What do four-year-olds have in common with Frege?

A theoretical and practical investigation of the false-belief task

Marian Counihan

October, 2002

Acknowledgements

Contents

1 Introduction

For just on twenty years, there has been an almost unbroken interest in a little test which two psychologists, Josef Perner and Heinz Wimmer, first conducted in 1983. Perner and Wimmer's Maxi task, which Perner called "the first systematic experimental investigation into children's understanding of how a person develops a false-belief" (Perner 1991), is these days only one version among many so-called 'false-belief tasks'. Since 1983, the results of the original experiment have been replicated scores of times, a myriad of modified versions have been conducted (with varied effects on results), and a host of other abilities have been tested for possible correlation with false-belief mastery. Tests have also been performed on autistic and deaf children, with very different results from normally developing children. And that is only the empirical data. The interpretations and hypotheses offered to explain the results of this task are just as numerous and varied. Why this massive interest and discussion around what at first appears to be a simple task?

I think the continued interest in this test indicates two things about false-belief testing. Firstly, it is still unclear exactly what capacity is indexed by successful performance on the task. Secondly, despite the first point, performance on false-belief tasks has been shown to be affiliated with competencies in language, counterfactual reasoning, understanding of causality, and theory of mind.

Analogously, this thesis has two aims. The first of these is to contribute to the discussion about what the false-belief tasks actually test. This end was partly served by a pilot study conducted by David Wood and myself. It is also served by a advancement of the original ideas propounded by Josef Perner *Understanding the Representational Mind*. The literature on false-belief is full of arguments using similar results to support diverging theories, but I think that a lot of the supposed divergence is illusory, and that a good deal of the theorizing about this task may not be as antagonistic as it likes

to present itself. As support for this, I offer the themes of intentionality and Fregean semantic theory as means by which diverse research may be interpreted in a unified way.

Secondly, this thesis aims to provide an overview of the research that has been done in attempting to establish the aforementioned correlations. Overview is too generous a word - rather let's say I will indicate several capacities which seem to play a role in enabling, or at least be correlated with, performance on the task, and evaluate the research that has been done to evidence these links. Exactly what the links look like: how these various capacities are related to false-belief mastery - for example the direction of influence, the mechanisms which implement them - is mostly unclarified. I will describe suggestions from the literature and in some cases make conjectures of my own. Further than that, this work offers principally suggestive conjectures for further research directions. The principal capacities which will be discussed are to do with the linguistic structure of complements and ability to reason with counterfactual statements.

There is a third point to be made here, which is more to do with us, than with the test itself. False-belief attribution is perhaps so interesting to us because it indicates something which we consider to be a significant and peculiarly human trait: understanding others' external states, their behaviour, through their internal states - specifically their epistemic states - and not just through their actions and preferences, or, for that matter, from the way the world is. Attributing a false belief to someone really allows them status as another mind, another intentional agent, and predicting another's behaviour on the basis of their false belief gives higher priority (in reasoning about that person) to their mental states than to the physical state of the world. As such, the research in this field also provides quite a fascinating glimpse of the complex interplay of faculties which are required to operate in a world of intentional agents. Thus discovering how, when and why false-belief understanding emerges, may really be discovering a rich source of information about human cognition. There is no third aim to this thesis,

to further this idea.

Turned on its head, however, this point provides motivation for continuing study of and experimentation with false-belief tasks (this paper included). Why these experiments? Well, if in fact, these tasks do provide a measure of our ability to attribute mistaken beliefs to others, then they are really a key indicator of a developing 'theory of mind' (usually called commonsense or folk psychology in the philosophical literature). Specifically, they indicate emergent understanding that what goes on in our minds often takes precedence over how things are in the world, in determining our behaviour. So anyone investigating the developmental sequence by which we come to understand ourselves and others as intentional agents, can find in the research around this subject a gold-mine of relevant data.

One last remark before I start. This thesis examines the subject of false-belief testing from two angles: firstly, by evaluating experimental work that has been done on the subject, and secondly, by discussing the philosophical issues underlying that work. It is often difficult to determine exactly what the relevant issues might be, until the empirical research has been evaluated. Moreover, these issues are usually wide-ranging, deep and hotly debated philosophical topics. Consequently, this work is by no means self-contained or complete. It is an exploratory work which has as its ultimate aim suggestive indication, and not comprehensive theory.

2   Our experimental work

David Wood and I conducted a pilot study with a group of eleven Dutch preschoolers, aged between 2;11 and 4;0, (mean age of 3;5 years) with the explicit aim of exploring the robustness of young children's inability to perform well on false-belief tasks. As the children in our group were all (bar one) below the reported passing age, we needed to introduce measures which would simplify the task if we were to get any correct responses at all.

Firstly, we conducted a variation on the traditional changed location task: instead of using dolls to enact the scenario, we showed the children an episode of a popular children's television programme. Second, we played hide-and-seek with the children. Both of these modifications of the original task have been reported to lower the passing age (Surian and Leslie 1999), but had no discernible effect in our experiments.

Due to the small size of the subject group, and the fact that this was our first foray into the world of experimental psychology, I don't propose to draw any definite conclusions from this study. Rather, I will use the data we collected as a suggestive indication and illustration of factors which potentially affect performance. This experimental work was valuable for me, as a means of making concrete evidence which forms the basis for any theorising on the subject of false belief, and thus I think it also renders the theorising more robust and more relevant, as it is tied to the data, so to speak. Actually listening to a three year old repeatedly insist that their teacher knows where they are hidden, somehow piques and focusses your attention on elements of the situation which might be missed with purely theoretical research!

2.1 Method, modifications, results and discussion

We first asked the child a series of introductory questions intended to test his/her understanding of and ability to respond to sentences with complement structures, future hypotheticals, counterfactuals, and true belief situations. This was also a means to accustom the child to question-answer situation and make them more comfortable with and responsive to the experimenters. The questions and the answers we obtained are given in tables in the appendix.

2.1.1 Bob the Bouwer test

We showed the child an episode of a popular children's television programme, 'Bob de Bouwer' (the Dutch version of the BBC's 'Bob the Builder'), in

which Bob gets stuck on the roof of a shed, after the naughty scarecrow Spud takes his ladder in order to pick apples off the tree. Originally we thought it would be best to let the children see the episode several times, or at least once all the way through, before we asked them questions about it. This would ensure that they were familiar with the storyline and characters, so that uncertainty in answering could not be attributed to uncertainty about who or what events the questions referred to.

However after a few pilot runs we realised that this strategy might also work against us, as one child in particular (Bambi, 4;0) got very confused about the story line. In answering she referred to events in the narrative which had not yet happened at the time we wanted her to refer to, but which she had already watched happen in a previous viewing. We realised that our approach may heighten confusion over the narrative structure and thereby add complexity to the interpretation of the question. It was difficult to specify that we wanted to know what Bob's epistemic state was at that particular point in the episode, and difficult to know how to get this across the child. After this we decided the problem could be minimised by stopping before the end of the episode, at the point where Bob gets down from the roof, but does not yet know what has happened to his ladder. Several researchers have discussed the effects of decreasing the saliency of the current, changed location on the children's responses to the false belief question (Perner 2000, Surian and Leslie 1999, Robinson and Beck 2000)- for example by making the actual current location unknown or unclear - but as yet there is no consensus on how much effect this has on performance. Certainly it did no harm to modify the experiment thus, as at most it made the task easier and thus would have eliminated false negative answers.

We however did show the child the crucial segment - where Spud takes Bob's ladder - twice. During the second viewing, we paused the video at key points in this segment to ask further questions. These included the so-called memory and reality control questions. We also asked the children a counter-factual question, after the key false-belief question. The vast majority of our

subjects answered the control question correctly, and both the false belief and counterfactual questions incorrectly. The full results can be also found in the appendix to this paper. They have been taken from the experimental summary: *Two Experimental Variations on the Standard False Belief Task*, written up by David Wood. Here it will suffice to say that the responses given matched those reported in the literature on this task. Most children made clear realist errors, confidently answering that Bob thought that the ladder was by the tree, where Spud had taken it. Thus our results confirmed the hypothesis that children below the age of 4 years old are unable to answer explicit questions about false belief.

2.1.2 Hide and seek

The second version of the task that we conducted was a simple game of hide and seek. At the end of the first task, on a prearranged cue the teacher (who had been sitting in with the child) got up and left the room, supposedly to fetch a letter for the experimenter. As soon as she had left, the experimenter asked the child if they would like to play hide and seek. Every child responded enthusiastically to this suggestion and David and the child proceeded to get into a play-tent, out of sight. Once hidden, the experimenter asked the child two questions: had their teacher had seen them hide?, and where did she think they were?

The responses on this task, although replicating other results, nevertheless provide a simple and very striking illustration of failure to attribute false belief. The full table of results is again available in the appendix. They show astonishing consistency: every child bar one answered the first question ('Did Debbie/Tamara [the creche leader] see us go and hide?') correctly with a confident (even delighted) 'no'. But the next question was 'Where does Debbie/Tamara think that we are?', to which the answer was 'here' or 'in the tent'! Almost every subject was unhesitating and unswerving in their response, even when David asked a more leading question: 'Does she think we are here, in the tent, or over there, by the TV?' (pointing to where they had been sitting). The children seemed very well to grasp the fact that they

were playing a game, which involved hiding from the teacher. False belief is really built into the game dynamic. And yet they couldn't attribute it to the seeker - or at least when directly asked, they answer as if they don't. This seemingly bizarre pattern of answering could be attributed to several factors, between which our experiment could unfortunately not distinguish. The following issues are relevant here, and will be addressed in later sections of this thesis:

- Chandler, Fritz and Hala (1989) suggest that participating in a game of hide and seek already demonstrates early workings of a theory of mind - the subjects are engaging in false belief ascriptions, they just aren't able to explain them yet. Their thesis proposes to use deceptive acts as an index of false-belief attribution, rather than standard verbal responses to questions. Chandler et al think that the standard task leads to underestimation of young children's theory of mind abilities because it "conflates the active capacity to entertain beliefs about beliefs with the altogether different ability to comment upon this understanding".

- Does the child realise that you have to see to believe? Seeing leads to knowing, belief is weaker - you can come to believe by inference, from evidence, or from others' utterances. Related to this is an account which has been defended by Heitner (1999, also Chandler et al 1989). This more modest explanation attributes failure on false-belief tasks to a failure to appreciate the importance of informational access for belief formation.

- The sureness of response indicates that the subjects were positively answering *some* question, if not the one we intended! Which aspect of the question do they misinterpret? Is it a lexical item? Is it the syntactic structure? Is the child just answering the question, 'Where are we now?' One child alternated his emphatic "here" with "hidden" - which would seem to indicate one of three things: either he was answering the simpler question, or he was getting confused about how far the pretence was extending - so that the teacher was also in on

it, or he was confused about the time the question referred to. It is a temporally sensitive question and the answer will change once, when the teacher first notices the absence of the child (when "hidden" becomes a perfectly acceptable answer), and again when the teacher finds the child (when "here" becomes the correct answer).

- Omniscience of the teacher, as a source/authority for knowledge. The teacher did in fact report an interesting observation: although repeatedly instructed to inform the teacher when they leave the room to go to the bathroom (not to ask for permission), the children never do this. They seem surprised when the teacher asks them where they have been[1].

The above comments provide a indication of the numerous and varied elements which come into play in testing false belief. The next sections will cover in more detail some of these elements and attempt to elucidate their role in both testing and performance on the task.

3   The original task and competing schools of explanation

Wimmer and Perner's original task is what is now called a 'displaced object task': two characters (usually dolls or puppets, sometimes drawn figures, rarely real people) put an object in location A, then one character (Maxi, say) goes away and is still absent when the other character moves the object to location B. When Maxi arrives back on the scene the subjects are asked either where Maxi thinks that the object is, or where he will look for the object (differing results of the effect that the form of the question has on performance have been reported - see for example Surian and Leslie, 1999). The other main type of false-belief task is the 'unexpected contents' task, where children are shown a familiar container, such as a Smarties box,

---

[1]Results from the Smarties test where children are asked about their own beliefs and their playmates' beliefs show that the failure to attribute false belief is robust, and thus argues against the idea that the teacher's role in the game affects results significantly.

and asked what they think is inside. They are then shown the unexpected contents - a plastic frog, say, and asked about their previous (mistaken) belief. They are often also asked what someone else (most often a playmate) would expect the contents of the box to be. The subjects on these tasks are typically 3 to 5 year olds, and there is a sharp improvement in performance somewhere after 4;0 years: 3-year-olds usually make 'realist' errors - that is answering with the actual location of the object, or the actual contents of the box. 5-year-olds mostly answer correctly: in the first task, they realise that the character will look where he mistakenly thinks the object is: in location A. In the second type of task, subjects correctly can recall their previous mistaken belief and also suppose their friend, on seeing the closed box, will have the same (mistaken) belief about the expected contents.

Normally developing children are only one group of subjects who have been tested. Results from work with autistic and deaf children provide fascinating complicating factors to any theorising on the subject of false-belief testing. Numerous studies have shown that autistic children fare very badly on standard false-belief tasks, and often show no improvement even into adolescence, when they have verbal mental ages way over 4 years. They typically also make realist errors. More recently, research with deaf children has shown they also suffer very delayed competence on false-belief tasks.

There are various schools of explanation for these results, most of which fall within two broad categories. The first category covers those who favour the plainest interpretation of the data - namely, that it indexes false belief mastery. These theorists propose that the sharp increase in performance noted in false-belief task data indicates a correlated shift in the child's conceptual understanding of the task. As such the task indexes a major development (Perner 1991, Perner 2000, Moses and Flavell 1990, Woolley 1995, Wellman 1991). The general idea is that there is a fundamental development in understanding the representational character of mental states, and specifically belief states, between the ages of three and four years old, and so the test does just what it says it does: tracks the child's understanding

of false belief. This view is usually evidenced by work with normally developing children, and the most prolific proponent of this type of view is Josef Perner, whose account is outlined below.

The other main category focuses on the continuity of the developmental sequence and supposes that the increase in performance between three and four years old is represented as an artificially discontinuous curve in the data. On these accounts, the task results should rather be seen as a reflection of the maturation of mechanisms, and a decrease in computational and processing limitations. Hence younger children's understanding of false belief may be masked by other cognitive deficits, and their failures on standard tasks should be interpreted as false negatives (Robinson and Mitchell 1995, Chandler et al 1989, Surian and Leslie 1999). Tasks with simplified format have been used to support this view: showing that 3-year-olds have heightened success on versions with lessened computational capacity requirements (achieved by simplifying the narrative structure, decreasing the saliency of reality, changing the form of the question, for example), strengthens the hypothesis that 3-year-olds have a basic conceptual grasp of belief[2].

A noted proponent of the latter type of view is Alan Leslie. Leslie describes what he calls a 'theory of mind mechanism' (ToMM), and proposes that it is the development of this mechanism which enables performance on false-belief tasks. This view postulates an earlier domain-specific competency which would explain the early-emerging ability to pretend and understand pretence: "According to the theory of ToMM, an early competence is embodied in a pre-structured representational system that provides the child with a domain-specific and probably modular learning device. ... For

---

[2]One such modification was to change the question from "Where will Maxi look ...?" to "Where will Maxi look *first* ...?" (Surian and Leslie 1999). Original experiments by Siegal and Beattie reported a significant improvement in 3-year-olds performance when asked the latter question, but there have been puzzling failures to replicate these results (Lewis and Osborne 1990). Even if the results are robust, however, we still need to know why 3-year-olds need extra pragmatic prompting, and how inclusion of the word 'first' accomplishes this prompting.

'continuity theories' such as ToMM, the failure of 3-year-olds on false-belief tasks is to be understood in terms of performance limitations of various kinds which prevent or attenuate the deployment of an inherent competence"(Surian and Leslie 1999). Impetus for this view comes from cognitive theories on autism. Autistic children perform very badly on the standard false-belief tasks, and this, coupled with their lack of spontaneous pretend play, has been taken as evidence for a specific module in the brain which implements theory of mind, by theorists such as Leslie and Baren-Cohen. Other 'continuity theories' focus on executive control mechanisms, and inhibitory processing ability (Russell 2002, Riggs et al 1998, Bartsch and Wellman 1995).

