does not know* where the toy is. While this may be a worry in Buttelmann et al. (2009), we do not think the same worry applies in Buttelmann et al.’s (2015) later version of the helping behaviour paradigm which makes use of the appearance-reality paradigm (Flavell et al. 1983; Gopnik and Astington 1988). That children retrieve a duck when the experimenter asks for an ambiguous object which the child, but not the experimenter, knows looks like a duck but is actually a sponge cannot be explained in terms of a rule such as ‘if ignorant, then gets it wrong’.

Finally, Prieuwater et al. (2018) have recently conducted a study in favour of a new ‘teleological’ interpretation of the data from the helping behaviour study according to which children attribute goals to the experimenter based on objective facts of the experimental set-up rather than the experimenter’s beliefs. They argue that the child forms the expectation that the experimenter will return for his toy given his previous interest in the false belief condition, while in the control condition it seems that the experimenter has lost interest in his toy. In order to distinguish between these two interpretations they carried out a helping behaviour paradigm with three boxes instead of two and a new false belief condition where the experimenter returned to open the third box in which the toy had neither been hidden, nor where it was located now. According to the teleological but not the ToM interpretation by Buttelmann et al. children ought nonetheless to interpret this as an attempt to retrieve the toy. Prieuwater et al. indeed found that this was what happened, which they concluded provided strong evidence in favour of their teleological interpretation over the alternative. Their findings, however, are considerably weakened by the fact that the majority of children retrieved the toy in all conditions, in particular they also went to the box with the toy in the true belief condition. Rather than providing convincing evidence in favour of their interpretation, it seems that what happened was that the pull of the toy was too strong on the children which led to them retrieving the toy regardless of the condition. One possible way of avoiding this problem would be to have a second independent toy in each box such that the objective appeal of opening each box is equal. In order to provide compelling evidence, therefore, it would be necessary to show not only that children retrieve the toy in the new false belief condition, but also that there are conditions, i.e. the true belief condition, in which they do not retrieve the toy. A further worry concerning this experiment is that the addition of a third box might have placed

too strong demands on children's executive processes, thus leading to a default action of retrieving the toy. For further discussion of these findings see Baillargeon et al. (2018). Our aim with this discussion is not to dismiss the teleological account, which we think may actually be compatible with our account, but rather to clarify that the findings of the Prieuwater et al. (2018) study do not undermine our interpretation of the helping behaviour.

In the remainder of this paper, we will begin to put forward an account of the development of false belief understanding based on these three stages and taking into account the role of both cognitive development and situational factors and their intertwinement. In order to do so we will make use of the mental files framework, which has recently been put forward by Recanati (2012) and further developed by Perner and Brandl (Perner and Brandl 2005, see also Perner et al. 2015; Perner and Leahy 2016) as a useful framework for thinking about the development of cognition. In the next section we will first introduce the structure and organisation of mental files which provides the basis of our account of the cognitive development. We will then introduce two – situational and cognitive - principles of activating mental files. Finally, having developed these tools, we apply the principles to provide our new situational mental file account of the development from implicit to explicit false belief understanding in children.

3 Mental Files

There are two central components of the mental files account (Perner et al. 2015, 78): firstly, the mental files themselves, which are to be thought of as “tool[s] for managing information about an object in the world” and, secondly, the links between files which enable the information to be shared between them. We will introduce this account using an example. Suppose I see an unfamiliar silver pen lying on my desk. I create a mental file for this pen, which is anchored in the object and contains the information I have concerning the pen, namely that it is silver and on my desk (Fig. 1).

Later on, I talk to my colleague Claire and she tells me that she is missing her favourite pen. Here, too, I create a mental file for Claire's pen. As we talk more, Claire describes the pen further and I come to realise that Claire's pen is actually the unfamiliar pen lying on my desk. If I know that the pen on my desk and Claire's pen are actually one and the same, the two files need to be linked such that information between the two files can be shared (Fig. 2). This way, when I think of Claire's pen, I

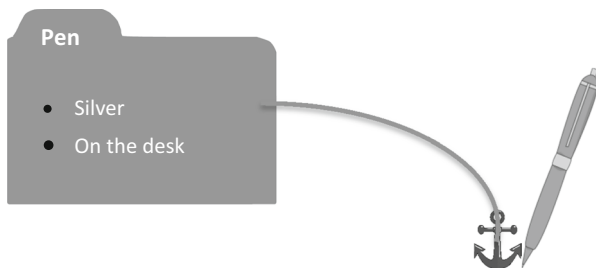


Fig. 1 The mental file is anchored in the object

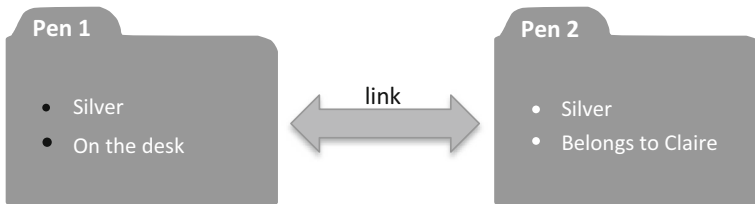


Fig. 2 Linking of co-referential mental files. In the following figures the anchoring relation is not shown for the sake of simplicity

am able to retrieve the information from the file of the silver pen on my desk and thus tell her where her pen is.

Perner and Leahy (2016), who have argued for the use of mental files as a tool for thinking about cognitive development, apply this idea to representing other people's beliefs. The key idea, they argue, is that we not only have 'regular files' which are representations of our own perspective of the world, but that we also have 'vicarious files' which are indexed to another person and instead represent their perspective on the world. So, for example, if I decide not to tell Claire that I have her pen and instead lead her to believe that the pen is in her pencil case, I would have a mental file indexed to Claire containing the information 'is in the pencil case' (Fig. 3).

Applying this to the FBT, this means that the child has two representations (or mental files) of the situation: one regular file with the information that the chocolate is in the green box, and one vicarious file indexed to Maxi with the information that the chocolate is in the blue cupboard. The problem, Perner and Leahy (2016) argue, is that children below the age of 4 are not yet able to link mental files (Fig. 4a). That is to say that the child is unable to switch between the vicarious and the regular mental file in a controlled and systematic way. It is only once the mental files are linked that the child is able to systematically access the information of Maxi's belief concerning the location of the chocolate (Fig. 4b). The reason why children fail the explicit FBT before the age of 4 is because, when confronted with the question where Maxi will search for the chocolate, they are only able to draw on the information from their own regular file (containing the information on where the chocolate actually is) and are not able to access the information from the vicarious file (containing the information where Maxi believes the chocolate to be).