I will focus in the next section on the first variety of explanation, and unpack what it involves. What does it mean to attribute a false belief to someone? What understanding of the mind is needed? I will draw attention to the usefulness of the notion of intentionality, which is taken to be key in cognitive science and philosophical literature, in understanding the nature of mental states, and yet which does not occur in any of the psychological literature on the subject of false belief. I believe that a description of the intentionality of the mental, and how it is expressed in language, can be also usefully invoked in two experimental studies that I discuss: that is, the work with deaf children done by de Villiers and de Villiers (2000), and the research into counterfactual aspects of false-belief reasoning done by Riggs et al (1998). This approach is nevertheless still aligned with the simple interpretation that the false-belief task indexes false belief mastery, since intentionality is a key aspect of representation. As such, I hope to illustrate that the concept of intentionality is a unifying theme of divergent strands of research.

There is another area of debate in the theory of mind literature, and this is focussed on how our theory of mind is implemented. Here the division is between what is called 'theory-theory' and simulation. Broadly, the former proposes that we reason about others' internal states by inferring them from

general principles, which compose a 'theory'. The latter proposes that we employ a strategy of what Peterson and Riggs call "adaptive modelling". The basics of this are that we reason about others' minds using our own mental apparatus, but with two modifications: a) we adjust the inputs suitably to make our systems similar in relevant respects, and b) we run a simulation, which means we somehow take our mental processes 'off-line', so that any conclusions reached don't result in the accompanying action, and result in limited affective reaction. The false-belief task seems to be neutral between these two views, and so I will not discuss this issue further here.

Here is a brief overview of the current major theories on the data from false-belief testing:

Theorist	Test subjects	Proposed reason for failure	Evaluation	Refer to
Perner (1991, 2000)	3- to 5-year olds	Lack of understanding of mind as representational	Compatible with most other theorists, underspecified thesis as yet	Chapter 4
Surian and Leslie (1999)	Autistic subjects	Lack of ToM module	Unlikely that a single module responsible for wide and varied range of autistic deficits	Not discussed further here
	3- to 5-year olds	Performance limitations	Underspecified at this stage	
Russell (2002)	Autistic subjects	Lack of executive control mechanisms	Unexamined here	Not discussed further here
Riggs et al (1998, 2000)	3- to 5- year olds	Inability to reason counterfactually	Counterfactuality needs to be considered more carefully	Chapter 7
de Villiers and de Villiers (2000)	Deaf children	Lack of requisite syntactic structure (complementation)	Relevant on level of enabling, probably compatible with most above theories	Chapter 5

4   The Representational Mind

It is of course a matter of huge debate whether or not our mind does work
as a representational device. How does that discussion bear on the topic of
this paper? I don't think it does. What matters here is not how the mind
actually works, but rather how we *think* it works, and this commonsense
conception of the mental is indeed presumptive of a representational mind.

Cummins' (1989) distinction between the problem of representations and the problem of representation is helpful here. The former is a concern of empirical science and covers debates about how mental representations are instantiated and what role they play in mental processes: the connectionist versus computational debate, for example.

The topic of this thesis, however, is the development of theory of mind in young children and hence we are concerned with the latter problem: the problem of representation. This is a concern of philosophy, and must explicate the notion of mental representation, taken as primitive in the above-mentioned discussions. Moreover we are not even concerned with the notion of representation as used in scientific theory, but that which informs the commonsense psychology, the layperson's theory of mind. And the central element of this is the idea of intentionality: that which attaches beliefs and desires to their contents.

It seems timely to specify here what I mean by 'commonsense psychology' or 'theory of mind', as it is called in the psychological literature. This is the structure or functioning by which we understand and reason about our own and others' minds, and how we use this understanding to predict and explain our own and others' behaviour and intentions. We routinely make reference to states such as desires and beliefs, to explain the actions of ourselves and others: I ate an apple because I wanted to relieve my hunger, and I believed that an apple would achieve that. This triad of beliefs, desires and actions is invoked in the common reasoning used in everyday life, but in fact only actions are directly observable. Ascribing beliefs and desires thus involves some kind of theoretical postulation, and it is the emergence of this ability which concerns theory of mind acquisition theorists. Moreover, if the false-belief task does in fact index competency in false-belief attribution, then it is a key test for theory of mind. This is because success shows quite explicitly that the child grants mental states, such as belief, primary status in determining behaviour. Hence success on the task is taken by some theorists as evidence of the last stages of acquisition of a fully-fledged theory

of mind. One such theorist is Josef Perner, who set out to show (1991) the developmental sequence of theory of mind acquisition, and thus also to explain the role of false-belief task data in indexing this.

4.1 The basics of a Perner-style thesis

An explanation for the notable difference in performance observed between 3- and 4-year-old subjects is offered by Josef Perner in his book, *Understanding the Representational Mind*. The bare bones of Perner's theory are this: our mind is a representational device. It has a three-tiered representational faculty. The first level of representation is primary representation, which involves a single model of the world, and the child acquires this level - that is, the ability to model reality - from birth. The second level builds on this, and allows the child to entertain multiple models of reality, enabling hypothetical reasoning about the future, the past, and alternative situations, including counterfactuals (more about this later). The capacity for secondary representation emerges at around 15-18 months of age. Finally, at around 4 years old, the child acquires the concept of representation (so can make 'models of models'), and thus can now 'metarepresent'. Perner's notion of metarepresentation is strictly second-order representation: representation of representation. He distinguishes it from what he calls 'metarepresentational comments'. Although these comments may modify the status of representation - such as marking them true or false - they don't qualify as metarepresentations proper[3]. This last stage is required to be able to understand misrepresentation, and this is precisely what 3-year-olds cannot do. They lack the conception of the mental as representational, and hence cannot conceive that someone is able to misrepresent some state of affairs. Thus good performance on the task indicates a significant conceptual shift, from what Perner calls a 'mentalistic theory of behaviour' to a fully-fledged 'theory of mind', as the child gains the concept of representation. This development (viewing the mind as a representational device) enables the

---

[3]These 'metarepresentational comments' are closer to Leslie's use of the term 'metarepresentation'.

child to understand two key notions, according to Perner. These notions are the fallibility of mental representation, and the causal origins of it (Perner, 2000).

If one is to take seriously the hypothesis that an inadequate conception of the mind as a representational device is responsible for younger children's failure on false-belief tasks, then a more detailed and more specific formulation of the hypothesis needs to be given. There are two main aspects to this: specifying the nature of mental representations, and specifying which aspects children need in order to attribute false beliefs. The former should answer questions such as: What makes a state a representation? What is it for a state to have a semantic content, a meaning? And further, what differentiates states such as believing from other mental states such as desires? The latter should answer questions such as: What is it about the nature of representing, and, more specifically, believing, that children come to learn when they pass the test? How does this interact with linguistic capabilities? Answering these last questions (at least partially) is really the task of the whole thesis. So now I turn to the former task: specifying the nature of mental representations.

There are two ways of specifying how something represents something else: either by specifying properties of a representation itself, or by specifying the *relation* which holds between a representation and the thing it represents. I will use the latter. The relation between a state of mind and its object is peculiar in several ways (Dennett and Haugeland, 1987): firstly, for physical relations, the things related exist independently of the relation. But for mental relations "what a belief is supposed to be about is crucial to which belief it is." So even though two beliefs may refer to one and the same object, they are different beliefs if they represent it differently. For example, my belief that the black spot in the distance is a horse is different from my belief that the black spot in the distance is a large black spot. Dennett and Haugeland give the example that "one and the same belief cannot at one moment be about a frog (that it is green, say) and at another moment

19

be about a house (that it is green). The latter is a different belief." Presumably both beliefs are 'about' greenness but are different because they represent greenness in different ways. Secondly, each of the things related may not exist: you can have beliefs about unicorns. Thirdly, whether or not the relation holds depends on how the relation is specified: although I may hold the belief that Cicero was an orator, if I don't know that Cicero and Tully are the same person, I do not also believe that Tully was an orator. This is the referential opacity of propositional attitudes such as belief[4].

For reasons of space and efficiency, instead of launching into a long-winded evaluation of the thesis offered by Perner, and attempting to first fix the broad and sometimes abstruse notion of representation, I will take his hypothesis only as a starting point and rather concentrate on the two aspects of representation which are seemingly most relevant when it comes to attributing false beliefs, and suggest how these figure in theory of mind acquisition. Those are, as mentioned, their fallibility as true representations and their causal formation (applicable specifically to beliefs). One incontrovertibly core aspect of mental representations, which I think offers a clear description of at least the fallibility aspect, is their 'aboutness'. You can't believe without believing *something*, you can't want without wanting *something* - you just can't represent without representing something. Representations are always *about* something. When referring to mental representations, this 'aboutness' is often called intentionality. Cummins (1989) says: "most philosophers.. assume that the problem of mental representations is just the problem of intentionality." Cummins goes on to argue that this is a mistaken assumption, but it nevertheless serves to illustrate how pervasive the connection is. This is especially relevant for a discussion of commonsense psychology, and in fact I think it is sensible to shift the analysis from that of representation to that of intentionality. The above mentioned features of the (mental) representing relation are captured by intentionality, and it is a more easily specifiable feature, as I will show. Moreover I propose that

---

[4]and it also is evident in the axioms of epistemic logic: knowledge of p, together with the inference $p \rightarrow q$, does not grant knowledge of q.

the results of false-belief testing can be quite comprehensively explained by invoking this key feature, and it can also be appealed to, to explain results in counterfactual reasoning tasks. It can be aligned with a linguistic notion of intentionality (first described by Chisholm) which in turn offers a concrete theoretical underpinning to the work done by de Villiers et al on deaf children. The second aspect - the causal formation of beliefs - I will only discuss briefly in the last section.

One of the philosophers who Cummins may have had in mind is Brentano, who held that intentionality is the distinguishing feature of mental states.

4.2   The 'aboutness' of mental representation

Brentano's intentionality thesis states that, for any mental activity, there is always some object to which that activity is directed: "In presentation something is presented, in judgment something is affirmed or denied, in love loved, in hate hated, in desire desired, and so on." (Brentano, p 88) But the distinctive feature of the objects of mental activities is that they need not exist. This, for Brentano, is the feature which marks off the mental from the physical: "This intentional in-existence is characteristic exclusively of mental phenomena. No physical phenomena exhibits anything like it." Any directed physical activity must be directed towards an existing object, but one can quite happily think of a unicorn, a non-existing object. A remark before going further: another way of stating this is to talk of the *aboutness* of mental *states*. Since this seems sufficiently similar to Brentano's formulation,[5] I will use this terminology, instead of talking of the 'directedness' of mental activities.

So there are two aspects of Brentano's doctrine of intentionality: one concerning the 'aboutness' of mental states, another the ontology of the objects of this 'aboutness'. Assigning ontological status to the objects of thought is where theories of intentionality diverge. Brentano assumed a

[5] for Brentano, "activity" means "being concerned with something", "having something as object" (Brentano 1973).

relational account of experience in which there are only two aspects of experience: the act and the object. This assumption meant for Brentano that the 'aboutness' of mental states entails their "intentional in-existence", which he takes to imply mentalistic ontology for the objects of thought: they only exist in the mind. This is neither a necessary nor a helpful consequence, because if objects of thought only exist in the mind, they are essentially private. This throws up epistemological difficulties: how can we come to know about physical objects? How can objective reference to entities which exist outside the mind, in the realm of the physical, be achieved (Locke's 'veil of perception')? It also throws up questions about how mental objects are related to physical objects: directly, or indirectly? And which has primary ontological status[6]?

In order to avoid mentalism, or rather the 'private' and thus objectively unreachable character being ascribed to mental states, we can turn away from the act-object dichotomy presumed by Brentano and rather posit a three-levelled theory of meaning. This is what Brentano's pupil, Husserl, did. I will begin by focussing on semantic concepts borrowed from language, specifically Frege's view of language, and then see how Husserl expanded on these ideas[7]. Frege's three-levelled semantic theory is ideally suited to mental state ascriptions. Chisholm (1957) has argued that the psychological

---

[6]It is much more useful here to follow Wittgenstein, who argues against the possibility of a private langauge, propounding rather that our mental vocabulary acquires its meaning from the public shared language to which it belongs. I think this is already hinted at by Frege when he says: "in order to justify mention of that which a sign means it is enough, at first, to point our intention in speaking or thinking" (On Sense and Reference) See section on language for elaboration on this idea.

[7]The tertiary level of representation that Perner describes, also relies on a Fregean approach to the nature of representational relations: "believing is not just a two place relation between organism and proposition but is a three place relation... that relates the organism to the proposition and the world against which the proposition is to be evaluated." (Perner 2000). Perner has a unusual starting point though - usually a dichotomous account of belief states distinguishes between sentence and world, or act and object, but not between organism and proposition. The third aspect is not usually the world against which the proposition is to be evaluated.

notion of intentionality can be translated into a linguistic one if we use a Fregean approach to language, and in the other direction, Husserl has argued that Frege's notion of *Sinn* can be generalised to apply to all acts, not just linguistic signs. So although Frege was strongly anti-psychologist, the insights generated by his introduction of the notion of *Sinn* may still be very useful in elucidating the mental concept of intentionality. In fact I think that it is not unreasonable to suppose that the structure of the language we use to describe the mental reflects our understanding of it, and thus examining the syntax and semantics of mental state ascriptions goes a long way in capturing the structure of theory of mind as we use it in every day life. We now come to Frege's idea of:

4.3   Distinguishing sense from reference

The motivation for Frege's three-levelled semantic theory was the observation that the two true identity statements '$a = a$' and '$a = b$' nevertheless differ in "cognitive value" - the former being a matter of *a priori* truth, and offering no new information, while the latter is informative (and thus the truth of it may not always be established *a priori*). Signs (for example linguistic signs such as words) are primarily designators - they designate or refer to something else. For instance the sign 'a' designates the object a (whatever it may be). But the statement '$a = b$', if true, says that the signs 'a' and 'b' refer to, or designate, the same thing (the identity relation holds between the objects they designate), and so if this is to be informative then there must be a difference between the signs 'a' and 'b' which is not a feature of what they designate. These considerations led Frege to introduce, apart from the 'reference' of the sign (what the sign designates) what he calls the 'sense' of a sign, "wherein the mode of presentation is contained", and further, the relation between a sign and its reference is always mediated by its sense. This distinction he applied firstly to proper names - replace 'a' and 'b' by 'evening star' and 'morning star' respectively, to see this illustrated. Their reference is the celestial body Venus but they have different senses. In Frege's terms: "A proper name (word, sign, sign combination, expression) *expresses* its sense, *refers to* or *designates* its reference." The

notion of sense can be extended to apply to whole sentences: the sense of its components combine compositionally to give the 'complete' sense of the sentence, which Frege calls a thought. Further, he chose (perhaps badly) that a sentence refers, not to a proposition but to the truth value of that proposition, which famously means that "in the meaning/reference of the sentence all that is specific is obliterated". This may not have been a wise choice - choosing truth conditions for the reference may have been better - but it won't have too much import on what we discuss here. I will focus on how the sense/reference distinction sheds light on several features of language.

Firstly, this distinction can be used to explain how it happens that we can talk about things which do not exist: unicorns, centaurs, the present king of France. A two-levelled semantics is either forced to judge these expressions meaningless[8], because they do not designate or refer to anything in the real world - as names, they don't have a bearer (as Russell propounded), or to suppose that they refer to objects which exist only in the realm of the mental (cf Brentano's act-object thesis). The latter type of account runs into all sorts of difficulties in explaining how these objects can nevertheless achieve objective reference, and transcend the privacy and subjectivity of the mental. The former type I will not consider here principally because I think it is not very useful. But on a Fregean account, each name necessarily has a sense, although it may not have a reference: "In grasping a sense, one is not certainly assured of referring to anything." So this explains how we can meaningfully talk about things which, although they do not have a referent, always have a sense. Secondly, a Fregean semantics can explain referential opacity. This is discussed further in the section on language.