How are mental files implemented? It is not our aim to work this out in detail but we can rely on the work of Kahneman et al. (1992) who describe in detail why we have to presuppose object files and how they may be structured from a cognitive psychological perspective.¹² The idea of systematic object representations is also well established in the developmental psychology describing early infancy with inborn or early acquired core cognitive abilities. Core cognition (Carey 2009; Kinzler and Spelke 2007) includes

¹² In the literature some authors distinguish between 'object files' and 'mental files' (Murez and Smortchkova 2014) such that the latter are based on long-term memory while the former are only based on working memory. If one considers this distinction then object files are inadequate to solve problem in philosophy of language with long-term reference understanding. For our purpose this distinction is not relevant. First of all, our work focuses mainly on representations which would be held in working memory due to the design of false-belief tasks. Second, we are open-minded to allow for a dynamics between files of objects, agents and situations which are either held in working memory or long-term memory. A full story of mental files, going beyond the scope of this article, would clearly integrate such dynamics.

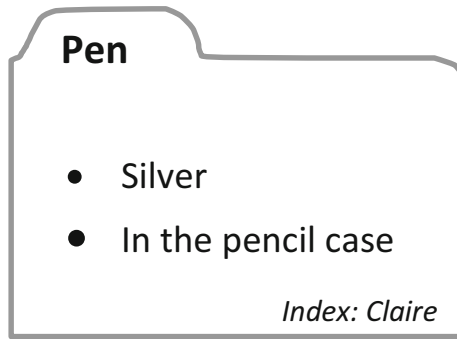


Fig. 3 Vicarious mental file of the pen, indexed to Claire

object representations as a central component. Furthermore, in philosophy mental files are used as a necessary presupposition to explain the mental organisation of the human mind (Perry 1990; Recanati 2012). Perner and Leahy (2016, 491) have also defended the use of mental files as a tool for thinking about cognitive development in psychology as they “capture important aspects of cognition”. Therefore, although mental files are not yet boiled down to neural correlates,¹³ there is a clear description of the functional role and the cognitive structure that they can justifiably be used as an explanatory tool.

In the previous section we outlined a number of challenges for any account of the development of children’s ToM abilities. In particular, we argued that the account must be able to explain not only the difference in performance between implicit and explicit FBTs, but also the distinctive developmental stage of the helping behaviour. In the next section we will address this issue and argue that if we consider both the role of cognitive and situational factors, the mental files account can provide a compelling explanation of the gradual development from implicit to explicit false belief understanding.

4 Mental Files and the Paradox of the False Belief Task

Expanding on Perner and Leahy (2016) we argue that the activation of mental files happens according to the following principles which we will elaborate below:

1. Situational factors principle – if a person has two mental files of one object, namely a regular and a vicarious mental file and *there is no linking relation between them*, then the activation of one of the two files is determined by situational factors alone: If situational factors trigger a focus on the object, the regular file is activated, if they trigger a focus on the person, the vicarious file is activated.
2. Cognitive factors principle – if a person has two mental files of one object, namely a regular and a vicarious file, and *there is a linking relation between the files*, then if one of the files is activated by situational factors, then

¹³ But see Arora et al. (2015) for a meta-analysis of patterns of neural activation and identity which suggests some possible areas of neural activity for linking mental files.

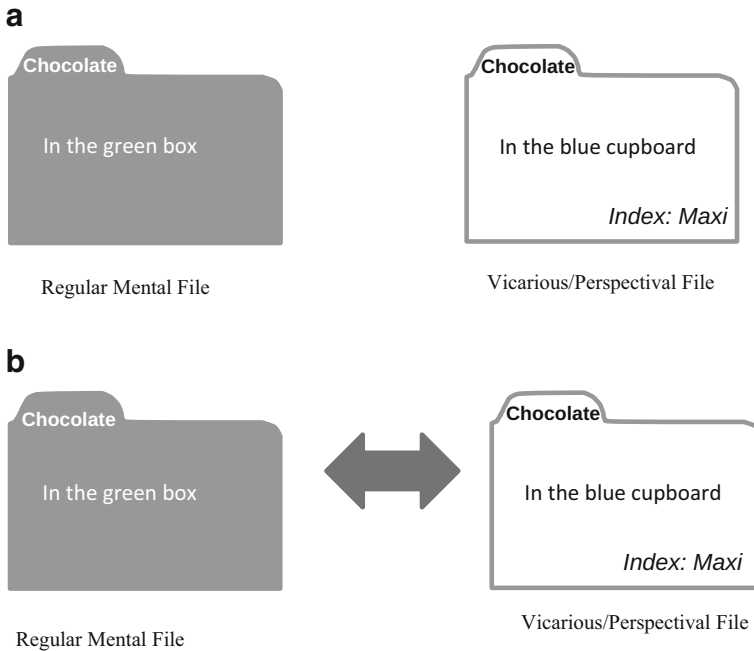


Fig. 4 a) Vicarious mental file and regular mental file are initially unlinked. b) Vicarious mental file and regular mental file become linked, enabling the flow of information between the files. Perner and Leahy (2016) argue that this linking underlies explicit false belief understanding

information from all the files accessible through the link (which may be directed) is available. The most adequate information from the accessible files is then *selected* according to the task.

We will first elaborate on these principles before showing how they can be used to model the three stages in the development of false belief understanding.

4.1 The Situational Factors Principle

There is considerable evidence that situational factors can have an effect on cognitive processes. For example, many priming studies have shown that attention can be strongly modulated by the context stimuli are presented in (Chun 2000). Similarly, contextual priming has also been shown to have an effect on language comprehension in the case of ambiguous words (Gennari et al. 2007). Putting this in terms of mental files, these findings can be re-described as the focus on the object due to the context (i.e. situational factors) facilitating the activation of the mental file of that object. Applying this principle to the FBT, if situational factors determine which mental file is initially activated, then factors which highlight the other person's perspective should improve performance on the FBT. There is some evidence that this is the case in the explicit FBT. In one study, Lewis et al. (2012) showed that children's understanding another person's false beliefs actually improves if another person is added to the scenario who also observes the change in location. That is to say that they were more accurate in determining the beliefs of the other person in a scenario in which there were

two seekers looking for a hidden object as opposed to one. This, at first sight, might seem puzzling, as we would expect that adding another person to the scenario makes the task more cognitively demanding. We want to suggest, however, that the presence of the other person highlights the different perspectives on the situation and thereby increases the chances of the vicarious file as opposed to the regular file being activated.¹⁴ This, in turn, means that more children are able to correctly report the other person's false belief. Similarly, Hansen (2010) conducted a study in which they showed that asking "you and I both know where the chocolate is, but where will Maxi look for the chocolate?" improved children's performance in the FBT. While this question may have been overly suggestive, this evidence is in line with our suggestion that children's ability to access their information on the other person's beliefs is facilitated if the perspective of the other person is emphasised. A different way of highlighting perspective is in the form of a memory aid, as done by Mitchell and Lacohée (1991) who also found that this improved performance.