In the meanwhile, let's compare this with a notion of representation proper. There are suggestive parallels to be drawn between the Fregean sense/reference distinction and the distinction Nelson Goodman describes as that between pictorial "representing" and "representing-as" (Goodman,

---

[8]or rather to be not names at all but descriptions

1976). We can recognise this distinction by observing the different uses of the locution "represents ... as". The first use is plain representation: "This picture represents Winston Churchill as an infant", where the picture shows Winston Churchill when in his infancy (one of his mothers' old photographs, say). "Here "... as" combines with the *noun* ... to form a description of one portion of the whole extended individual" (Goodman, 1976). The second use of the locution is, for example, illustrated by "This picture represents Winston Churchill as a bulldog", where the picture referred to is a political cartoon, for example. In this case, the "... as" combines with and modifies the *verb*, and is the genuine case of 'representing-as'. Further, "while [ordinary representing] concerns only what a picture denotes, and [a fiction depiction] only what kind of picture it is, [representing-as] concerns both the denotation and the classification." The denotation is the person, object or situation that the representation represents, and the classification is *how* it represents it. But this is only pictorial representation - what about mental representation[9]?

Consider now the more cognitively-oriented version of Frege's semantic distinction: propositional attitude reports - for example belief attributions - must indicate the way, or mode, in which the content of the belief is represented by the believer. Thus, in an embedded context, the referent of the expression is no longer its referent in the world (an individual or situation, say), but rather the way it is represented in the believer's mind - for instance, as a proposition. Thus to fully understand the structure of propositional attitude reports, one needs to realise that they refer to propositions, not situations in the real world. This is the core of Perner's thesis: "younger children have procedural knowledge of evaluating propositions against the world. They can assign truth values; they can decide whether the propo-

---

[9]Empirical research which addresses this issue is available in the data from on false photograph tests, which are largely designed to test if failure on false-belief tasks is due to a failure to appreciate the nature of representation in general, and not specifically mental representation. There are very mixed results from these studies, some confirming equal difficulty for mental and physical representation tasks, and some reporting non-mental representations - such as pictorial representation - easier.

sition is true or false, but they do not understand that propositions are being evaluated. Consequently, they cannot conceive of the possibility that the proposition could be evaluated differently... The children who pass the false-belief task have achieved this level of understanding" (Perner 2000). Here a big qualification needs to be made: the conversion from a theory of linguistic meaning to a theory of cognitive meaning is not immediate by any means. In fact, Frege was vigorously anti-psychologistic. He specifically says "Rules for asserting, thinking, judging, inferring, follow from the laws of truth. And thus one can very well speak of laws of thought too. But there is an imminent danger here of mixing different things up. ... one might come to believe that logic deals with the mental process of thinking and the psychological laws in accordance with which it takes place. This would be a misunderstanding of the task of logic, for truth has not been given the place which is its due here. Error and superstition have causes just as much as genuine knowledge" (Frege 1967).

Can we respect Frege's anti-psychologism and yet expand the idea of sense to apply to mental acts too? According to Føllesdal (1969), Husserl's notion of noema does exactly this.

4.4 Husserl's notion of noema

> The noema is nothing but a generalization of the idea of meaning (Sinn) to the field of all acts. (Husserl, Ideas)

Following his teacher Brentano, Husserl held that intentionality is the essential characteristic of consciousness: "the peculiarity of consciousness is to be consciousness *of* something." But Husserl then departs from his teacher: whereas Brentano struggled to explain the possible non-existence of the object of consciousness, Husserl overcomes this difficulty by inserting an intermediate concept into Brentano's act-object dichotomy: the *noema*. Briefly, Husserl's account of the of the intentionality of perception is: the noema (or meaning) of a perceptual act is the mode in which the object is intended in that particular act, so the relation between an act and its intended object

26

is mediated by the noema.

Føllesdal (1969) furthers the thesis that this notion is a generalisation of Frege's notion of *Sinn* to a theory of all acts, and thus that Husserl's trichotomy of

   act - noema - object

is a parallel of Frege's trichotomy of

   sign - sense - referent.

The following aspects of the notion of noema are described by Føllesdal in support of his thesis: associated with each act is a noema in virtue of which the act is directed towards an object. There need not exist an associated object (such as one when thinks of a unicorn), and each object may be associated with several different noema. However each noema is associated with only one object. Noemata are abstract entities and thus not perceivable through the realm of senses, but rather through special reflection, the "phenomenological reflection", which can be iterated. The noema is an intentional object which provides the means of achieving objective reference.

Compare these qualities with the following aspects of Frege's sense: Every sign necessarily has a sense but only contingently a reference. Moreover, many different senses may be associated with one reference. Senses are not private: they are abstract yet objective entities which occupy the 'third realm'; they are what is grasped in understanding language, and can act as possible referents. Now we are in a position to see how this theory might account for the mind-relatedness of linguistic meaning: analogously to above, the relation between a sign and its referent is mediated, this time by linguistic meaning which is captured by the notion of *Sinn*.

Husserl's tripartite theory of meaning is surely only one among others which pay heed to Fregean semantics. Here it serves to indicate that the intuitions which a distinction between sense and reference captures, can be

expressed in more than linguistic terms. We can express the intentionality of mental acts in terms of Husserl's noema, which connects in linguistic acts with Frege's *Sinn*. I further surmise that advancement in false-belief mastery may indicate the replacement of an act-object dichotomy with a act-noema-object trichotomy, in understanding the relational account of experience. It is certainly compatible with the patterns observed in the data and hypotheses presented in the literature - for instance, inability to conceive of two differing experiences of the same reality can be interpreted a two-levelled theory of meaning, and acquisition of a three-levelled theory then enables competencies such as false-belief attribution. This could also be understood as acquisition of the concept of intentionality.

5   Language

There are several results which indicate children have an implicit understanding of the nature of belief long before they can answer explicit questions about it correctly. O'Neill's illustration of this (1989 - unpublished) was even reported in Perner's 1991 book: 2 year olds watched while a researcher put a toy out of reach on a high shelf. If a parent was absent while the researcher placed the toy, and only came into the room subsequently, the subjects gestured towards the shelf much more than if the parent was present while the toy was put out of sight. Perner and Clements (Perner 2000) have conducted further investigations into this with a modified version of the false-belief task, using children's eyes as an indicator of anticipation. Sam the mouse puts his cheese at one hole (location A), and then disappears. The cheese is moved to another hole (location B). Sam only reappears when he pops out of either of these holes, and so children's gaze can be monitored as an indication of where the child expects Sam to appear. According to Perner, "the finding was striking:... after a sharp developmental onset almost 80% of children at the age of 3 years (2;11 - 3;2) did show implicit knowledge. And there was a large gap at this age between visually orienting to A and answering with 'A' the explicit question where Sam will go to look for his cheese " (Perner 2000). This result has been replicated using

video animations and observing spontaneous reactions. Further anecdotal evidence comes from our empirical data. While watching the episode for the first time, Minne spontaneously commented, at the moment that Bob sees his ladder is gone, that "he can't find it". He nevertheless went on to answer the false belief question incorrectly, and unhesitatingly so.

So, although 3-year-olds may have some understanding of false belief, the ability to answer explicit questions about false belief emerges much later - past the 4 year mark in most cases. What is this lag due to? What extra capacity is required for the jump from implicit to explicit representation of false beliefs?

A possible answer to this question is that it is the specific cognitive demands placed on the child by the formal tasks which inhibit correct answering. A large part of this task-load is the linguistic capability required for the task. The following sections discuss this in further detail.

5.1 The syntax of mental state explanations: is the medium part of the message?

"You have a new conception and interpret it as seeing a new object. You interpret a grammatical movement made by yourself as a quasi-physical phenomenon which you are observing... But there is an objection to my saying that you have made a 'grammatical' movement. What you have primarily discovered is a new way of looking at things" (Wittgenstein, 1958).

De Villiers and de Villiers (2000) have reported results from several studies which show that the most significant (linguistic) predictor variable for both spontaneous and future performance on the false-belief test, with normally developing children, is the production of sentential complements. This in itself isn't anything more than suggestive, but it is a possible explanation for the large amount of realist errors which are always recorded in false-belief studies. If the child cannot comprehend the whole question, in

29

complement form, ('Where does Bob think that the ladder is?'), then they may just process the parts they can, and answer the simpler form 'Where is the ladder?'[10]. But the startling data from work with deaf children provides stronger support for their claim that the language of complements is causally relevant in explicit false-belief attribution. The strongest position indicated by the evidence is that syntactic structures in statements about mental states may in part determine how we understand mental states in the first place, or at least, how we format that understanding. De Villiers and de Villiers are careful not to overstate their claim: "it has become strikingly evident in much recent work that developments in theory of mind, including false-belief understanding, may be a prerequisite for understanding in complex linguistic tasks..... The overall process is undoubtedly one of mutual facilitation between language and theory of mind." Nevertheless, they defend a position which imputes a "much more significant role to language development that is currently being discussed", namely that "a child who becomes capable of the language of complementation, namely embedded propositions, might have available a new *representational capacity* for propositional attitudes. That is, perhaps the complete syntax that is used for describing mental events makes possible the representational changes that allow for understanding false beliefs."

To assess this position, they conducted a series of experiments. Firstly, they did a longitudinal study with normally developing children aimed at showing the order of language and false-belief development. They collected data four times over the period over a year (although only the first three results are published in the mentioned article), from a series of tasks testing both language and false-belief competency. On their analysis the production of sentential complements was the most significant predictor variable for both current and future performance on false-belief tests.

---

[10]This observation is probably the motivation for the de Villiers' claim that "we cannot but sure that complementation in child language is *genuine* until the second clause *can* be false". I will come back to this later.

The language tasks were designed to test use of complementation through memory and spontaneous speech. This may be the closest empirical access one can gain, but I nevertheless think these tests are both too narrow (the memory tasks)and too broad (spontaneous speech), to gauge complementation ability. The difficulty on the memory tasks arises because, at least according to de Villiers, the two primary classes of verbs which take embedded sentential complements are communication verbs ('say' and 'tell') and mental state verbs ('think', 'believe' and 'know') (de Villiers and de Villiers 2000). So the questions used de Villiers and Pyers in the longitudinal study, seem to presuppose understanding of mental states. The children were asked:

> He thought he found his ring, but it was really a bottle cap.
> What did he think?

> She said she found a monster under her chair, but it was really the neighbour's dog.
> What did she say?

> This girl saw something funny at a tag sale and paid a dollar for it. She thought it was a toy bird but it was really a funny hat.
> What did she think she bought?

These questions enquire about propositional attitudes, or at least reports of propositional attitudes. At this point, the test seems to require something remarkably similar to what false-belief tasks test. The child must realise that the subject represents the actual situation in another way to answer the questions successfully. According to de Villiers and de Villiers, it is difficult to eliminate this similarity without undermining the test, because even if we don't accept that embedded complements are usually taken by communication or mental state verbs (they don't mention perception or desire verbs), "we cannot be sure that complementation in child language is *genuine* until the second clause *can* be false" - that is, the child realises

that the truth of the embedded clause does not determine the truth of the whole utterance. Hence other verbs - such as perception or desire verbs - which may embed sentential clauses, are not suitable to measure this, because these verbs when used with complements do not allow for referential opacity - if Mary sees that the shop is closed, then the shop really has to be closed. So there is no way to distinguish when the child is reporting Mary's propositional attitude from when the child is mistakenly reporting reality - as is so often the case in false-belief testing. This argumentation is a step too quick, and there is an intermediate option: there are certainly desire verbs which do allow referential opacity and thus are an ideal vehicle for testing. Consider the following:

> He hoped the sun was shining outside, but it was really raining.
> What did he hope?

> He wished that it was Saturday, but it was really only Friday.
> What did he wish?

To answer correctly the child must report the content of a propositional attitude and not the actual situation in the world, but no understanding of belief is necessary[11]. Since desire-psychology is known to develop earlier than belief-psychology, these questions seem much more suitable to test understanding of the syntactic structure of complements than the ones used by the de Villiers. As it stands in the tasks used in the reported studies, it is not surprising that succeeding on these tasks is a predictor for performance on false-belief tasks proper - they are just a watered-down version of the real thing. de Villiers and de Villiers claim a partial defence of this is

---

[11]There are also some perception verbs which are borderline cases. For example:

> John smelt the smell of pear-blossom but it was actually just Mary's new perfume. What did he smell?

I don't know what a correct answer would be in this case - depends on whether you take a wide or narrow view of content. Also, the distinction between 'see' and 'see that' could be exploited for testing whether children understand the possibilities of complement structures.

provided by the evidence of concomitant spontaneous use of complements in speech, which do not have to be about mental events, nor do they have to be false, but nowhere give concrete data about the pattern of emergence of this spontaneous use.

There also remains the possibility that the child is merely repeating what he/she has just heard in giving the correct answer - i.e. is just 'parroting', and has not understood the question (although the last question posed by De Villiers et al ask does vary the form thereby avoiding this charge). The data argue against this, however, since younger children are inclined to make the typical 'realist error' and report the true clause (eg "Friday" in last example), which is the opposite of parroting, and thus when older children do report the embedded clause it seems more likely this is due to increased comprehension and not the more regressive seeming 'parroting' strategy.

Actually how can you test only understanding of complements? It is very difficult to see how you could tease apart comprehension of the syntactic structure without at least also partial interference from the content of the utterance. Woolley (1995) has shown that children answer questions about what someone has imagined (even though it differs from reality) earlier than they correctly answer questions about false belief. Three-year-olds performed well on tasks which involved questions such as "What is [the experimenter] imagining is inside this box - is she imagining there's an apple inside [imagined contents of box] or is she imagining (there's a pencil inside/that it's empty) [actual content]?" (Woolley 1995). Thus it appears children can handle complement sentences with desire/imagination verbs before they can handle complement sentences with belief verbs. How this interacts with the de Villiers proposal is unclear, but it does suggest that mastery of complements, although possibly necessary, is clearly not sufficient and must be augmented by understanding of belief before false-belief attribution is possible. I will come back to this point later.

More exhortative evidence comes from another source:

5.1.1 Studies with deaf children

Most deaf children have hearing parents, who are not fluent signers, and this shows up especially in the first few years of the child's language development, which is known to be significantly delayed compared to normal children (de Villiers 2000). Moreover, several studies (Russell et al 1998, de Villiers and de Villiers 2000) have shown that deaf children raised in a spoken environment show a delay in performance in false-belief tests. This of course is a fairly pedestrian observation with respect to verbal testing - interpretation of the question is an immediate factor - but there is evidence from non-verbal testing to suggest deeper interaction between linguistic competence and theory of mind capabilities. In particular, it raises the possibility that linguistic structures play a role in theory of mind development: that language itself gives us an idea of what we are using language to talk about.

Deaf children have been shown to be socially impaired: egocentrism and delayed role-taking abilities are two primary impairments. However, these deficits are primarily correlated to language skills and show up much less in non-verbal tasks, and so "....any subtle social deficits in deaf children seem to be causally tied to their development of language and communicative skills" (de Villiers and de Villiers, 2000). If this is accepted, studies with deaf children provide an excellent opportunity to partition out the contribution made by language structure, to our understanding of the mental. De Villiers and de Villiers argue thus:

> If we are wrong in attaching importance to language as a catalyst for developing false-belief understanding, then deaf children will succeed at these tasks on the basis of their nonverbal understanding of social interaction, which might be acute, given their dependence on it rather than language for predicting others' behaviour. Indeed, it could be that such an essential ingredient for proper social interaction is a "robust" module of mind, buffered against the vagaries of cultural and biological accidents, and hence accessible by a variety of developmental routes, much

34

like language itself.

In order to test this hypothesis, slightly modified versions of the displaced object task and the unexpected contents task were carried out with a group of moderate to profoundly deaf children with normal non-verbal IQs, aged 4 to 9 years old. They were also tested for language ability. This was done by showing the children video clips depicting events which involved mistakes or deception and thus the narrative could only be adequately explained with reference to characters' mental states: desires, lack of knowledge and false beliefs. The children were asked increasing specific questions to elicit mental state explanations. The children were then scored according to the spontaneity and complexity of the explanations given. Performance on standard false-belief tasks[12] was compared with scores from this task, and also with age, degree of hearing loss and verbal mental age. The results showed that "producing complex sentences with cognitive state verbs like think and know was by far the strongest predictor of the children's false-belief reasoning on the standard verbal tests... and remained significant even when the effects of the other three variables were partialled out."