Additionally, in a highly influential experiment, Rubio-Fernández and Geurts (2013) provided evidence that children before the age of 3 can pass even the explicit FBT if the task is modified such that, firstly, the child is frequently reminded of the other person's perspective throughout the experiment and, secondly, the question the child is asked is changed to "what happens next?". This modified version of the FBT was referred to as the Duplo Task. Rubio-Fernandez and Geurts argue persuasively that asking the question "where will Maxi look for her chocolate?" introduces a focus on the object and hence disrupts the children's tracking of the other person's perspective. Putting this in terms of mental files, because the question mentions the objects there is a focus on the object which leads to the activation of the regular file instead of the vicarious file. Children are therefore only able to access the information where the chocolate actually is as opposed to where *Maxi* believes it to be.¹⁵

So far we have considered cases where the perspective was highlighted. However, it is also possible to improve performance by reducing the salience of reality, for example if the children themselves do not know the real location of the object or the object is removed from the scene (Mascaro and Morin 2015; Mascaro et al. 2017; Wellman and Bartsch 1988). Reducing the salience of reality – and hence reducing the propensity towards activating the regular file – facilitates the activation of the vicarious mental file.

While the evidence so far considered is based on the explicit FBT, this can be used to explain children's early success in the implicit FBT: in the implicit FBT the return of the agent places the focus on the other person. This leads to the activation of the vicarious file, allowing children to pass the task. In the explicit FBT, however, the question "where will Maxi look for her chocolate?" places the emphasis on the object (Rubio-Fernández and Geurts 2013) and hence activates the regular file at the cost of the vicarious file.¹⁶ If the files are unlinked, as is the case in children before the age of

¹⁴ To be clear, our claim is not that this effect is something which is found every time at every occasion. Our claim, rather, is that there within a group there will be an overall improvement due to situational factors.

¹⁵ In a recent study attempting to replicate the Rubio-Fernández and Geurts (2013) study, Kammermeier and Paulus (2018) found that children were better at the Duplo Task than the standard FBT for ages 3 and 4, but that only the 4 year olds were above chance at the Duplo Task. This is in line with our view that situational factors are important in improving performance, but that cognitive development is necessary for the development of false belief understanding too and seems to constrain the effects of situational factors.

¹⁶ The idea is that activation of mental files is competitive, at least for co-referential but unlinked mental files.

4, this means that the information in the vicarious file is not accessible and therefore children fail the FBT. One point to note here, however, is that there is no systematic link. The problem lies not in representing another person's perspective, but rather in being able to access this in a systematic, situation independent way.

4.2 Cognitive Factors Principle

We base our cognitive factors principle off the work of Perner et al. (2015) and Perner and Leahy (2016) who argues that children develop the ability to link mental files at the age of 4. The main motivation for this principle is that this provides the best explanation of the correlation between performance on the FBT and the seemingly unrelated alternative naming task: both of these tasks depend on the ability to link mental files. That is to say that both require the ability to switch between different perspectives (or ways of considering) one and the same object (Perner et al. 2002). Following Perner and Leahy (2016) we suggest that the development of the ability to fully link mental files allows 4 years olds to pass the explicit FBT (i.e. in the absence of situational facilitation). However, we want to suggest that this linking relation can be *unidirectional* (i.e. information from file A is accessible from file B, but information from file B is not accessible from file A) as well as reciprocal (i.e. information from file A is accessible from file B and vice versa). There is some intuitive plausibility to this idea that the link between vicarious mental file and regular mental file is not fully reciprocal, even in the case of adults in which mental files are linked. When thinking about someone else's perspective, one's own perspective naturally suggests itself (especially if the other person's perspective differs from mine). Remembering to take into account someone else's perspective when thinking about my own perspective requires considerably more effort (Bradford et al. 2015). Moreover, postulating such a unidirectional link allows us to account for the helping behaviour which, as we argued earlier, should be seen as a distinct developmental stage.

A further point to note is that linking mental files goes hand in hand with general executive function abilities: linking mental files means that more information is available to the child which needs to be selected and inhibited.

4.3 Interaction between Situational and Cognitive Factors

We have suggested that both situational and cognitive factors play a role in the development of false belief understanding. We want to go further than this and argue that there is a substantial co-dependence between the two factors. A central claim of our view is that situational and cognitive factors do not just impact false belief understanding independently, but that these factors interact in important and interesting ways. In particular, we argue that situational factors themselves have an impact of cognitive development.

So far, we have highlighted the role of situational factors within the FBT itself. That is to say, that we have argued that various aspects of the FBT may make the task more or less demanding for the child. We have seen direct influences of situational factors which can facilitate FBT performance by highlighting the person involved. There is, however, also a further role in which situational factors can impact false belief understanding which we might think of as being a more *indirect*, namely by triggering the cognitive development. The helping behaviour

provides a good example of such an interaction between situational and cognitive factors. As we will explain in more detail below, in the helping behaviour the vicarious mental file is still activated through situational factors (direct influence) and it is through this situational activation combined with a need to act on reality that an initial link between mental files can be set up (indirect influence). This is also a good way of capturing an indirect impact from factors like maternal mental state discourse or having an older sibling which have been shown to improve performance on the FBT (Ruffman et al. 2002; Perner et al. 1994): this situational highlighting of other people's perspectives over time leads to an increased sensitivity to other people's perspectives. While factors like maternal mental state are not related to the task itself, they may play a role in highlighting children's overall sensitivity to other people's perspective. This increased sensitivity to other people's perspectives then facilitates cognitive development in the form of the linking. Putting this in terms of mental files, this highlighting of other perspective through the mother's discourse leads to activation of the vicarious mental file or even facilitates being able to construct vicarious mental files (direct influence). This in turn leads to a cognitive development facilitating the retrieval of the perspectives file in other situations too (indirect influence).

5 A New Account of the Paradox of the False Belief Task

Having introduced the mental files framework and our principles for the activation of the mental files, we are now in a position to make use of these tools to provide a new account of the development from implicit to explicit false belief understanding, considering both the role of situational and cognitive factors in order to map out the three stage development.

5.1 A Characterisation of the Three Stages

5.1.1 Stage 1: looking behaviour in the implicit false belief task (triggered by situational factors)

At the initial stage the mental files are still unlinked. This means that which file is activated is entirely dependent on situational factors. If the situation dominantly emphasises the other person, this primes the vicarious file and activation of the vicarious file allows children to pass the implicit FBT. The vicarious mental file contains the information where Maxi expects her chocolate to be and this is all the child needs in order to correctly predict Maxi's action in terms of their looking behaviour in the implicit FBT.

In this first stage, situational factors determine which mental file is activated and hence which information is actually available to the child. Nonetheless, we do not want to rule out completely a role for inhibition mechanisms even at this early stage. The reason for this is that the child is constantly confronted with reality which might lead to a dominance of the regular file. Therefore, some inhibition of the regular file *along with the situational facilitation of the vicarious file* might be necessary in order to allow for the activation of the vicarious file. More advanced inhibitions are not available yet.

5.1.2 Stage 2: active helping behaviour in the helping behaviour paradigm (triggered by cognitive and situational factors)

The helping behaviour paradigm requires coordinating information from both the vicarious and the regular mental file: the child needs to determine the experimenter's goal from the information on the experimenter's perspective in the vicarious file, but then in order to help the experimenter they need to consider the way the world actually is. To provide an account of this, we again have to begin with situational factors which initially lead to the activation of the vicarious file. In this case it is the experimenter coming in to open the box which places an emphasis on the experimenter and therefore the vicarious file is activated. The child is able to use the vicarious file to determine the agent's goal.¹⁷ But she can only help by acting on reality, however, so in aiming to help the agent the child reverts back to the regular file. The move from vicarious mental file to regular file is triggered through action, which necessitates the use of the regular file. By this we mean that it is this call for action which leads to the systematic activation of the regular file after a prior activation of the vicarious file. In other words, there is a unidirectional link from the vicarious file to the regular file (fig. 5): following the activation of the vicarious file due to situational factors, the information from the regular file is also accessible.

Given that this requires the coordination of more than one mental file, executive processes clearly also play an important role in this initial, unidirectional linking between mental files. It involves first, the same inhibition of the actual information that is already in place in implicit gaze behaviour, but also, second, a working memory ability such that the behavioural goal of the other taken from the vicarious file can remain activated when switching to the regular object file.¹⁸

It is important to note, however, that so far in the developmental story this link is only unidirectional: there is a link from vicarious mental file to regular mental file through action, but no link from regular mental file to vicarious mental file. As we will show below, this allows us to explain that children who pass the helping behaviour task still fail the explicit FBT.¹⁹

¹⁷ This step of determining the agent's goal and keeping it in mind is a central feature of the teleological account defended by the group of Perner (Priewasser et al. 2018, see above). Our analysis of unidirectional linking is compatible with this teleological analysis of the helping behavior since we argue that in the false belief condition there is a shift from the perspectival goal (sensitivity to the goal-directed behaviour of the experimenter to pick up the object) to the reality in which the helping behaviour has to take place (to get the object from where it actually is). But this needs unfolding in a different paper.

¹⁸ In section two, we discussed the possibility of the distinction between the two different types of implicit ToM collapsing. If this is the case, then unidirectional linking would underlie both stages. The important point to note here is that there is typically nothing about the looking behaviour task which requires unidirectional linking, while this is required for the helping behaviour task. For this reason we do not think that the looking behaviour tasks provide evidence of any linking of this sort.

¹⁹ A further example of this unidirectional linking is found in Mascaro and Morin's (2015) study where three year olds were able to infer the correct location of an object based on being told another's belief about the location of the object and that this belief is mistaken. In this scenario, the fact that the child did not know the real location of the object and the discussion of the other's belief would have served to highlight the other's perspective and activate the vicarious file. From this, however, children were then able to draw conclusions about reality.

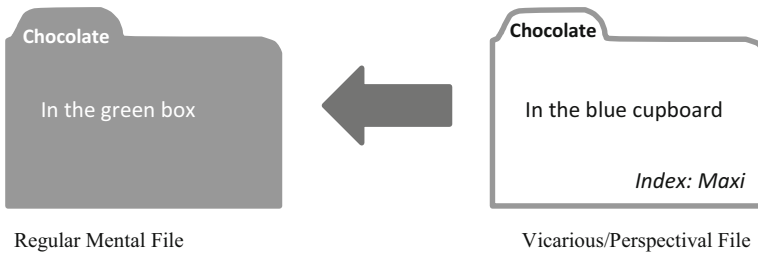


Fig. 5 Unidirectional link from vicarious file to regular file

5.1.3 Stage 3: the explicit false belief task (triggered by cognitive factor)

Like the helping behaviour, the explicit FBT requires the child to coordinate both the vicarious and the regular mental file. The direction, however, of coordination here is from regular mental file to vicarious mental file: as the question posed to the child places the focus on the object, there must be a switch from the regular mental file to the vicarious mental file. Children therefore do not succeed in the explicit FBT till the mental files are fully linked and they are able to access the information from the vicarious mental file in a systematic and controlled fashion after the activation of the regular file. The data in Rubio-Fernández (2013) support this view by providing evidence that an initial disruption of belief tracking also takes place in adults when they are asked the classic false belief question. The difference to children, however, is that adults are able to recover from this disruption whereas children do not. This fits with our suggestion that, due to the link between mental files, adults (and children older than 4) are able to revert back to the vicarious file following a distraction, whereas children before the age of 4 with unlinked mental files are dependent on situational factors and are therefore unable to recover from the distraction. The linking of the mental files leads to some independence from the situation itself thereby allowing children to succeed in the task even in the absence of situational facilitation. This idea of linking also includes a number of more sophisticated executive processes, namely the deliberate selection of the relevant perspectival information (the vicarious object file) such that this information is simultaneously available with the regular object file and the child can switch between both.²⁰

This leaves us with the question how the link in the reverse direction might be achieved. One thought here might be that executive abilities as well as symbolic representations acquired in the context of learning a natural language play a crucial role in this increased cognitive flexibility. But this remains a hypothesis which needs unfolding in a separate paper.