What does this actually show? Of course you need sufficient linguistic ability to pass a verbal test. What needs to be further specified is exactly where linguistic competency comes into play: is it just in parsing questions and producing answers? Or does it also affect the reasoning process? To address this, de Villiers et al propose non-verbal versions of the tests. If linguistic demands in interpretation and expression are lessened, do deaf children perform better in false-belief tasks?

Two non-verbal tests were used. The first tested understanding of the relationship between seeing and knowing, and among both hearing and deaf

---

[12]with the key false belief question in changed location version of the form "Where will the boy first look for his cake?" This version of the question was used because it is not in a complement form, and the word *first* was inserted because this has been shown by some theorists to lower the passing age on the false belief task. Although these results are have not been consistently replicated, at the very least this variation of the question does not make the task more difficult.

subjects, the average age of the passers on this task was almost identical to that of the passers on standard tasks. The second test involved attributing appropriate facial expressions to cartoon characters on the basis of their expectations - basically a version of the unexpected contents task, except with an extra step. Subjects had to infer a what a character's emotional reaction would be, on the basis of their epistemic state, instead of just reporting that epistemic state (what the standard task requires). This proved to be slightly more difficult than the standard version, for both hearing and deaf subjects, which is not unexpected, since new task-processing demands are placed on the subjects (although verbal demands lessened). Nevertheless these two tasks seem sufficiently similar to the original verbal task to be comparable in terms of understanding of theory of mind. Yet the results again indicate a delay in deaf children's performance (production of sentential complements and age proved to be the two significant predictors of performance, general language ability was not), and further, performance on the two tasks was delayed to the same degree. This indicates that it is not just language requirements on the task which lead to the delay on false-belief tasks in deaf children.

There are other studies which provide evidence for this hypothesis - for instance Russell et al's (1998) paper confirms that false-belief mastery among deaf subjects is severely delayed and that performance is age-related (and not IQ-related). Russell et all used the 'changed location' test. A qualified teacher of the deaf signed and enacted the standard story using two characters (John, Mary) and a toy aeroplane, and also posed the questions. Since this test uses verbal narrative and questioning, the possibility is introduced that it is linguistic inadequacy which impairs performance, and not cognitive inadequacy. To lessen interference from linguistic difficulties with the test and thereby counteract this possibility, Russell et al used within-task control questions, the so-called "reality" and "memory" questions. Subjects were deemed to have passed only if they answered all three questions (reality, memory and false-belief) correctly. The subjects were split into three groups and the results were as follows: of the youngest group (mean age

6;7), two of ten passed the test. In the middle group (mean age 10;11), one of nine children answered correctly. In the oldest group (mean age 15;5), six out of ten children passed. Unfortunately, Russell et al don't give the figures for the children who answered only the control questions correctly - they only specify that "a large majority of the children correctly answered control questions". Yet the discrepancy between those who answered the two control questions correctly, and those who didn't answer any correctly, needs to be significantly smaller than the difference between those who passed and those who only answered control questions correctly, for the results to be remarkable. If this discrepancy doesn't reach significance, then the possibility that the results index linguistic competency (or lack thereof), as mentioned earlier, re-emerges. Studies with deaf children should benefit from using non-verbal tasks such as those outlined by de Villiers and de Villiers.

In any case, a more precise specification of the developmental sequence in the acquisition of complements is now needed, to add substance to this proposal. De Villiers and de Villiers outline a possible model, but specify that it is only at the level of speculation. It looks like this:

Step 1: the child masters the basic sentence forms: a simple sentence is mapped onto a simple event. The child encounters true sentences that match reality.

Step 2: the child first encounters discrepancy between sentences and reality: the child learns to recognise pretence as well as mistakes.

Step 3: the child masters first embedded structures under verbs of communication/mental state/desire: child acquires the fundamental syntax of embedding but makes no accommodation of meaning within that structure. That is, the complement retains its truth value as a simple sentence independent of the matrix verb. So, if the child hears a sentence such as "Jim said he ate the broccoli", the child thinks it is true *both* that Jim said something, and that he ate the broccoli.

37

Step 4: the child first notices occasions with verbs of communication that suggest the complement can be false when embedded e.g. reports of lying, mistakes. For example, the child notices that what Jim *said* he ate is not what he really ate. Because statements are overt and can be compared to reality, the semantic accommodation is made evident.

Step 5: the discovery about semantic accommodation mastered with the complements of verbs of communication can now be extended to verbs of mental states, e.g. beliefs. The child can then understand a statement such as "Jim thought he ate broccoli", to imply that Jim had a false belief. The "thought" is not overt, so it must be inferred from actions or statements, but the analogy with communication structures allows the sentence to be understood, thus the possibility of other minds with thoughts that do not map onto one's own reality is given expression.

De Villiers and de Villiers note that this model focusses on the role of exposure to acts of "mis-speaking", such as lying and mistakes, as the means to master sentence complements[13]. Further they propose that the situations in which the embedded clause is intensional are marked by some feature in the complement phrase (CP), and so that if young children lack that feature, or "have it "set" to the default form, then they will not be able to represent reports of lies or mistakes. When children can finally represent that feature, it is only the first of a series of features in an articulated CP feature set that eventually must accommodate at least tense perspective,..., speaker beliefs / referential opacity [this is where false-belief attribution would come in] and factivity... In other words, it is argued that the full articulation of linguistic *structure* eventually accommodates all the other special *semantic* features of

---

[13]Realist errors are prevalent in this kind of attribution until at least 3.5 years of age. de Villiers and de Villiers report that "children have difficulty with questions such as: "What did he say he drank?", if the character said he drank something other than what he really drank. The strong tendency is for young children to respond as if they were answering instead the question: "What did he drink?" (de Villiers and de Villiers, 2000).

certain propositional attitude reports." (de Villiers and de Villiers, 2000). Hollebrandse (1999) has investigated the correlation of sequence of tense acquisition[14] with false-belief mastery for Dutch children, but his paper unfortunately does not give clear concrete data to evidence his claims about this correlation. Further research in these areas will allow a more thoroughgoing critical evaluation of the contribution of linguistic ability to false-belief attribution. Certainly the existence of a developmental delay in deaf children

---

[14]A past tense embedded under a matrix past tense can be interpreted in two ways: firstly, as past with respect to the matrix event, and secondly, as simultaneous with the matrix event. This occurs in both Dutch and English. Hollebrandse gives the example: "Koekiemonster zei dat hij een rood bordje had", which translates as "Cookie Monster said that he had a red plate". An embedded future tense does not have the same flexibility: different tenses are used to differentiate between two different readings. There is one for when the embedded event is future only with respect to the matrix event ("Cookie Monster said that he would have a red plate"), and another for when it is future also with respect to time of utterance ("Cookie Monster said that he will have a red plate"). Correct use of sequence of tense distinguishes these four different possible sequences of events, but requires a distinction between time of utterance and time referred to by the utterance. Hollebrandse bases his experiment on the premise that "A child lacking complementation can only interpret the "embedded" past tense directly to UT [utterance time]. This child will then also allow the forward shifted reading before UT." In other words, the child will mistakenly allow a third reading of "Cookie Monster said that he had a red plate": the forward shifted reading, which places the embedded event - Cookie Monster's possession of a red plate -*after* the matrix event - Cookie Monster's utterance. Hollebrandse tests for this using the following script:

  B: Zal ik eens kijken of ik een banaan voor je kan vinden, Cookie Monster (sic)?

  (Let me have a look whether I can find a banana for you, CM)

 KM: Ja Bert, ik will een banaan op mijn bordje hebben

  (Yes Bert, I will have a banana on my plate)

  [B puts the banana on CM's plate]

 Exp: [to child] Zei Cookie Monster dat hij een banaan op zijn bordje had?

  Did Cookie Monster say that he had a banana on his plate?

The results are unfortunately sketchily reported, but anyway I think the wording of the experiment is ambiguous - 'wil' is both 'will' and 'want' in Dutch, and so Cookie Monster's utterance can also mean that he wants to have a red plate. This renders the question unclear, and thus diminishes the significance of results obtained.

has implications for theories of mechanisms underlying the development of theory of mind, but exactly what those implications are, is at this stage, a matter of speculation. There is also the possibility that deaf children fail false-belief tasks for different reasons to normally developing children. Early signs of this come from studies of an emerging sign language, which is developing in Nicaragua (de Villiers et al 2002), and indicates that deaf adults may fail because they cannot express mental representations in their simple language. Younger deaf children, on the other hand, might be more inclined to fail because they lack a proper understanding of causal relations. This suggestion needs to be investigated further before any conclusions can be drawn from it.

Apart from experimental indications, are there theoretical grounds to attribute this specific construction (namely complementation) such a key role in false-belief mastery? The semantics of complements allow the possibility of talking about a fundamental distinction between actual situations and represented situations. In particular, the situation represented in the embedded clause may not match reality. De Villiers makes the dubious claim that this is the only construction which has this feature: "only embedded complements have the property that they can be false yet the sentence that contains them can still be true. In all other instances where propositions combine to make sentences, if any of the propositions is false then so is the whole" (de Villiers and de Villiers 2000). This is clearly far too simplistic, as any conditional statement illustrates - in particular, as counterfactuals illustrate. But the semantics of propositional attitude reports are nevertheless 'special', as Frege first pointed out. This is the topic of the next section.

5.2 A Fregean approach to propositional attitude reports

The problem of propositional attitude reports, first systematically tackled by Frege in his paper 'On sense and reference', has remained a hot topic in philosophy since then: Barwise and Perry (1983) call it the "bane" of semantic theory. But the psychological literature on belief attribution pays little attention to this highly relevant issue. I argue that a Fregean approach

is very helpful here, at the very least by providing a formal framework into which de Villiers and de Villiers' claims about the semantics of propositional attitude reports can be fitted.

The introduction of the notion of sense, as discussed above, helps to explain another feature of language - namely that co-referring expressions cannot always be substituted for one another without disturbing the truth value of the sentence in which they occur. The principle occasion of this is in propositional attitude reports. This is because "If words are used in the ordinary way, what one intends to speak of is what they refer to. It can also happen, however, that one wishes to talk about the words themselves *or their sense*". To see this, consider the sentence

  (1) "Mary thinks that the butler did it"

Now suppose that the butler is in fact the illegitimate son of Lord Edmund, and further that Mary doesn't know this - would one still assent to

  (1′) "Mary thinks that Lord Edmund's illegitimate son did it"?

Reasonably, no. This is the puzzle of propositional attitude reports: co-referring terms cannot be substituted *salva veritate*. According to a Fregean framework, this failure of the principle of substitutivity occurs because the embedded clause no longer refers to its customary reference (a truth-value, since it has a complete sense) but rather to the thought it expresses, which is its usual sense[15]. This is called "indirect reference" and occurs when "the subordinate clause has for its reference a thought, not a truth-value; as sense not a thought, but the sense of the words 'the thought that (etc.),' which is only a part of the thought in the entire complex sentence." Hence the words "The butler did it" designate not the truth value of the proposition they assert, but instead they designate the thought which is expressed by the sentence "The butler did it". Now it becomes clear that "The butler" and "Lord Edmund's illegitimate son" have different customary senses, since

---

[15]The meaningful use of non-referring terms in propositional attitude reports: "Mary believes the king of France is bald" is explained in the same way.

they express different thoughts, and therefore the clauses containing them will have different indirect references. Unsurprisingly, then, the compositional reference for the whole utterance is different for (1) and (1′), and we can happily assent to the truth of the former without committing ourselves to the truth of the latter. "In such cases it is not permissible to replace one expression in the subordinate clause by another having the same reference, but only by one having the same indirect reference, i.e. the same customary sense." Words refer to their sense whenever they occur in subordinate clauses, introduced by 'that', and "can also be recognised by seeing that it is indifferent to the truth of the whole whether the subordinate clause is true or false."[16] This is not to say that they are unrelated, but rather that the truth of the whole depends on the sense of the subordinate clause, and not its reference (which is a truth value).

Introducing this notion of indirect reference is a neat way to capture what is grasped when statements of, for example, false belief, are understood. This suggests that the research done on deaf children's understanding of false belief may benefit from a Fregean approach to the semantics of belief and other mental state sentences. I think Frege is helpful here because his framework enables a finer-grained description of the functioning of complement sentences. This is something which is clearly called for. De Villiers and de Villiers (2000) observe that "the structures necessary in language to represent desire do not typically involved a full, tensed *that*-clause in English, ... but rather a *to*-clause that represent an event that is "irrealis" rather than true/false". Mental verbs of belief *do* customarily take a tensed *that*-clause[17], and more importantly, in these cases, the embedded clause

---

[16]This is true only for "de dicto" attitude reports: cases with main verbs like "see that", "remember that", where the truth of the whole *does* rest on the truth value of the subordinate clause. It also rules out (for a different reason) sentences like "Whoever discovered the elliptic form of the planetary orbits died in misery", where the sense of the subordinate clause is not complete and the clause gets its customary reference (i.e. the referent of "whoever discovered the elliptic form of the planetary orbits", Kepler)

[17]And note however, that verbs for pretence usually also take a tensed *that*-complement - exactly the same as the structures for mental verbs of belief.

has its indirect reference, in Fregean terms. But this observation does not hold cross-linguistically - for instance in Dutch, the above mentioned case of desire verbs taking a *to*-clause is no longer applicable - desire verbs also take a *that*-clause in this case. For example, the English "I want you *to* tidy up your room" is "Ik wil *dat* je je kamer opruimt" in Dutch. Even within English there are desire verbs which take a full tensed *that*-clause: 'hopes that', 'wishes that', 'imagines that', and communication verbs ('says that') certainly do too, and hence these structures should suffer the same delay as belief-sentences do, on de Villiers and de Villiers' hypothesis. There is experimental evidence (Woolley 1995) to suggest that this is in fact not the case. It has been well-documented that belief-psychology develops later than desire-psychology, and this developmental lag points to the possibility that there is something more than mastery of the syntactic structures of complement sentences which enables false-belief mastery. The interaction of this developmental lag with delayed linguistic development in the deaf children's theory of mind acquisition needs to be investigated more fully before claims of linguistic determinism can be appraised.

Again, there is philosophical work which can be put to use here, to possibly elucidate what it means to understand and produce a statement of false belief attribution. Roderick Chisholm (1957) formulated a linguistic equivalent of the mental concept of intentionality, "a working criterion" by which we can distinguish intentionality in language, and thereby re-express Brentano's thesis by reference to intentional sentences. Here these criterion serve to give a clearer picture of the interaction of language structures with the structuring of folk theories of mind.

5.3   Intentionality in language

One way to avoid the issue of "intentional in-existence" of mental states and thus also recourse to a mentalist doctrine is to consider the intentionality of the sentences we use to talk about the mental, instead of the intentionality of the mental itself. Quine would call this 'semantic ascent': recourse to a logical analysis allowing for the problems of intentionality to hopefully be

reformulated in more accessible terms and explicit criteria to be described. In particular, one can use a linguistic reformulation of Brentano's notion of intentionality - due in large part to the work of Roderick Chisholm[18]. The formulation given here is based on that outlined by M. Harney in *Intentionality, Sense and the Mind* (1984).

Now, which sentences are to be considered intentional? It seems natural to want any sentence expressing a mental state/attitude to be considered intentional, since we have described every mental state (and only mental states) as being intentional. So we need to cover sentences such as

- 'Maxi wants a chocolate', 'Maxi is thinking about the chocolate' (where the object of mental attitude is an object in the world)

- 'Maxi believes that the chocolate is in the fridge', 'Maxi hopes that the chocolate is in the fridge' (where the object of the attitude is a proposition), and

- 'Maxi remembers that the chocolate is in the fridge', 'Maxi knows that the chocolate is in the fridge', 'Maxi sees that the chocolate is in the fridge' (again the object is a proposition, but the truth or meaningfulness of the main sentence is dependent on the truth of the embedded clause).

Chisholm identifies three criteria of intentionality to cover these cases, which are:

> 1. ... A simple declarative sentence is intentional if it uses a substantival expression - a name or a description - in such a

---

[18] Note that this does not immediately imply that ontological priority is given to linguistic phenomena over mental phenomena: Chisholm himself, in correspondence, says "With respect to intentionality,... Among the central questions, ..., are these:

1. Can we explicate the intentional character of believing and of other psychological attitudes by reference to certain features of language; or

2. Must we explicate the intentional characteristics of language by reference to believing and to other psychological attitudes?

In my ... paper, I answer the first of these questions in the negative and the second in the affirmative; " (Chisholm-Sellars, 1956)

way that neither the sentence nor its contradictory implies either that there is or that there isn't anything to which the substantival expression truly applies.

2. Any noncompound sentence which contains a propositional clause is intentional provided that neither the sentence nor its contradictory implies either that the propositional clause is true or that it is false...

3. A third mark of intentionality may be described this way. Suppose there are two names of descriptions which designate the same things and that E is a sentence obtained merely by separating these two names and descriptions by means of 'is identical with'... Suppose that A is a sentence using one of those names or descriptions and that B is like A except that, where A uses the one, B uses the other. Let us say that A is intentional if the conjunction of A and E does not imply B.

The first condition on this list re-expresses the possible non-existence of the objects of thought, as a linguistic criterion for intentionality. Thus 'Maxi is thinking of a river of chocolate' is intentional because neither this sentence, nor its contradictory, 'Maxi isn't thinking of a river of chocolate', allow us to infer anything about the existence (or non-existence) of the river of chocolate[19].

The second condition is the analogue for propositional attitudes: it tells us that a sentence like 'Maxi thinks that the chocolate is in the fridge' is intentional because we cannot infer from the sentence or its negation either the truth or falsity of the embedded clause 'the chocolate is in the fridge'.

The third condition covers mental verbs like 'perceiving', 'remembering that', 'seeing that'[20], and 'knowing'. Unlike those verbs mentioned above,

---

[19]although I hope there is one.

[20]Note that 'remembering' and 'seeing' do *not* fall into this category, but are rather covered by the second criterion.

these are epistemically neutral: correctly attributing these attitudes to a subject entails the truth of the embedded clause, and moreover commits the subject to the knowledge of this clause. Maxi cannot be said to remember where the chocolate is unless it is in fact where he remembers it to be, and this non-accidentally - he can remember *that* it is there because it actually is there, and he knows this. This condition also has another feature, which is the opacity of context[21]. Suppose that the chocolate is actually a present from Maxi's mother for him. Then, given that Maxi knows where the chocolate is, can we also say that Maxi knows where his present is? No - if Maxi doesn't know that the chocolate is his present, then we cannot assent to the truth of 'Maxi knows where his present is'. Yet the expressions 'the chocolate' and 'Maxi's present' are co-referring in this context: they both designate the bar of chocolate sitting in the fridge.

This feature of these sentences was the original motivation for Frege's introduction of the notion of indirect reference and the criterion is derived directly from this notion. In fact Harney's (1984) central claim is that "it is only by appeal to Frege's notion of sense that a satisfactory notion of intentionality can be constructed". She advances this by describing advantages to be gained from a linguistic reformulation of intentionality: namely that questions about the existence (or non-existence) of objects of thought become questions about the way in which expressions pick out (or fail to pick out) some object. Secondly, 'objects of thought' sometimes occur as contents (what is thought) and as objects (what is thought about) - for example, propositions judged, and desired objects respectively. This analysis allows these to be distinguished, provided an intensional account of language is used. This in turn recognises a difference between the proposition expressed by a sentence and the truth-value of the sentence - since "propositional equivalence is not the same as material equivalence." This is where

---

[21]provided you take a narrow view of content, that is. On a narrow view, it is opaque, but if you take a wide view of content, then the context is no longer opaque: if the chocolate is Maxi's present then he also knows that his present is in the fridge, for example. For mental state verbs a narrow view of content is usually taken, as it captures better our intuitions about how mental states work . Here I assume a narrow view.

a Fregean semantics comes in. Contents are analysed as intensional objects, and for Frege intensional objects are senses.

It may prove worthwhile to use Chisholm's criteria to devise a test of children's understanding of intentionality. Although only linguistic criteria, they do capture the most puzzling aspects of the intentionality of the mental: the possible non-existence of the objects of thought (whether extensional or intensional objects), the failure of substitutivity in attributing attitudes, and the independence of correct attribution from the truth of the content so attributed (for example, correctly attributing a false belief). De Villiers and de Villiers tested only for this latter aspect. One could also test whether deaf (and normally developing) children are sensitive to the other criteria, especially the failure of substitutivity. Are they just as susceptible to wrongly gloss over this feature of certain complement structures as they are to make realist errors on false belief attribution? Are they aware that one can quite happily think of unicorns? Answering these questions elucidates exactly how delayed language development affects mental understanding.

One could go as far as to say that the child just realises the intentionality of language when they learn to attribute false beliefs. The import of this in theory of mind acquisition depends on what role you accord language in thought. Does the word give the concept, structure the concept, or just label the concept? When the child learns to answer correctly in a false-belief test, is he/she merely learning to express in language what they already know - namely that beliefs can be false; or are they learning *from* the structure of language that beliefs can be false?[22] Use of a linguistic reformulation of the mental concept of intentionality allows for a clearer picture to be drawn from empirical research, of the acquisition of the concept of intentionality, and how it affects and is affected by the linguistic structures which express it.

[22]Chandler (1989) argues for the former, de Villiers (2000) for the latter.

5.4   Use of deixis

Almost all of our subjects, when asked where the ladder was, pointed at the screen and said "there". This of course was the most natural response, especially given that the image was right in front of them, but when we asked them to say further where "there" was, they responded hesitantly - also understandably so, since the answer they had given was perfectly adequate for a reality question. But if deictic reference, is the only, or at least primary, means they had of answering, then correct response to the false-belief task suddenly becomes difficult for a completely different reason. To answer correctly, the child has to make reference to a non-actual situation, and that is not possible with deixis alone. Deictic expressions direct attention towards a referent in the actual situation and cannot be used to direct attention towards a non-actual situation, unless accompanied by some linguistic indications that the location is the referent, and not the object - the ladder. To answer correctly, the child has to point to where the ladder is not. Another way of stating this: the child has to indicate Bob's epistemic state by referring the actual situation. It is much harder to distinguish physical reality from mental entities when the primary means of referring to states of affairs is by deixis - which means relativising the utterance to the extra-linguistic *physical* context - and not purely linguistic. Tanz (1980) makes the following comment about the use of the words *this, that, here* and *there*, which unwittingly makes exactly my point: "On the one hand, the pointing may facilitate children's decoding of the general meaning of these terms as indicators of entities and locations. But, on the other hand, it means that crucial information is being provided in a second [extra-linguistic] medium, perhaps to the detriment of the children's attention to the information available in the linguistic medium."

At this point it is also of interest to go back to Frege, and examine a gap in his theory of propositional attitude semantics, namely how it applies to embedded clauses which contain demonstratives or indexicals. Recall: "the case of an abstract noun clause, introduced by 'that', .. in which we have seen the words to have their indirect reference, coincident with what

is customarily their sense" (Frege 1997). But indexicals (such as 'I', 'now') and demonstratives ('this', 'here', 'they') get their reference from the context in which they are used (often by an accompanying demonstration - such as pointing - in the case of demonstratives), and it is unclear what the sense of these expressions is. Moreover, the context-sensitivity and lack of descriptive content characteristic of these instruments would seem to indicate that the notion of sense is just not going to be appropriate here. As Barwise and Perry (1983) put it, " For the theory to work, expressions that appear in the embedded sentences in meaningful attitude reports must have a sense or "conceptual content" which the agent's mind grasps. But many expressions [indexicals, names] in such reports seem to have no such content to provide, and others seem to provide the wrong content [pronouns, and even definite descriptions]"[23]. This shows up especially in cases where the embedded clause contains demonstratives or indexicals, , and the 'sense' of the expression becomes obscure, and thus also the 'indirect reference' of the clause. In fact the use of demonstratives and indexicals in propositional attitude reports highlights the more general difficulty they pose for a Fregean approach to semantics. Basically, "Frege's basic semantic framework, ... would seem to be constitutionally incapable of recognising demonstratives - whose very essence resides in their context-sensitivity - as more than semantic epiphenomena" (Yourgrau 1990). In this case, the interpretation of demonstratives and indexicals would become a pragmatic responsibility.

For our subjects, who were using deixis, to correctly ascribe a false belief meant producing sentences such as "Bob thinks that the ladder is *there* (by the roof)". But in a sentence like this, only the referent of "there" seems necessary for the correct attribution of belief to Bob. So Frege's notion of the sense of the embedded clause becomes apparently redundant, at least in its function as the reference of the clause. One possible remedy for this is to suppose that interpretation of deictic expressions *is* primarily a

---

[23]For example consider "Today is Wednesday" - true on Wednesday, false otherwise - nothing in the utterance has changed, only the context in which it is uttered. It is unclear how this can be accounted for in Fregean terms.

pragmatic issue: demonstratives and indexicals become shorthand for the situation or object they refer to, from which the 'sense' of the expression can be recognised[24] A starting point for this approach would be the consideration of sense as that which is grasped by a linguistically competent hearer when (s)he comes to understand an utterance. Clearly, this needs to be argued more fully but the gist of it would be that there *is* a way to extend Frege's theory to include the use of demonstrative and indexical expressions.

There is one more aspect of the use of deixis which is relevant here, if only to make a suggestive comment at this stage of the investigation. Tanz proposes that learning the correct use of deixis precedes and helps to prepare for non-egocentric operations outside the sphere of language - such as is required in role-taking, perspectivising or reasoning about others' minds. This is a supplement to the direct suggestion of other agents provided by the content of mental terms. So there are two means, provided by language, by which we can come to experiences others' points of view. The first is formal: deictic terms, production and comprehension of which requires what Tanz calls 'decentering'. Consider for example, personal pronouns. "Although the child must produce *I/you, my/your*, etc with himself at the deictic center, he cannot have grasped this egocentric formula without having understood at some level that other speakers organize the system with themselves at center." But notably, deictic terms are useful only to indicate variety in physical perspectives.

---

[24]Perry (1977) argues against this possibility. When we use demonstratives we seem to use a rule which takes us from the context of utterance to the referent of the demonstrative - for instance using "today" refers to the day of utterance. This is what Perry calls the *role* of the demonstrative, and it is unchanging. What does change is the *value* of the demonstrative - the object or situation which the demonstrative picks out on each occasion of use - as "today" picks out a different day according to when it is uttered. He then goes on to argue that neither the role nor the value of the demonstrative is the same thing as its sense. But utterances containing demonstratives can express thoughts - that is, complete senses - so how does the demonstrative contribute to it? Evans (1990) offers a rebuttal to Perry and defends the idea that these expressions can be incorporated into a Fregean theory of meaning.

For non-physical/mental perspectives, a second means is used. This is, as mentioned, the much more obvious contribution of language to reasoning about others: the use of mental terms in linguistic communication. But we use mental terms primarily to talk about motives, feelings, attitudes - things which are inherently subjective and figurative - much more often than we use them to talk about our physical perspective on things. Now belief falls uncomfortably between the two stools: while it is tied to a physical perspective it also has an inherently subjective quality to it. It relies on perception yet is not fully determined by it. Thus it is not fully captured by either linguistic means of 'perspectivising', and this alone could lead to a developmental delay in correct ascription of beliefs to others, since it requires both means. Conceptually, ascribing a false belief involves the compound difficulty of taking anothers' perceptual perspective and then computing a mental attitude about this perspective.

6 Lexical items: modal terms and mental state verbs

If the hypothesis that it is the development of theory of mindwhich enables performance on the false-belief task (as posited by theorists such as Perner) is correct, then this development should surely be evident in other capabilities which emerge concomitantly, around the age of 4 years old. And indeed, there are preliminary indications that the related concept of epistemic modality, which indicates degree of speaker certainty, is acquired around the same time, as the following studies evince.

Linguists standardly make a distinction between deontic and epistemic modalities. Most modal verbs can be used in both readings: as in the following examples (Papafragou 2001), where the first use is deontic, the second epistemic:

may The candidates *may* leave the room now.
   You *may* be wondering why I disappeared.

must She *must* find a way to help her son.

You *must* be tired.

should After such a scandal, the mayor *should* resign.

Looking for Hamlet? It *should* be on the top shelf.

Papafragou groups deontic modality with other non-epistemic readings involving ability and intention and refers to these together as *root* modality. She further reports that epistemic readings of modality emerge much later in language than root interpretations: "in language after language, the first modal notions to be expressed in children's speech are (physical) ability and volition." Desire verbs (*wanna*, *needta*) emerge about the same time as these root uses of modals.

Order of acquisition of different modal verbs is also relevant here: Papafragou reports that in English the first modals to appear are *can* and *will* (between 1;0 and 2;6) while *should* only appears after 2;6, and epistemic interpretations of *should* "have negligible frequencies till about the fourth year, or even later". Similar acquisition sequences have been observed cross-linguistically. This is evidenced in work done by Harris et al (1996) where the percentage of children using the word *should* in a why-question about a character's action rose from 12.5% in the 3-year-old group, to 51.2% in the 4-year-old group[25]. Indeed, Harris et al comment that "the main difference between two age groups was that the 4-year-olds had converged on two linguistic formulae... (*didn't* and *should*) whereas the 3-year-olds used a more heterogeneous set of formulae" - the main formulae being *didn't* (37.5%), *has to* (21.9%, not used by any of the 4-year-olds) and *need* (18.8%, has

---

[25]The question was asked after the children had heard a story in which the protagonist chose an option that would play a causal role the story outcome, which was a "minor mishap", and rejected an outcome that would have prevented this mishap from occurring. They were then asked why this outcome had occurred. For example, the protagonist chose to use a black pen, instead of a pencil or a blue pen, and this resulted in the protagonist getting ink on their fingers. Children's replies were examined for reference to the rejected option, and then classified according to how they referred to it: for instance, "She should have used a pencil", or "She needs to use a pencil".

also disappeared in 4-year-old group). These figures indicate that the use of deontic modal verbs emerges first and is gradually replaced by more appropriate epistemic modals.

Papafragou proposes that the acquisition of epistemic modal concepts indicates a fully formed theory of mind in the speaker, hence reversing the direction of determinism in the langauge-cognition relationship, in this area. As she sees it, "the successful use and comprehension of epistemic modal operators involves actively considering one's beliefs as representations of reality, as well as assessing their adequacy and accuracy as representational means. These processes jointly presuppose a representational theory of mind."

A study by Moore, Pure and Furrow (1990) investigated the correlation between comprehension of the relative 'strength' of modals (that is the amount of speaker certainty expressed by them) and performance on false-belief tasks. They conducted two experiments: the first to assess the development of the understanding of modal terms as expressions of degrees of speaker certainty (that is, epistemic interpretation of the terms), and the second to investigate the relation between competence with modal terms and competence with mental terms, and other aspects of the nature of belief.

In the first task, they used modal verbs *must*, *might* and *could* and modal adjuncts *probably*, *possibly* and *maybe* in a variation on the changed location task: the child was told she would have to try and find a sweet hidden in one of two boxes. Two statements, differing only in modal terms employed and locations indicated ("It must be in the blue box" and "It might be in the red box" for example) were voiced by two different puppets, and the child was then asked to pick a box, on the basis of what the puppets had said. Children were scored according to whether or not their choices accorded with the most common choices in the whole study. These choices gave overall preference to *must* and *probably*. Further, *could* was chosen over *might*, and *maybe* over *possibly*. Results showed main effects for age in the *must-might* and *must-could* contrasts, as well as for the *probably-possibly*

and *probably-maybe* contrasts. There was no significant effect for age in the *might-could* and *possibly-maybe* contrasts. 3-year-olds did very badly on all tasks, and between 3- and 4-year-olds, significant improvements were observed for the *must-might*, *must-could* and *probably-maybe* contrasts. Results for the *probably-possibly* contrast improved significantly only between the 5- and 6-year-old groups. Moore et al conclude that "differentiation of modal terms with respect to the expression of degrees of certainty starts to develop at about 4 years." They further remark that this falls within the window of development for mental terms, which has been reported to be the years from 3 to 5, and that "this correspondence implies that it is not just particular lexical distinctions that are developing during this period, but an underlying understanding of the mental state concept of relative certainty."

To render their results more than suggestive in this regard, Moore et al conducted a second experiment to test correspondence within the same subjects. The experimental set-up was similar to that described above, except a task testing understanding of the contrast between *know* and *think*[26] was used, and compared with competence on *must-might* contrast. Performance on the two tasks was highly correlated, and performance on the mental term task was again highly correlated with that on a standard false-belief task (the unexpected objects version), even when the effect of age had been partialled out.