The linking of the mental files enables children to access perspectival information in a systematic and controlled manner and they are therefore able to pass the FBT even in unfavourable circumstances. Cognitive development in the form of the ability to link mental files allows for a decoupling from the situation such that children can reason

²⁰ De Bruin and Newen (2012) distinguish between an association module and an operational system to explain the development of the ToM ability. The structuring of information of objects and persons in mental files is the explication of what was called the association module while the development of execution and inhibition processes is exactly what determines the handling of the information stored in mental files. Thus, this explicates an operational system.

about the beliefs of others in some independence of the immediate situational factors. In the account we have presented, however, this decoupling itself originates in the early successes due to situational factors. Moreover, the initial systematic link is generated through a sensitivity for the goal-directed action of the other person in the situation such that one can adjust one's action e.g. to adequately help the other. The development of the decoupling from the situation, implemented as a linking between mental files, is triggered by situational factors and conditions for action, and supported by the development of executive processes. In terms of executive processes it involves the acquisition of 1. (automatic) inhibition of actual information, 2. working memory to keep activated relevant information about goals of others and 3. the deliberate selection of relevant perspectival information of a person.

5.2 Advantages of the Situational Mental File (SMF) Account

This account has a number of advantages. Firstly, we propose a detailed account of the gradual cognitive development underlying the development of children's ToM abilities and we have shown how this relates to the role of situational factors. While there are many theories which argue that cognitive development underlies the paradox of false belief understanding, many of these do not make explicit what the nature of this development might be and none describes how the development of all ontogenetic stages can be understood as part of a gradual development. Our SMF account is promising in that it offers such a theory in terms of the gradual development of linking of mental files. Furthermore, it includes principles which describe the interplay between situational and cognitive factors play a role in the development of this ability to link mental files.

Secondly, our account is able to provide an explanation for children's performance in the helping behaviour task, which we argued poses a problem for many of the previous accounts of children's ToM abilities, especially for the two system accounts: we need to presuppose *three central stages* in a rather continuous development including (a) the implicit looking behaviour, (b) the helping behaviour and (c) the explicit attribution of false beliefs.

Thirdly, we offer a framework to answer the open question whether passing implicit FBTs already involves a representation of *beliefs*. This question is so far usually answered in a stipulative way, e.g. while Baillargeon et al. (2010) claim that passing the looking behaviour FBT already involves belief representations while Apperly and Butterfill (2009) describe those as only belief-like representations but not yet beliefs, both positions do not offer a clear definition of having a belief representation. The SMF account includes the tools to defend the answer in a non-stipulative way. Given that we could argue that a belief representation presupposes a bi-directional linking between mental files (see Perner and Leahy 2016), then the SMF account implies that belief representations are developed not before age 3 and normally around age 4 in line with Rakoczy (2017) since this is the age when bi-directional linking underlies performance in the FBTs. We could offer a clear definition of having belief representations and have a story of the gradual development of mental files from being unrelated to bi-directional linking via unidirectional linking. Thus, we could integrate the answer when belief representations are active into a systematic description of the development.

Lastly, an advantage of our account is that it generates distinct and testable predictions. For example, we predict that having a task which emphasises the perspective of the agent over the object would facilitate the activation of the vicarious mental file and thereby improve children's performance on the FBT. This would include in particular tasks which have a strong interactive element or that are carried out in a cooperation or competition context, but also those in which reality is made relatively less salient such as by removing the object. Conversely, if the task emphasises the object, for example if the task involves a very interesting object or is presented as a more reality based problem solving task, children's performance will be impaired. One way in which one might test this is if there is a difference in performance between children who are introduced to the experimenter and told about them (agent condition), as opposed to children who are only introduced to the object and perhaps even allowed to play with it before administering the FBT (object condition). We would predict that children in the agent condition would perform better than those in the object condition as the emphasis on the agent as opposed to the object would increase the likelihood of the vicarious mental file being activated and thereby enabling them to pass the FBT. If we are correct in thinking that children pass the implicit false belief looking tasks because there is a focus on the other person in these tasks which is disrupted in the explicit FBT (Rubio-Fernández 2013), then we would also predict that if the implicit false belief is modified to break this focus on the other person, then infants should fail the implicit FBT too. One way of doing this would be to hide an object which is very attractive for the infants (for example a flashing toy). While initially children's performance is initially strongly influenced by the context of how the task is framed, we would predict that this effect of situational factors is reduced (as it can be compensated for) in older children. Generating such predictions for concrete experiments is especially important as it means that our account goes beyond post-hoc explanations.

6 Conclusion

We have argued that the development of children's performance in the FBT is neither due to purely cognitive nor purely situational factors, but rather that there is an interplay between these factors. The interplay can be described with two principles, namely the situational factors principle and the cognitive factors principle. While situational and cognitive factors remain active at all stages in ontogeny our account predicts that they develop a new dynamic at each stage. Using these principles children's early success on the implicit FBT can be explained by the situational factors principle. Violation of expectation FBT is a consequence of situational factors highlighting the person and thereby facilitating the activation of the vicarious mental file. While initially children's responses are determined purely by situational factors there is a cognitive development which allows for the gradual detachment from the situation²¹ characterised by the cognitive factors principle, thereby enabling children to pass the explicit FBT. In the SMF account we analysed the development of passing the helping behaviour FBT and then the explicit FBT as follows: due to situational aspects the person is in the focus of the child such that the child activates the vicarious mental file and is with 18 months of

²¹ This is not to suggest that situational factors do not continue to play an important role in ordinary social cognition but merely that it does not depend primarily on situational factors.

age able to represent the chocolate as belonging to Maxi or as desired by Maxi (in line with Wellman's (2002) desire psychology). It is through acting in a situation that a first systematic unidirectional link established between the vicarious file and the regular file. That is to say that it is through the challenge of acting in relation to the other goal-directed behaviour that children come to establish a unidirectional link which enables a systematic switch between accounting for someone else's perspective, i.e. wanting the chocolate, and then acting on their own reality by handing it over. In a second step the bi-directional link is established which enables the rather unconstrained use of the information of the files which are connected through the linking.

We have proposed the SMF account as a way of providing a satisfactory account of the developmental data from the FBT, but further investigation of the role of situational factors and how these relate to cognitive development along these lines could provide support for the account and allow for further elaboration of the account.

Acknowledgement: We would like to thank the German Research Foundation DFG which supported this research in the context of funding the Research Training Group "Situating Cognition" (GRK 2185/1).

Acknowledgements We would like to thank the German Research Foundation DFG which supported this research in the context of funding the Research Training Group "Situating Cognition" (GRK 2185/1)/gefördert durch die Deutsche Forschungsgemeinschaft (DFG).

Funding Information Open Access funding provided by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Apperly, I.A., and S.A. Butterfill. 2009. Do humans have two systems to track beliefs and belief-like states? *Psychological Review* 116 (4): 953–970. <https://doi.org/10.1037/a0016923>.
- Arora, A., B. Weiss, M. Schurz, M. Aichhorn, R.C. Wieshofer, and J. Perner. 2015. Left inferior-parietal lobe activity in perspective tasks: Identity statements. *Frontiers in Human Neuroscience* 9 (June): 1–17. <https://doi.org/10.3389/fnhum.2015.00360>.
- Baillargeon, R., R.M. Scott, and Z. He. 2010. False-belief understanding in infants. *Trends in Cognitive Sciences* 14 (3): 110–118. <https://doi.org/10.1016/j.tics.2009.12.006>.
- Baillargeon, R., D. Buttelmann, and V. Southgate. 2018. Invited commentary: Interpreting failed replications of early false-belief findings: Methodological and theoretical considerations. *Cognitive Development* 46 (May): 112–124. <https://doi.org/10.1016/j.cogdev.2018.06.001>.
- Baron-Cohen, S. 1995. *Mindblindness: An essay on autism and theory of mind*. Cambridge: MIT Press.
- Baron-Cohen, S., A.M. Leslie, and U. Frith. 1985. The autistic child have a "theory of mind"? *Cognitive Development* 21: 37–46. [https://doi.org/10.1016/0010-0277\(85\)90022-8](https://doi.org/10.1016/0010-0277(85)90022-8).
- Bartsch, K., and H.M. Wellman. 1995. *Children talk about the mind*. New York: Oxford University Press.

- Bradford, E.E.F., I. Jentzsch, and J.C. Gomez. 2015. From self to social cognition: Theory of mind mechanisms and their relation to executive functioning. *Cognition* 138: 21–34. <https://doi.org/10.1016/j.cognition.2015.02.001>.
- Buttelmann, D., M. Carpenter, and M. Tomasello. 2009. Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition* 112 (2): 337–342. <https://doi.org/10.1016/j.cognition.2009.05.006>.
- Buttelmann, D., H. Over, M. Carpenter, and M. Tomasello. 2014. Eighteen-month-olds understand false beliefs in an unexpected-contents task. *Journal of Experimental Child Psychology* 119: 120–126. <https://doi.org/10.1016/j.jecp.2013.10.002>.
- Buttelmann, F., J. Suhrke, and D. Buttelmann. 2015. What you get is what you believe: Eighteen-month-olds demonstrate belief understanding in an unexpected-identity task. *Journal of Experimental Child Psychology* 131: 94–103. <https://doi.org/10.1016/j.jecp.2014.11.009>.
- Buttelmann, D., F. Buttelmann, M. Carpenter, J. Call, and M. Tomasello. 2017. Great apes distinguish true from false beliefs in an interactive helping task. *PLoS One* 12 (4): 1–13. <https://doi.org/10.1371/journal.pone.0173793>.
- Butterfill, S.A., and I.A. Apperly. 2013. How to construct a minimal theory of mind. *Mind and Language* 28 (5): 606–637. <https://doi.org/10.1111/mila.12036>.
- Carey, S. 2009. *The origin of concepts*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195367638.001.0001>.
- Carpenter, M., J. Call, and M. Tomasello. 2002. A new false belief test for 36-month-olds. *British Journal of Developmental Psychology* 20 (3): 393–420. <https://doi.org/10.1348/026151002320620316>.
- Carruthers, P. 2013. Mindreading in infancy. *Mind & Language* 28 (2): 141–172. <https://doi.org/10.1111/mila.12014>.
- Carruthers, P. 2016. Two Systems for Mindreading? *Review of Philosophy and Psychology* 7 (1): 141–162. <https://doi.org/10.1007/s13164-015-0259-y>.
- Chun, M.M. 2000. Contextual cueing of visual attention. *Trends in Cognitive Sciences* 4 (5): 170–178. [https://doi.org/10.1016/S1364-6613\(00\)01476-5](https://doi.org/10.1016/S1364-6613(00)01476-5).
- Clements, W.A., and J. Perner. 1994. Implicit understanding of belief. *Cognitive Development* 9 (4): 377–395. [https://doi.org/10.1016/0885-2014\(94\)90012-4](https://doi.org/10.1016/0885-2014(94)90012-4).
- De Bruin, L.C., and A. Newen. 2012. An association account of false belief understanding. *Cognition* 123 (2): 240–259. <https://doi.org/10.1016/j.cognition.2011.12.016>.
- De Bruin, L.C., and A. Newen. 2014. The developmental paradox of false belief understanding: A dual-system solution. *Synthese* 191 (3): 297–320. <https://doi.org/10.1007/s11229-012-0127-6>.
- Dörrenberg, S., H. Rakoczy, and U. Liszkowski. 2018. How (not) to measure infant theory of mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development* 46 (February 2017): 12–30. <https://doi.org/10.1016/j.cogdev.2018.01.001>.
- Flavell, J.H. 2004. Theory-of-mind development: Retrospect and Prospect. *Merrill-Palmer Quarterly* 50 (3): 274–290. <https://doi.org/10.1353/mpq.2004.0018>.
- Flavell, J.H., E.R. Flavell, and F.L. Green. 1983. Development of the appearance-reality distinction. *Cognitive Psychology* 15: 95–120. <https://doi.org/10.1177/016502548901200407>.
- Gennari, S.P., M.C. MacDonald, B.R. Postle, and M.S. Seidenberg. 2007. Context-dependent interpretation of words: Evidence for interactive neural processes. *NeuroImage* 35 (3): 1278–1286. <https://doi.org/10.1016/J.NEUROIMAGE.2007.01.015>.
- Gopnik, A. 1993. How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences* 16 (1): 1–14. <https://doi.org/10.1017/S0140525X00028636>.
- Gopnik, A., and J.W. Astington. 1988. Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Source: Child Development* 59 (1): 26–37. <https://doi.org/10.2307/1130386>.
- Gopnik, A., and H.M. Wellman. 1992. Why the Child's theory of mind really is a theory. *Mind & Language* 7: 145–171. <https://doi.org/10.1111/j.1468-0017.1992.tb00202.x>.
- Gopnik, A., and H.M. Wellman. 2012. Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin* 138 (6): 1085–1108. <https://doi.org/10.1037/a0028044>.
- Grosse Wiesmann, C., A.D. Friederici, T. Singer, and N. Steinbeis. 2016. Implicit and explicit false belief development in preschool children. *Developmental Science* 1–15. <https://doi.org/10.1111/desc.12445>.
- Hansen, M.B. 2010. If you know something, say something: Young children's problem with false beliefs. *Frontiers in Psychology* 1 (JUL): 1–7. <https://doi.org/10.3389/fpsyg.2010.00023>.
- Helming, K.A., B. Strickland, and P. Jacob. 2016. Solving the puzzle about early belief-ascription. *Mind & Language* 31 (4): 438–469. <https://doi.org/10.1111/mila.12114>.