Taken together, these results indicate that ability to attribute false beliefs emerges around the same time as recognition that beliefs can be held

---

[26]Different uses of word 'think' need to be distinguished, as it is used in various diverse ways:

- to indicate belief: "I think the chocolate is in that box"
- to express desire: "I think you should sit next to me"
- as exhortation/order: "I think it is time for bed"
- as detached consideration "I am thinking about going to Spain"

Although 'think' first emerges in child language as early as the third year (Moore et al 1990, Papafragou 2001), differentiating between these uses will only happen later.

with differing degrees of certainty, at around 4 years of age. This suggests that general understanding of the fallibility of belief underlies performance on the false-belief task. It may also facilitate sensitivity to the epistemic/deontic modality distinction. The emergence of the epistemic use of modal verbs at the very least indicates awareness of the fallibility of belief, but more work needs to be done to yield a more precise description of the sequence of emergence and its correlation with performance on traditional false-belief tasks.

7 Counterfactual reasoning

Recent research has focussed on the idea that false-belief statements are a species of counterfactual statements, and so false-belief mastery may just be part of broader ability to reason with counterfactual premises. The basic idea is this: in order to correctly answer the key question in a false-belief task, the subject must set aside what he/she knows to currently be the case, and consider a situation that does not currently obtain - namely that which forms the content of the other's false belief. Older children do this: by successfully reporting someone else's (or their own previous) mistaken belief, they refer to a non-actual or outdated state of affairs, they "invoke some counterfactual situation" (Harris and Leevers, 2000). Younger children, on the other hand, appeal to the situation which does obtain: current reality, instead of invoking some situation which doesn't obtain - and thus make typical realist errors. Given this formulation of task performance, there appears an alternative hypothesis to explain the improvement observed in false-belief attribution between the ages of 3 and 5 years of age. This hypothesis attributes the improvement to an improvement in ability to invoke counterfactual alternatives to the situation at hand: " the important step taken between 3 and 5 years... is not the discovery that the mind is a representational device, but rather the appreciation that mental states ... can be directed at situations that the child rules out as part of reality as it stands. This discovery is part of a more wide-ranging ability to think about and describe counterfactual substitutes for current reality" (Harris, 1991).

In a recent paper, Riggs, Peterson, Robinson and Mitchell (1998) explored the possibility that counterfactual reasoning is needed to succeed on false-belief tasks. On their account, false-belief reasoning is difficult because it involves consideration of a situation which is directly counter to current reality, and not because of any special complexity that consideration of the nature of belief may introduce into in reasoning processes. This they argue for by investigating correlation between performance on false-belief tasks and ability to reason with counterfactual antecedents. They conducted a series of experiments designed to compare 3 and 4 year old children's performance on standard false-belief tasks with what they call "physical state" tasks, which "required similar handling of counterfactual situations but which did not require understanding about beliefs or representations: Children were asked to report what the state of the world might be now had an earlier event not occurred". They report a strong correlation between performance on false-belief and counterfactual tasks, even when age and verbal ability have been partialled out, and conclude that these results provide evidence that younger children's difficulty with false belief is merely part of a broader difficulty with consideration of counterfactual situations.

Here I will focus on one of their experiments in particular. Their 'Experiment 4' is designed to rule out the possibility that the difficulty is tied to *conditional* reasoning in general, and not just counterfactual conditional reasoning. But Riggs et al choose the lamest version of the conditional to do this: precisely the future predictive, which is most common and emerges first. The experiment ran as follows: each subject was given two tasks. In the first task, two boxes were shown to the children: one for items with pictures on them, and another for blank items (such as blank pieces of paper and card). The experimenter was aided by a toy called 'Tidy Teddy'. Tidy Teddy put a piece of blank paper into the latter box (for blank items), and then went to sleep in the experimenter's bag. In the counterfactual version of the task, the experimenter then took the paper out again and drew on it, before asking the child which box it should go into. Once correctly placed

by the child (in the box which holds items that have pictures on them), the experimenter asked the counterfactual question: "If I had not drawn on the piece of paper, which box would it be in?". In the future hypothetical version, the experimenter asked the key question before drawing on the paper: "If I draw on this piece of paper, which box will it go into?". The modelling task was similar except the two boxes were designated for round objects and long objects respectively. Of 28 subjects, 10 answered both questions correctly, but another 15 answered correctly in only the hypothetical version of the task, whereas only two answered just the counterfactual version correctly. Riggs et al. conclude that "The future hypothetical question was significantly easier than the counterfactual question."

This isn't surprising. As mentioned above, this experimental set-up is undermined by observations offered in Bowerman (1986) and Comrie (1986). Bowerman reports that the earliest conditional forms to emerge in children's language are those with low hypotheticality (which means greater probability), such as future predictives , and counterfactuals (off the hypotheticality scale) are the latest to emerge. In his typology of conditionals Comrie reports that the most common form of conditional is the future hypothetical. This is a conditional with low hypotheticality, where the protasis is present tense and the apodosis is expressed in future tense. Yet Riggs et al reason as follows: since children do better on tasks involving future predictives, than on those involving counterfactual conditionals, the typical realist error that children make must be due to "difficulty with counterfactuality, and not with reasoning conditionally" (Riggs et al 1998). In effect, they have taken future predictives to be the control version of the conditional. But a future predictive isn't a 'pure' or default conditional any more than a counterfactual is, and indeed it is unclear how (or if) you could separate hypotheticality (or counterfactuality, in the extreme case) from conditionality anyway. In sum, it isn't really a fair fight: of course children will perform better on tasks which use statements of low hypotheticality if they master the comprehension and production of them earlier than that of counterfactual conditionals. The prevalence of future predictives (as reported by Comrie) might alone

account for the fact that children master them first. Of course, it may well be that counterfactual conditionals emerge latest because they are more difficult than other types of conditional statements, but this is a moot point and even if it weren't, it isn't enough to establish what Riggs et al conclude from this experiment. They want to use their results to suggest that it is not conditional reasoning, but consideration of counter-to-fact situations, which the children find difficult. The results of this experiment suggest only that *future hypothetical* conditionals are easier than counterfactual conditionals. The experiment could at least be improved by considering conditionals with a wider range of degrees of hypotheticality, and intersecting this with various tenses - for instance having future predictives with higher hypotheticality and conditionals phrased in past tense with unknown hypotheticality. This modification itself needs to be qualified however: Comrie (1986) reports that, cross-linguistically, increasing hypotheticality is associated with backshifting in tense, and thus new combinations of tense and hypotheticality may be difficult simply because they are unfamiliar constructions. Conditionals of this form have in fact been used in experiments by Robinson and Beck (2000) - they call them 'counternormals': future conditionals with very high hypotheticality. Details of their experiments are given below.

The situation is further complicated by other data which indicates that children can readily report counterfactual situations under certain conditions. One such condition seems to be the support of a counterfactual belief by a counterpart in reality. Wimmer and Hartl (1989) conducted a variation on the 'unexpected contents' task, and contrasted this with the traditional formulation. Recall that in a usual 'unexpected contents' tasks children are asked what they think is in a container before being shown its unexpected contents. They are then asked about their previous (mistaken) belief. In the modified version of the task, the experimenter asked children what they thought was in the container, but then opened the container to reveal that it did in fact contain the expected contents (Smarties in a Smarties box), thus confirming the child's belief. The experimenter then replaced the contents with others (pencils) while the child watched, and then asked the standard

58

test question about what they had initially thought was in the container. To perform successfully on the modified "change of state" task, the child must, in the same terms that Riggs et al specify, suppress their own knowledge of current reality and report their previous belief, whose content no longer obtains. Yet 3-year-olds performed exceptionally well on this task, and poorly on the standard task. The wording for the two tasks was exactly the same. Thus it seems much easier for children to report a mistaken belief provided the content of it has a basis in past reality: they can invoke a situation which runs counter to their current reality if that situation was once factual. What bearing does this have on the results presented above? It precludes positing difficulty in avoiding current reality as an *prima facie* explanation of realist errors. It is however compatible with Harris and Leever's qualified proposal that children find counterfactual reasoning difficult, but only in some situations - namely those in which counterfactual alternatives are for some reason not spontaneously invoked.

Harris and Leevers (2000) consider two objections to the counterfactual hypothesis. First, there is a body of evidence that children as young as 2 years old are fairly competent with counterfactual antecedents and consequences (Harris and Leevers (2000), Perner (1991), for instance)[27]. Also, there doesn't seem to be a major improvement in this ability between the ages of 3 and 4 years of age. So the developmental sequence for counterfactual reasoning ability doesn't seem to mirror that observed for false-belief mastery. Secondly, high-functioning children with autism usually fail standard false-belief tasks, but there is evidence that autistic children may readily invoke counterfactual alternatives in certain circumstances. How can this be aligned with the data offered by Riggs et al? Here the little 'certain circumstances' clause is crucial, propose Harris and Leevers. They suggest that these apparently opposing sets of findings can be reconciled once one notes

---

[27]Bowerman (1986) reports the much later emergence of counterfactual conditionals in children's speech. These two sets of findings are not necessarily incompatible, however, because Harris and Leevers (2000) are investigating the ability to reason with given counterfactual statements, while Bowerman is concerned with observing spontaneous use of specific counterfactual conditional forms.

a key procedural difference. In Riggs et al's experiments, the changes to objects which occurred in the experiments (such as blank paper being drawn on, clay balls being moulded into clay sausages) were not associated with either a positive or negative outcome and thus may not spontaneously provoke consideration of alternatives by the children. In the experiments reported by Harris and Leevers, on the other hand, the outcomes would probably be construed as negative or regrettable (a clean floor dirtied by muddy footprints, for example) and this may cue consideration of ways the situation could have been avoided - that is, counterfactual alternatives. Stated more broadly, children may make use of or refer to counterfactual alternatives to the current situation only or primarily when it is inherently called for by the situational set-up. The data from Wimmer and Hartl could fit in with this proposal if it includes situations where counterfactual alternatives are also somehow considered 'relevant'. The children watched the change of contents and thus it is possible their previous belief becomes somehow 'effective' or 'relevant' in describing this event sequence. This is a very tenuous suggestion however, and needs to be supported by further research into conditions under which children can indeed manage counterfactual situations.

There is another aspect of the experiment which needs to be queried. Riggs et al (1998) state that "In the future hypothetical version of the task, the test question was asked at the point when the paper was still blank: "If I draw on this piece of paper, which box will it go into?"," a correct response to which they claim "did not require counterfactual thinking, but did require the suppression of known reality." So the more modest conclusion - that the difficulty of counterfactuals lies not only in imagining alternative scenarios, but more specifically scenarios which directly counter reality - may still be drawn. However it is unclear to me what exactly "suppression of known reality " means here - I would assume that correct answering of the future hypothetical question is entirely compatible with consideration of current reality (namely blank paper). The reasoning required involves only the ability to predict future events, based on current reality. The children did have to resist reporting their current view, but it is yet to be seen why successful

60

answering requires them to entertain an *alternative* scenario. So the conclusions that Riggs et al draw - that it is the consideration of a counter-to-fact situation, and not just alternative situations, which flummox the child - are not justly supported by their experimental set-up. The results of this experiment are not surprising, but I think their relevance is stretched by Riggs et al, beyond their actual significance.

The counterfactual hypothesis is usually presented as an incompatible alternative to the representational hypothesis. Certainly there is an apparent clash with Perner's original account of false belief mastery. If children do badly on false-belief tasks because of a lack of ability to handle misrepresentation, there is no reason why they should not perform better on similar non-representational versions of the task (Riggs et al's 'physical-state' tasks). Perner (1991) in fact supposes counterfactual reasoning to be enabled upon acquisition of what he terms 'secondary representation' - at around 15-18 months of age. The correlations reported by Riggs et al suggest that it is not a deficit in the concept of representation, but a deficit in counterfactual reasoning ability, which is responsible for failure on standard false-belief tasks. Although this is posited as an alternative hypothesis, by Riggs et al and other researchers, this is not necessarily the case. If we focus on the intentional aspect of representation, it becomes evident that counterfactual reasoning skills may in fact rely on an understanding of the nature of the aboutness of mental states. Consider again the quote from Harris (1991): " the important step taken between 3 and 5 years... is not the discovery that the mind is a representational device, but rather the appreciation that mental states ... can be directed at situations that the child rules out as part of reality as it stands. This discovery is part of a more wide-ranging ability to think about and describe counterfactual substitutes for current reality". Harris is describing something very close to Brentano's "intentional inexistence" of the mental, specifically that the object of a mental state ascription need not exist. I think that appreciation that mental states can be directed at situations other than current reality is *part* of discovering that the mind is a representational device, which therefore can be directed at situations

61

which are not "part of reality as it stands".

A related point is that clarification needs to be given to exactly what exactly 'non-representational' means in the context of Riggs et al's experiments (1998), and thus also that no understanding of representation is required on the physical state tasks. Certainly, consideration of a counterfactual state of affairs does not entail supposing someone else considers that state of affairs to be true of the world. But it is also plausible to conjecture thus: to successfully reason with counterfactuals, one must consider a 'misrepresentation' of the world in one's own mind - in other words, posit a distinction between the represented situation (reality) and the description of a state of affairs which represents it differently at key points (the counterfactual case). Does this not require an understanding of beliefs as fallible? Or rather, as capable of misrepresenting the world?

Research in this area would benefit from closer attention from what it means to 'invoke' a counterfactual situation. The distinction between alethic and epistemic modalities used by linguists and semantics (see Papafragou 2001, Rott 1999) in particular should prove useful here, by making more precise how counterfactuals are interpreted, and exactly when and how different modalities come into play. A similar distinction that the research reported above is insensitive to, and which could prove very relevant, is that between degrees of hypotheticality and degrees of speaker certainty[28]. Comrie makes the point that "by choosing a given degree of hypotheticality within conditional constructions, the speaker *expresses* a certain degree

---

[28]Rott (1999) has discussed this distinction - between ontic and epistemic conditionals. He allows that "grammatical mood tells us something about whether a conditional is concerned with learning (changes of beliefs about the real world) or with hypothetical courses of events in the real world. Moreover, he formulates the following rule as a default to distinguish this:

> Indicative conditionals are usually to be interpreted as doxastic conditionals. Subjunctive conditionals are usually held to be interpreted as ontic conditionals.

of hypotheticality; this expressed degree of hypotheticality need not correspond to his actual belief, much less to the real world." (Comrie 1986). He gives as example the following dialogue:

> A: Are we in Bolivia now?
> B: If Brasilia is the capital of Bolivia, then we're in Bolivia.

Now if B's reply is taken to be sarcastic, and thus B knows that they are in fact in Brasilia, then: both the antecedent (protasis) and consequent (apodosis) are counterfactual, B knows this, and further, B believes that A knows at least that the antecedent is counterfactual. But, crucially, B's *utterance* leaves open whether or not the antecedent is true, and thus it does not express counterfactuality itself. It must be complemented by both A and B's belief that Brasilia isn't the capital of Bolivia, for the antecedent to get a counterfactual status, and thus for the sarcasm to work[29].