- Heyes, C. 2014. False belief in infancy: A fresh look. *Developmental Science* 17 (5): 647–659. <https://doi.org/10.1111/desc.12148>.
- Heyes, C. 2018. *Cognitive gadgets : The cultural evolution of thinking*. Cambridge: Harvard University Press.
- Kahneman, D. 2003. A perspective on judgment and choice. *American Psychologist* 58 (9): 697–720. <https://doi.org/10.1037/0003-066X.58.9.697>.
- Kahneman, D. 2011. *Thinking, fast and slow*. New York: Farrar, Straus & Giroux.
- Kahneman, D., A. Treisman, and B.J. Gibbs. 1992. The reviewing of object-files: Object specific integration of information. *Cognitive Psychology* 24: 174–219.
- Kammermeier, M., and M. Paulus. 2018. Do action-based tasks evidence false-belief understanding in young children? *Cognitive Development* 46 (November): 31–39. <https://doi.org/10.1016/j.cogdev.2017.11.004>.
- Kano, F., C. Krupenye, S. Hirata, and J. Call. 2017. Eye tracking uncovered great apes' ability to anticipate that other individuals will act according to false beliefs. *Communicative & Integrative Biology* 10 (2): 1–7. <https://doi.org/10.1080/19420889.2017.1299836>.
- Kinzler, K.D., and E.S. Spelke. 2007. Core systems in human cognition. *Progress in Brain Research* 164: 257–264. [https://doi.org/10.1016/S0079-6123\(07\)64014-X](https://doi.org/10.1016/S0079-6123(07)64014-X).
- Knudsen, B., and U. Liszkowski. 2012. 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy* 17 (6): 672–691. <https://doi.org/10.1111/j.1532-7078.2011.00105.x>.
- Kovács, Á.M., E. Téglás, and A.D. Endress. 2010. The social sense: Susceptibility to others' beliefs in human infants and adults. *Science* 330 (6012): 1830–1834. <https://doi.org/10.1126/science.1190792>.
- Kulke, L., and H. Rakoczy. 2018. Data in brief implicit theory of mind – An overview of current replications and non-replications. *Data in Brief* 16: 101–104. <https://doi.org/10.1016/j.dib.2017.11.016>.
- Kulke, L., M. Reiß, H. Krist, and H. Rakoczy. 2017. How robust are anticipatory looking measures of theory of mind? Replication attempts across the life span. *Cognitive Development* (August): 0–1. <https://doi.org/10.1016/j.cogdev.2017.09.001>.
- Kulke, L., B. von Duhn, D. Schneider, and H. Rakoczy. 2018. Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. *Psychological Science* 29 (6): 888–900. <https://doi.org/10.1177/0956797617747090>.
- Leslie, A.M. 1987. Pretense and representation: The origins of “theory of mind.” *Psychological Review* 94 (4): 412–426. <https://doi.org/10.1037/0033-295X.94.4.412>.
- Leslie, A.M., O. Friedman, and T.P. German. 2004. Core mechanisms in “theory of mind.”. *Trends in Cognitive Sciences* 8 (12): 528–533. <https://doi.org/10.1016/j.tics.2004.10.001>.
- Lewis, S., V. Hacquard, and J. Lidz. 2012. The semantics and pragmatics of belief reports in preschoolers. *Salt* 22 (1): 247–267.
- Mascaro, O., and O. Morin. 2015. Epistemology for beginners: Two- to five-year-old Children's representation of falsity. *PLoS One* 10 (10): e0140658. <https://doi.org/10.1371/journal.pone.0140658>.
- Mascaro, O., O. Morin, and D. Sperber. 2017. Optimistic expectations about communication explain children's difficulties in hiding, lying, and mistrusting liars. *Journal of Child Language* 44 (5): 1041–1064. <https://doi.org/10.1017/S0305000916000350>.
- Mitchell, P., and H. Laco  e. 1991. Children's early understanding of false belief. *Cognition* 39 (2): 107–127. [https://doi.org/10.1016/0010-0277\(91\)90040-B](https://doi.org/10.1016/0010-0277(91)90040-B).
- Moll, H., A. Khalulyan, and L. Moffett. 2017. 2.5-year-olds express suspense when others approach reality with false expectations. *Child Development* 88 (1): 114–122. <https://doi.org/10.1111/cdev.12581>.
- Moses, L.J., and J.H. Flavell. 1990. Inferring false beliefs from actions and reactions. *Child Development* 61 (4): 929–945. <https://doi.org/10.1111/j.1467-8624.1990.tb02832.x>.
- Murez, M., and J. Smortchkova. 2014. Singular thought: Object-files, person-files, and the sortal PERSON. *Topics in Cognitive Science* 6 (4): 632–646. <https://doi.org/10.1111/tops.12110>.
- Onishi, K.H., and R. Baillargeon. 2005. Do 15-month-old infants understand false beliefs? *Science (New York, N.Y.)* 308 (5719): 255–258. <https://doi.org/10.1126/science.1107621>.
- Perner, J. 1991. *Understanding the representational mind*. Cambridge: The MIT Press.
- Perner, J., and J.L. Brandl. 2005. File Change semantics for Preschoolers: Alternative naming and belief understanding. *Interaction Studies* 6 (3): 483–501.
- Perner, J., and B. Leahy. 2016. Mental files in development: Dual naming, false belief, identity and Intensionality. *Review of Philosophy and Psychology* 7 (2): 491–508. <https://doi.org/10.1007/s13164-015-0235-6>.
- Perner, J., and T. Ruffman. 2005. Infants' insight into the mind: How deep? *Science* 308 (5719): 214–216. <https://doi.org/10.1126/science.1111656>.