I suggest that in experiments of the sort mentioned above (where everything is uttered with utmost seriousness), the opposite obtains. The questions are posed in the subjunctive, which expresses counterfactuality, but the experimenter does not intend this to correspond any expression of his/her beliefs. Nowhere in tasks requiring counterfactual reasoning is the attitude of belief, towards the content (namely the counterfactual situation), intentionally invoked by the experimenters - but is it interpreted by the child that way? Consideration of an alternative state of affairs does not require any commitment to belief about that state of affairs. Physical-state tasks refer to counterfactual situations in the world, and not to counterfactual beliefs, but if the child does not distinguish between these two, they will fare badly on these supposed 'non-representational' tasks too. Anecdotal evidence for this proposal comes from our experimental work: one child shook his head vehemently after the first clause of the counterfactual question had been uttered - as if he hadn't heard the 'if' part at all and was responding to an assertion. This may serve as an illustration of what Leevers and Har-

---

[29]In fact, Comrie will go on to argue that conditionals are incapable of expressing the counterfactuality of a proposition (at least in English).

ris (2000) have recorded as a difficulty in testing ability in counterfactual syllogistic reasoning: getting the subjects to reason further without first accepting the truth of the premises. Leevers and Harris label the problem a pragmatic issue, a failure to grasp the experimenter's intentions that they treat the premises as hypothetical and not as actual. Maybe this difficulty has a more conceptual basis - namely the lack of understanding of the mental as representational - which in turns gives rise to the reported difficulty in interpretation. In fact this opens up another discussion. Pragmatic studies work from the assumption that a large part of the comprehension of utterances is inferring the speaker's intentions. In studies of the kind done by Leevers and Harris, and in fact any studies of counterfactual reasoning, the experimenters must communicate to the subject that the any counterfactual statements are intended not as assertions but as premises for further reasoning. If young subjects do not, or cannot infer this intention, they will not be able to succeed on the task. Inferring intentions is really a form of mental state attribution. This is an area for further exploration but will not be discussed further here, except for mentioning two directions for future research: firstly, revisiting counterfactual and false-belief comparisons but with more attention paid to the above mentioned distinction, and secondly, investigating the extent to which this distinction is observed by subjects - since this subtle interpretive effect has been neglected by experimenters, it doesn't seem unreasonable that 3 year old children have also failed to make the distinction! Harris and Leevers put it thus: poor performance on counterfactual problems "does not stem from an ability to reason with false premises, but rather it reflect confusion over the task requirements in the pragmatically anomalous reasoning context of a [counterfactual] reasoning task."

In our experiment, a further pragmatic difficulty must be taken into consideration - Dutch 'als' means both 'if' and 'when'. Experiments by Fillenbaum (1974) indicate that adults will happily drop a negation while paraphrasing disjunctive conditionals, often without even detecting the change in meaning they have brought about, in order to make the utterances 'make

sense'. University students were asked to paraphrase sentences, amongst which were disjunctions with 'normal' order reversed (for example they violated temporal entailments - "John finished and wrote the article on the weekend"), and disjunctions expressing 'perverse' conditional threats ("Don't print that or I won't sue you"). They were then asked if they saw any difference in meaning between the original sentence and their paraphrase of it. In both of these types of sentences, the majority of students changed the meaning. The most common changes related to order change, and change in temporal or causal sequence. In the latter type of sentence, the "perverse threat", the changes were overwhelmingly 'normalisations' which reversed the original meaning (for example as above to "Print that and I'll sue you") or converted it into a straight conditional ("If you print that, I'll sue you"). Further, the majority of those who changed the meaning failed to detect it even when asked explicitly. Those who did detect the changes they had made were asked "How come?". The grounds for these changes were mostly comments which claimed that "the paraphrases made things more clear and sensible,... and ... "they knew what the original sentences were trying to say and so they said it." This experiment indicates a tendency, when presented with an utterance describing events that are somehow extraordinary, to assume the utterance is flawed and is intended to convey information about ordinary events, instead of supposing it to be a faithful description of extraordinary events.

Could this tendency have been present in our 3-year-olds' interpretation of the counterfactual question, which was expressed using 'als'? If the child has a primitive process of normalising utterances, they may interpret, or re-interpret, "If Spud hadn't taken the ladder, where would it be?", to be "When Spud took the ladder, where did it go?"[30] . In this case, the realist answer which many of our subjects gave is correct. This tendency to normalise utterances to give an alethic reading of the modality is also an area for exploration, but will not be discussed further here. Even if this is the

---

[30]Future predictives with *when* emerge before those with *if* in children learning English (Bowerman 1986).

case, though, it only says *that* children misinterpret the subjunctive counterfactual, and not *why*. This strategy of normalising the utterance would only be used if it doesn't make sense to them in the first place - and so the question still stands - why not?

The evidence offered by Robinson and Beck (2000) indicates that these 'counternormals' are as easy as future conditionals with low hypotheticality. Robinson and Beck suggest that reasoning involving counterfactuality is more difficult than that using conditionals with high hypotheticality, for reasons other than just that of the general problem of inhibiting knowledge of current reality - for instance, that consideration of counterfactuals places higher demands on working memory than that placed by consideration of future predictives. They tested this by comparing children's ability to reason counterfactually about a past situation with their ability to reason counterfactually about the current situation. These tasks proved to be equally difficult, and this suggests that it is not just the influence or saliency of current reality which affects the child's ability on these tasks. In another experiment they investigated whether the difficulty extends to consideration of likely future situations - would children be more able to imagine alternatives to future reality (at least, likely aspects of it, such as which way water flows) than to that of past or current reality? The questions used in this task combine consideration of the future events, with high hypotheticality. Interestingly, no. The results indicate that the future hypothetical is easy, "even when it involved contradiction of an established norm". Children found it just as easy to answer the question "What if next time the water runs the other way [uphill]?" which has a very low likelihood, (this was controlled for: it was checked beforehand that children expect water to run downhill), as to answer "What if next time the car drives the other way?" (no implications about degree of certainty here). They also did just as badly in both tasks with counterfactual versions of the same questions. What can explain these results? They may be just an inadvertent side-effect of question wording - explicit mention of the relevant features of the alternative scenario may compensate for the low likelihood of them occurring, and thus enable the

child to entertain more easily the alternative situation. They also indicate that the difficulty may lie with the subjunctive tense.

In a similar vein, Perner (2000) has suggested that the difficulty in counterfactual reasoning tasks arises from the complexity of the parallel temporal referencing required by them. It is not enough, when reasoning with counterfactual conditionals, to consider alternative or even counter-to-reality scenarios. The use of the subjunctive tense indicates that these scenarios are to be precisely temporally synchronised with current reality. Perner argues thus: "If one thinks in possible worlds then the subjunctive remains an incomprehensible adjunct. The indicative conditional ... describes a possible world. Similarly, if the subjunctive conditional.. . describes a possible world then why do I need the subjunctive if it is just a possibility? It only starts to make sense if it is seen not just as a possibility but as a point for point alternative to real events"[31]. Consider the counterfactual question we asked: "If Spud hadn't taken the ladder, where would it be?" To answer this correctly, the child needs to go back to the point at which Spud takes the ladder, and then imagine an alternative world which differs from the real world only from that point onwards in time - maybe you could say branches from the real world only from that point onwards. So the consequences of Spud's action (or inaction in this case) need to be unravelled/tracked in parallel to the actual situation until the current moment is reached, and then the correct answer can be given by referring to the same point (after the key branching) in the alternative situation. Any difficulty with this process may in fact be a function of working memory limitations, and if so, this is not a new suggestion (see Robinson and Beck 2000, also Russell 2002, for example). But anyway Perner's point applies to counterfactual conditionals in the subjunctive mood, and whether that is precisely the territory 3-year-olds have difficulty with, remains to be established.

---

[31]Although there may be a valid point to be made here, it is obvious that Perner's possible world semantics needs to be given a far more rigorous treatment if the point is to stand!

The above discussion serves to illustrate the large and varied group of factors which may prove relevant in understanding counterfactuals, and their link to theory of mind ability. Apart from the empirical data, I have offered an array of underspecified suggestions of how the ability to reason counterfactually is related to false-belief attribution, but there is already enough to suggest further specification would be useful and fruitful. In particular, there are two aspects which look hopeful: firstly, the distinction between degrees of objective probability and degrees of speaker belief, which has been neglected thus far, and may be explicated within frameworks of deontic and epistemic logics, and secondly, the relationship between intentionality of mental representation and consideration of counterfactual situations, which remains unspecified as yet. I have suggested here that the *skill* of reasoning with counterfactual situations may require an understanding of the mind as intentional, and thus correlations observed are not necessarily indication that the original interpretation of the task results was wrong - they may just be a manifestation of the understanding of the intentionality of the mind at the level of skill, as opposed to concept, acquisition.

Among the theorists who think that the false-belief task does not have to do with inadequate conception of representational states is Reese Heitner, who thinks performance on the task requires an understanding of mental causation. But understanding causal relations is plausibly strongly linked with ability to conceive of counterfactual situations, and this provides a possible further aspect of the hypothesis which posits counterfactuality as a requirement for false-belief reasoning.

7.1   Mental causation

Heitner (1999) propounds a more modest explanation of failure on false-belief tasks: "it is both unnecessary and misleading to consider false-belief attribution to be a uniquely representational problem. Instead, false-belief attribution is better viewed as a cognitive milestone marking a more mature understanding of the *causal* relationship between perceptual exposure and belief formation". Heitner suggests that failure is an indication only of an

inadequate understanding of the causal origins of beliefs, since "an inability to correctly attribute a false belief clearly fits into a systematic pattern of children's difficulties with uniformly correlating belief states with their antecedent perceptual states". Certainly, our results could *prima facie* be taken as support for this idea. Recall the rather bizarre but consistent pattern of answering in the hide and seek task, where children first answered a firm 'no' to the question 'Did [the creche leader] see us hide?' and then promptly went on to say that the teacher thought they were in the tent. This is either striking illustration of blatant ignorance of the importance of informational access for belief formation, or the children have misinterpreted the question. Although the latter view is a broader and more flexible suggestion, the results are certainly compatible with both.

There is experimental evidence that between 3 and 5 years of age, children's understanding of the role of visual and linguistic information in belief formation develops significantly (Heitner 1999, Papafragou 2001). Papafragou reports results from a study in Turkish. Turkish has explicit evidentiality marking which makes it ideal for testing understanding of the origins of beliefs. For all past tense expressions there is a choice of two verb suffixes: one is used if the speaker was an eyewitness to the reported event/situation, and the other if the speaker only knows about the event/situation through hearsay or inference. The study indicated that Turkish children start to use these suffixes accurately from around 3;8 years, and diagnose them accurately from around 4;3 years. Acquisition of the semantics of these evidential terms would seem to index advances in the understanding of the causal nature of belief formation, since they indicate sensitivity to the various ways beliefs are formed.

Potential further support for Heitner's thesis is to be found in other areas. For instance, it is widely accepted that desires are understood earlier than beliefs (Bartsch and Wellman 1995), and further, there is evidence that value beliefs are mastered earlier than belief states about matters of

fact (Flavell et al 1990)[32]. Both desires and value beliefs have a much less immediate and obvious link to specific perceptual causes, whereas as fact beliefs are dependent on perceptual or linguistic information for their formation. If you want an apple, it is probably primarily because you are hungry, not because you saw one in the fruit bowl. If you believe apples are good for you, it is probably the result of years of parental propaganda, and not one instance of hearing or reading about it. Even the early emergence of pretence could be taken as support for the thesis, if one accepts that pretence has no direct causal connection with current reality. Consider the situation in which you pretend the banana is a telephone. The 'cause' of this pretend situation is most likely just your desire to pretend this - the actual situation, banana and all, has probably little causal influence on this, or what influence it does have is contingent and open to interpretation (for instance you think the banana looks like a telephone).

I think that Heitner's arguments are not as radical as he would have. Specifically, it seems to me that sensitivity to the causal formation of beliefs indicates understanding of the representational nature of belief, since it opens up the distinction between the world that the beliefs are about, and the beliefs themselves. Understanding beliefs as real things with causal properties - and specifically as more important in determining behaviour than the world itself - means that the child distinguishes between what the representation is about, and how it shows that thing as being. This is a key element of the notion of representation (Perner 2000, Goodman 1976).

As mentioned above, there is a relevant connection to be investigated: how does one come to understand that beliefs are caused by perceptions? By considering contrastive situations, on a counterfactual theory of causality. If we agree with Heitner, this throws up an alternative explanation of why counterfactuality and false-belief attribution are mastered contemporaneously: because counterfactual thinking enables causal thinking about

---

[32]A qualification here: value beliefs are most likely not distinguished from preferences or desires by younger children.

mental states, which in turn enables false-belief attribution. This is the topic of the next section.

7.1.1 Counterfactuality and Causality

Hume's original account of causality was in terms of regularity: we observe constant conjunction and from that (wrongly) infer a causal connection. But more recently, an alternative emerged: theorists such as Mackie and Lewis propounded accounts which describes causality in terms of *counterfactual* situations. The latter species of accounts uses the basic idea that a cause of an event can be understood as that which makes a difference to the occurrence of the event. This also requires some notion of background conditions, which one would not ordinarily want to count as a causal factor. For instance, the presence of oxygen would generally be considered a background or pre-condition for ignition of a match. But this is a context-sensitive assumption - in a laboratory where oxygen is excluded, the presence of oxygen suddenly becomes a highly relevant factor in establishing the cause of a fire. Precisely how this difference is explicated is not relevant here. For present purposes, it will suffice to consider the following analysis: we judge c to be a cause of event e, if it is the case that: if c had not occurred, then e would not have happened, in an otherwise standard context (which may include other contributing factors).

This is also a significant observation for experimental work, and in fact recent research (mostly with adult subjects) lends credence to the thesis that counterfactuality is a central element of our everyday conception of causality (Harris et al 1996). Our judgements about causation seem to involve contrastive judgements: we compare the observed events with alternative scenarios, and these alternative scenarios of course are counterfactual in some respect. It is precisely the features on which the alternative differs from observed reality, which we label 'causes'. Harris et al explore this idea further, and focus their attention on children's reasoning. This is in part motivated because "the study of causal thinking in infants and young children has been dominated by Hume's account. Yet, ... we would expect causal

thinking involving a contrast case to emerge at some point in development."
The experiments were designed to test this emergence: do younger children
engage in counterfactual thinking when they reach causal conclusions? Is
causal reasoning contrastive from the very beginning?

The results offered by Harris et al offer support for the thesis that causal
reasoning has an integral counterfactual aspect, and this is interesting be-
cause it ties together ability to reason with counterfactuality and the under-
standing of causality. Robertson and Beck (2000) have argued along these
lines, but in the opposite direction: they suggest that the key feature that
makes counterfactual reasoning more difficult than reasoning with pretend or
imaginary circumstances is the fact that the causal relations come into play
in counterfactual reasoning, and not in the latter type of situation. Causal
origins and consequences of the counterfactual situations have more rele-
vance than do those of pretend or imaginary situations[33]. So further ques-
tions concern the order of play of these two elements in judgements about
causation: Do we use counterfactual situations to identify causal agents?
Or do we first need to make causal connections before we can draw any
counterfactual inferences?

I make this seemingly tangential venture into the link between counter-
factual and causal reasoning because it may not be so tangential at all. If
failure on false-belief tasks is simple failure to appreciate the *causal* nature
of belief formation, and if we consider causal reasoning to require mastery
of counterfactuality, then the correlation described by Riggs et al (1998) be-
tween counterfactual and false-belief reasoning might also have to do with
a third factor - causal reasoning, and specifically causal reasoning about
beliefs. Younger children may fare badly on false-belief tasks because they
cannot reason counterfactually about the causes of belief: "If I hadn't seen
the chocolate being put in the fridge, where would I think it was?" for ex-
ample[34]. Here counterfactual reasoning provides the *means* by which one

---

[33]But Amsel and Smalley (2000) argue the opposite.

[34]A comment: Harris et al (1996) propose that it is easier to imagine alternative desires

reasons about beliefs, and specifically their causes - and this is different from being the umbrella capacity under which false-belief reasoning falls. A space opens up between the two capacities on the former account - one may conceivably find another way to reason about beliefs without using counterfactuality. Thus this suggestion is not the same as the hypothesis put forward by Riggs et al, and illustrates that exactly *how* counterfactual reasoning is used in false-belief tasks remains a relevant and open question.

---

than alternative beliefs, since differing desires about a given situation are more common than differing beliefs. This can explain why desire-psychology emerges before belief-psychology: because desire-contrasts are observed and understood earlier than belief-contrasts.

8 Conclusion

False beliefs involve attributing to another the attitude of belief towards a counterfactual proposition. Here I have considered various aspects of the attribution of false beliefs: firstly, that it involves recognition of the possible non-existence or falsity of objects of thought (objects are objects in the world and propositions respectively). This is captured by the concept of the intentionality of the mental. Advancement in false-belief mastery thus may indicate acquisition of the concept of intentionalty. This can also be understood as the replacement of an act-object dichotomy with a tripartate account of experience.

Secondly, studies with deaf children raise the possibility that linguistic structures play a role in false belief mastery. Using linguistic criteria devised to ascribe intentionality to sentences, one may test children's understanding of other aspects of the concept, such as the failure of substitutivity in mental state ascriptions. There are also results from the psycholinguistic literature which suggest that the ability to attribute false beliefs emerges at the same time as recognition that beliefs can be held with differing degrees of certainty. This may also indicate sensitivity to the distinction between epistemic and deontic modalities, something the literature on this subject has so far not addressed. The late emergence of false belief attribution may also have to do with the dual nature of belief: it has traceable perceptual or inferential origins but nevertheless is distinctly tied to individual perspectives and thus is subjective. Growing awareness of the former quality of belief is one of the suggested interpretations of improving performance on the false belief task.

Thirdly, false belief mastery may rely on skill in reasoning with counterfactual statements. Research suggests that these two competencies are highly correlated, but there are aspects of this proposal which need further elucidation, and which can be elucidated by discussions in semantic and philosophical literature. For example, the distinction between degree

of speaker belief and hypotheticality should prove very relevant in interpretations of counterfactual reasoning tasks. Also, skill on these tasks may be underpinned by understanding of the nature of mental states, as captured by the concept of intentionality.

I hope I have demonstrated some of the semantic, linguistic, and logical features of this standardly named psychological test, which it would prove fruitful to consider in future research. It seems that the false-belief task may indeed index competencies with mental concepts, but there is much work still to be done in establishing exactly which concepts, and how these are relevant in false-belief attribution.

References

[1] Amsel, Eric and J. David Smalley (2000). Beyond really and truly: Children's counterfactual thinking about pretend and possible worlds. In Mitchell P. and K.J. Riggs (eds) *Children's Reasoning and the mind.*

[2] Bartsch, Karen and Henry Wellman (1989). Young Children's Attribution of Action to Beliefs and Desires. In *Child Development* 60, pp 946-964.

[3] Barwise, John and John Perry (1983). *Situations and Attitudes.* MIT Press, Cambridge, Massachusetts.

[4] Bohnemeyer, J. (2001). Deixis. In *The International Encyclopedia of the Social & Behavioral Sciences* N. J. Smelser and P. B. Baltes (eds), Section 17: Linguistics (Section Editor B. Comrie), Elsevier.

[5] Bowerman, Melissa (1986). First steps in acquiring conditionals. In E. Traugott et al (eds) *On Conditionals.*

[6] Brentano, Franz C. (1973). *Psychology from an empirical standpoint*, translated by Kraus, Oskar and Linda L. McAlister, Routledge & Paul.

[7] Chandler, Michael, Anna S. Fritz and Suzanne Hala (1989). Small-Scale Deceit: Deception as a Marker of Two-, Three-, and Four-Year-Olds' Early Theories of Mind. In *Child Development* 60, pp 1263-1277.

[8] Chisholm, Roderick M. and Wilfrid Sellars (1956). Chisholm-Sellars Correspondence on Intentionality. In *Minnesota Studies in the Philosophy of Science*, Vol. II, University of Minnesota Press, pp 521-539.

[9] Chisholm, Roderick M. (1957). *Perceiving: A Philosophical Study*, Cornell University Press, Ithaca, New York.

[10] Comrie, Bernard (1986). A Typology of Conditionals. In E. Traugott et al (eds) *On Conditionals.*

[11] Cummins, Robert (1989). *Meaning and Mental Representation*, MIT Press, Cambridge.

[12] Dennett, Daniel C. with John Haugeland (1987). Intentionality. In R.L. Gregory (ed) *The Oxford Companion to the Mind*. OUP, Oxford.

[13] von Eckardt, Barbara (1999). Mental Representation. In *MIT Encyclopaedia of Cognitive Science*, MIT, Massachusetts.

[14] Evans, Gareth (1990). Understanding Demonstratives. In P. Yourgrau (ed) *Demonstratives*.

[15] Fillenbaum, Samuel (1974). Pragmatic Normalization: Further results for some conjunctive and disjunctive sentences. In *Journal of Experimental Psychology*, 102, pp 574-578.

[16] Flavell, John H., Eleanor R. Flavell, Frances L. Green, and Louis J. Moses (1990). Young Children's Understanding of Fact Beliefs versus Value Beliefs. In *Child Development* 61, pp 915-928.

[17] Føllesdal, Dagfinn (1969). Husserl's Notion of the Noema. In *Journal of Philosophy* 66, pp 680-687.

[18] Frege, Gottlob (1997). On Sense and Reference. In *The Frege Reader* translated by M. Beany, CTLI.

[19] Frege, Gottlob (1967/1918) The Thought: a logical inquiry. translated by A.M and Marcelle Quinton. Reprinted in P.F. Strawson (ed.) *Philosophical Logic*, OUP, Oxford.

[20] Garfield, J.L., C.C. Peterson and T. Perry (2001). Social Cognition, Language Acquisition and the development of the theory of mind. In *Mind & Language* Vol 16 No 5 , pp 494-541.

[21] Goodman, Nelson (1976). *Languages of Art*, Hackett Publishing, Indiana.

[22] Harney, Maurita J. (1984). *Intentionality, Sense and the Mind*, Martinus Nijhoff Publishers, The Hague.

[23] Harris, Paul L. (1991). The work of the imagination. In Andrew Whiten (ed) *Natural Theories of Mind*.

[24] Harris, Paul L., Tim German and Patrick Mills (1996). Children's use of counterfactual thinking in causal reasoning. In *Cognition* 61, pp 233-259.

[25] Harris, Paul L. and Hilary J. Leevers (2000). Reasoning from false premises. In Mitchell P. and K.J. Riggs (eds), *Children's Reasoning and the Mind.*

[26] Heitner, Reese (1999). Solving Putnam's False Belief Task: Representational Complexity or Perceptual Causality? Proceedings of the Naturalism, Evolution and Mind Conference 1999.

[27] Hollebrandse, B. (1998). On Theory of Mind and sequence of tense in Dutch. In Hollebrandse, B. (ed), *New Perspectives on Language Acquisition*, UMopp 22, Amherst, Massachusetts.

[28] Kaplan, David (1978) Dthat. Reprinted in Palle Yourgrau (ed) *Demonstratives.*

[29] Leevers, Hilary J. and Paul L. Harris (2000). Counterfactual Syllogistic Reasoning in Normal 4-Year-Olds, Children with Learning Disabilities, and Children with Autism. In *Journal of Experimental Child Psychology* 76, pp 64-87.

[30] Leslie, Alan M. and Pamela Polizzi (1998). Inhibitory processing in the false-belief task: Two conjectures. In *Developmental Science*, 1, pp 247-254.

[31] Lewis, Charles and Amanda Osborne (1990). Three-year-olds problems with false-belief: conceptual deficit or linguistic artifact? In *Child Development*, 61, pp 1514-1519.

[32] Lillard, Angeline S. and John H. Flavell (1990). Young Children's Preference for Mental State versus Behavioural Descriptions of Human Action. In *Child Development* 61, pp 731-741.

[33] Mitchell, P. and K.J. Riggs (eds) (2000). *Children's Reasoning and the Mind.* Psychology Press, East Sussex.

[34] Moore, Chris, Kiran Pure and David Furrow (1990). Children's Understanding of the Modal Expression of Speaker Certainty and Uncertainty and Its Relation to the Development of a Representational Theory of Mind. In *Child Development* 61, pp 722-730.

[35] Moses, Louis J., and John H. Flavell (1990) Inferring False Beliefs from Actions and Reactions. In *Child Development* 61, pp 929-945.

[36] Papafragou, Anna (2001) Linking Early Linguistic and Conceptual Capacities: The Role of Theory of Mind. In A. Cienki, B. Luka, and M. Smith, eds. *Conceptual and Discourse Factors in Linguistic Structure*, Standford: CSLI Publications, pp 169-184.

[37] Perner, Josef (1991) *Understanding the Representational Mind*, MIT Press, Cambridge, Massachusetts.

[38] Perner, Josef (2000) About + Belief + Counterfactual. In Mitchell P. and K.J. Riggs (eds) *Children's reasoning and the mind*.

[39] Perner, Josef., Gschaider, A., Schrofner, S., and Khberger, A. (1999). Predicting others through simulation or by theory? A method to decide. In *Mind & Language*, Vol 14 no1, pp 57-79.

[40] Perry, John (1977) Frege on Demonstratives. Reprinted in Palle Yourgrau (ed.) *Demonstratives*.

[41] Peterson, Donald M. and Kevin J. Riggs (1999) Adaptive Modelling and Mindreading. In *Mind & Language*, Vol.14 no.1, pp 80-112.

[42] Riggs, K.J., D. Peterson, E.J. Robinson and P. Mitchell (1998). Are errors in false-belief tasks symptomatic of a broader difficulty with counterfactuality? In *Cognitive Development* 13, pp 73-90.

[43] Robinson, Elizabeth J. and P. Mitchell (1995). Masking of children's early understanding of the Representational Mind: Backwards Explanation versus Prediction. In *Child Development*, 66 pp 1022-1039.

[44] Robinson, Elizabeth J. and Sarah Beck (2000) What is difficult about counterfactual reasoning? In Mitchell P. and K.J. Riggs (eds) *Children's reasoning and the mind.*

[45] Rott, Hans (1999). Moody Conditionals: Hamburgers, switches, and the tragic death of an American president. In J. Gerbrandy et al (eds.) *JFAK: Essays dedicated to Johan van Benthem on the occasion of his 50th Birthday*, Vossiuspers AUP, Amsterdam.

[46] Russell, James (2002). Cognitive Theories of Autism. In *Cognitive Deficits in Brain Disorders*, Harrison J. and A. Owen (eds), Martin Dunitz Publishing.

[47] Russell, P.A., J.A. Hosie, C.D. Gray, C. Scott, N. Hunter, J.S. Banks and M.C. Macaulay (1998). The development of theory of mind in deaf children. In *Journal of Child Psychology*, Vol 39, No 6, pp 903-910.

[48] Stenning, Keith (2002). *Seeing reason: image and language in learning to think.* In press.

[49] Surian, L. and A. Leslie (1999). Competence and performance in false belief understanding: A comparison of autistic and normal 3-year-old children. In *British Journal of Developmental Psychology*, 17, pp 141-155.

[50] Tanz, Christine (1980). *Studies in the acquisition of deictic terms*, Cambridge University Press, Oxford.

[51] Traugott, E., A. ter Meulen, J.S. Reilly, and C.A. Ferguson (eds) (1986). *On Conditionals.* Cambridge University Press, Cambridge.

[52] de Villiers, Jill G. and Peter A. de Villiers (2000). Linguistic determinism and the understanding of false beliefs. In Mitchell P. and K.J. Riggs (eds) *Children's reasoning and the mind.*

[53] de Villiers, Peter, Gill de Villiers, Kimberley Broderick, Nicole Eddy, Jennifer Friedman, Sarah Freedman, Jennie Pyers, Brenda Schick, Elizabeth Caldwell and Robert Hoffmeister (2000). The language of mental

state explanations: conceptual, semantic, and syntactic issues in normal and delayed development.

[54] Yourgrau, Palle (ed) (1990). *Demonstratives*, OUP, Oxford.

[55] Wellman, Henry (1991). From Desires to Beliefs: Acquisition of a Theory of Mind. In A. Whiten (ed) *Natural Theories of Mind*.

[56] Whiten, A. (ed) (1991). *Natural Theories of Mind*, Basil Blackwell, Oxford.

[57] Wilson, Deirdre (2000). Metarepresentation in Linguistic Communication. In D. Sperber (ed) *Metarepresentations*, OUP, Oxford.

[58] Wimmer, H., and Hartl, M. (1991). Against the Cartesian view on mind: Young children's difficulty with own false belief. In *British Journal of Developmental Psychology*, 9, pp 125-138.

[59] Wittgenstein, L. (1958). *Philosophical Investigations*, Basil Blackwell, Oxford.

[60] Wooley, Jacqueline D. (1995). Young Children's Understanding of Fictional versus Epistemic Mental Representations: Imagination and Belief. In *Child Development* 66, pp 1011-1021.

Appendix: test results

Prior to watching video:

Question	Minne (3;2)	Bambi (4;0)	Lila (3;8)	Dide (2;11)	Joris (3;5)	Finbar (3;11)	Floortje (3;11)	Guurtje (3;9)	Chaz (3;1)	Geert (2;11)	Kyra (3;3)
Do you like cookies?	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
Do you think that [the creche leader] (present) likes cookies?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Do you like coffee?	N	N	N	Y	Y	Y	Y	N	Y	When I'm big.	Y
Do you think that [the creche leader] (present) likes coffee?	Y	N	Y	Y	Y	Y	Y	Y	Y	(asks) Do you like coffee?	Y
If you were very big, would you like coffee?	N	Y	N	Y	Y	Y	Y	N	Y	Y, very much.	Y
If [the creche leader] were very small would she like lle like coffee?	N	N	N	Y	N	N	Y	Y	Y	A little bit.	Y
Do you like teriyaki?				Y	Y	Y	Y	N	Y	Y	(no answer)
Do you know what teriyaki is?				Y	N	Y	N	N	N	Y	(no answer)
What is teriyaki?				(no answer)		A kind of juice.				I don't know. That is a difficult word for me.	(no answer)
Where do you think that [the creche leader] (Not present) is?	Outside	Downstairs	Upstairs	Kitchen	Downstairs	Downstairs	Downstairs	Downstairs	Outside	Downstairs	Downstairs
Where does [the creche leader] (Not present) think that you are?	Don't know	Upstairs	Upstairs	I'm upstairs.	Debbie doesn't know that very	Upstairs	Upstairs	Upstairs	(no answer)	I'm upstairs.	(no answer)
Do you think you will come back to play tomorrow?	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y

First experiment: after at least one viewing of video. Video is paused at key moments and the following questions are asked:

Question	Minne (3;2)	Bambi (4;0)	Lila (3;8)	Dide (2;11)	Joris (3;5)	Finbar (3;11)	Floortje (3;11)	Guurtje (3;9)	Chaz (3;1)	Geert (2;11)	Kyra (3;3)
(ladder is still against the house)											
Where is the ladder now?	In the tree.	Edge of the roof.	By the house.	By Bob the Builder.	Against the roof.	Against the roof.	Against the shed.	Against the roof.	I don't know.	By the house.	(indicates only by pointing and touching TV.)
Where does Bob think that the ladder is?	In the house.	There. By the house.	By the house.	(no answer)	Bob stops.	By the tree.	Against the wall.	Against the wall.	I don't know.	By the tree.	(no answer)
(Spud takes the ladder away)											
Did Bob see that Spud took the ladder?		N	Y(corr.)	N	Y(corr.)	Y(corr.)	N	N	Y(corr.)	N	(no answer)
(ladder is no longer visible)											
(False belief question) Where does Bob think that the ladder is?	Tree.	By the tree.	Spud has it. By the tree.	By the tree.	The ladder is gone. By the tree.	By the tree.	Spud took it. (Question repeated.) He thinks that it's fallen.	Gone.	I don't know.	In the mud.	(no answer)
If Spud hadn't taken the ladder, where would the ladder be?	In the tree.	By the tree.	By the tree.	(no answer)	By the tree.	By the tree.	Fallen.	By Spud.	I don't know.	I don't know.	By Spud.

Second experiment. Once David and child are hiding in playtent the following questions are asked:

Question	Minne (3;2)	Bambi (4;0)	Lila (3;8)	Dide (2;11)	Joris (3;5)	Finbar (3;11)	Floortje (3;11)	Guurtje (3;9)	Chaz (3;1)	Geert (2;11)	Kyra (3;3)
Did [the creche leader] see us go and hide?		N	N	N	N	N	N	N	N	N	Y(corr.)
Where does [the creche leader] think that we are?	Hiding.	In the tent.	In the tent.	In the tent.	I'm in the tent.	In the tent.	Here.	Here.	Here.	In the tent.	Upstairs.
Does [the creche leader] think that we are here (in the tent), or there (on the sofa)?	Here.	Here.	Here.	Here in the tent.	(no answer)	Here	There.	Here.	Here.	Here.	In the tent.

Notes:

Y' indicates either a verbal affirmative or decisive nodding. Similarly for 'N'. (corr.) indicates that the child was explicitly corrected by the experimenter after they had answered incorrectly (on so-called control questions). A blank cell means no question was asked. Testing took place in a playroom above the main creche area, hence the reference to upstairs and downstairs. At all times (except during hide and seek), exactly one creche leader was present. .