- Perner, J., S.R. Leekam, and H. Wimmer. 1987. Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology* 5 (2): 125–137. <https://doi.org/10.1111/j.2044-835X.1987.tb01048.x>.
- Perner, J., T. Ruffman, and S.R. Leekam. 1994. Theory of mind is contagious: You catch it from your sibs. *Child Development* 65 (4): 1228.
- Perner, J., S. Stummer, M. Sprung, and M. Doherty. 2002. Theory of mind finds its Piagetian perspective : Why alternative naming comes with understanding belief. 17: 1451–1472.
- Perner, J., M. Huemer, and B. Leahy. 2015. Mental files and belief: A cognitive theory of how children represent belief and its intensionality. *Cognition* 145: 77–88. <https://doi.org/10.1016/j.cognition.2015.08.006>.
- Perry, J. 1990. Self notions. *Logos* 11: 17–31.
- Priewasser, B., E. Rafetseder, C. Gargitter, and J. Perner. 2018. Helping as an early indicator of a theory of mind: Mentalism or teleology? *Cognitive Development* 46 (February): 69–78. <https://doi.org/10.1016/j.cogdev.2017.08.002>.
- Psouni, E., A. Falck, L. Boström, M. Persson, L. Sidén, and M. Wallin. 2018. Together I can ! Joint attention boosts 3- to 4-year- olds ' performance in a verbal false-belief test together I can ! Joint attention boosts 3- to 4-year- olds ' performance in a verbal false-belief test. *Child Development* 00 (0): 1–16. <https://doi.org/10.1111/cdev.13075>.
- Rakoczy, H. 2012. Do infants have a theory of mind? *British Journal of Developmental Psychology* 30 (1): 59–74. <https://doi.org/10.1111/j.2044-835X.2011.02061.x>.
- Rakoczy, H. 2017. In defense of a developmental dogma: Children acquire propositional attitude folk psychology around age 4. *Synthese* 194 (3): 689–707. <https://doi.org/10.1007/s11229-015-0860-8>.
- Recanati, F. 2012. *Mental Files*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199659982.001.0001>.
- Roby, E., and R.M. Scott. 2018. The relationship between parental mental-state language and 2.5-year-olds' performance on a nontraditional false-belief task. *Cognition* 180: 10–23. <https://doi.org/10.1016/j.cognition.2018.06.017>.
- Rubio-Fernández, P. 2013. Perspective tracking in progress: Do not disturb. *Cognition* 129 (2): 264–272. <https://doi.org/10.1016/j.cognition.2013.07.005>.
- Rubio-Fernández, P. 2018. What do failed (and successful) replications with the Duplo task show? *Cognitive Development*. <https://doi.org/10.1016/J.COGDEV.2018.07.004>.
- Rubio-Fernández, P., and B. Geurts. 2013. How to pass the false-belief task before your fourth birthday. *Psychological Science* 24 (1): 27–33. <https://doi.org/10.1177/0956797612447819>.
- Ruffman, T. 2014. To belief or not belief: Children's theory of mind. *Developmental Review* 34 (3): 265–293. <https://doi.org/10.1016/j.dr.2014.04.001>.
- Ruffman, T., L. Slade, and E. Crowe. 2002. The relation between children's and mothers' mental state language and the theory-of-mind understanding. *Child Development* 73 (3): 734–751.
- Siegal, M., and K. Beattie. 1991. Where to look first for children's knowledge of false beliefs. *Cognition* 38 (1): 1–12. [https://doi.org/10.1016/0010-0277\(91\)90020-5](https://doi.org/10.1016/0010-0277(91)90020-5).
- Sodian, B. 2005. Theory of mind. The case for conceptual development. In *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind*, ed. W. Schneider, R. Schumann-Hengsteler, and B. Sodian, 95–130. Mahwah: Erlbaum.
- Southgate, V., C. Chevallier, and G. Csibra. 2010. Seventeen-month-olds appeal to false beliefs to interpret others ' referential communication. *Developmental Science* 13 (6): 907–912. <https://doi.org/10.1111/j.1467-7687.2009.00946.x>.
- Tomasello, M. 2018. How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences* 115 (34): 8491–8498. <https://doi.org/10.1073/pnas.1804761115>.
- Vierkant, T. 2012. Self-knowledge and knowing other minds: The implicit/explicit distinction as a tool in understanding theory of mind. *British Journal of Developmental Psychology* 30 (1): 141–155. <https://doi.org/10.1111/j.2044-835X.2011.02068.x>.
- Warneken, F., and M. Tomasello. 2007. Helping and cooperation at 14 months of age. *Infancy* 11 (3): 271–294. <https://doi.org/10.1111/j.1532-7078.2007.tb00227.x>.
- Wellman, H.M. 1990. *The Child's theory of mind*. Cambridge: MIT Press.
- Wellman, H.M. 2002. Understanding the psychological world: Developing a theory of mind. In *Handbook of childhood cognitive development*, ed. U. Goswami, 167–187. Oxford: Blackwell Publishing Inc.
- Wellman, H.M. 2014. *Making minds : How theory of mind develops*. New York: Oxford University Press.
- Wellman, H.M., and K. Bartsch. 1988. Young children's reasoning and beliefs. *Cognition* 30 (3): 239–277. [https://doi.org/10.1016/0010-0277\(88\)90021-2](https://doi.org/10.1016/0010-0277(88)90021-2).

- Wellman, H.M., D. Cross, and J. Watson. 2001. Meta-analysis of theory-of-mind development : The truth about false belief. *Child Development* 72 (3): 655–684. <https://doi.org/10.1111/1467-8624.00304>.
- Westra, E. 2016. Pragmatic development and the false belief task. *Review of Philosophy and Psychology* 8 (2): 235–257. <https://doi.org/10.1007/s13164-016-0320-5>.
- Westra, E., and P. Carruthers. 2017. Pragmatic development explains the theory-of-mind scale. *Cognition* 158: 165–176. <https://doi.org/10.1016/j.cognition.2016.10.021>.
- Wimmer, H., and J. Perner. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13: 103–128. [https://doi.org/10.1016/0010-0277\(83\)90004-5](https://doi.org/10.1016/0010-0277(83)90004-5).

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