

Meaning theory and autistic speakers

Kathrin Glüer and Peter Pagin

Abstract

Some theories of linguistic meaning, such as those of Paul Grice and David Lewis, make appeal to higher order thoughts: thoughts about thoughts. Because of this, such theories run the risk of being empirically refuted by the existence of speakers who lack, completely or to a high degree, the capacity of thinking about thoughts. Research on autism during the past 15 years provides strong evidence for the existence of such speakers. Some persons with autism have linguistic abilities that qualify them as speakers, but manifest a severely impaired capacity to understand what it is to have beliefs.

1. Introduction

Some accounts of linguistic meaning, or of what it is to speak meaningfully, appeal to higher order thoughts: thoughts about thoughts. The most typical and straightforward example is Paul Grice's (1957) account of what it is to mean_{NN} something with an utterance. In Grice's account meaning_{NN} something by an utterance involves having an intention of producing a belief (as a special case of response) in the addressee of the utterance, that is, having a second order thought. Moreover, it also involves having an intention regarding the audience's belief about that very second order thought, that is, a having fourth order thought. Grice's account therefore implies the claim

(G1) If S means_{NN} that p by utterance u , then S has a fourth order thought.

The term ‘means_{NN}’ is a technical term in Grice’s paper and in later work, but it is intended to apply to ordinary linguistic utterances, i.e. to speech acts. We can therefore ascribe a second claim to Grice:

(G2) If S asserts that p by utterance u , then S means_{NN} that p by utterance u .

Together these two claims yield the consequence that

(G3) If a speaker makes an assertion, then that speaker has a fourth order thought.

Is (G3) true? The truth of (G3), or (G1), has been disputed.¹ Clearly, if (G3) is understood as a claim about thoughts which we are aware of and have articulated explicitly in the course of performing ordinary speech acts such as assertions, then it is false. That much can be established by introspection.

However, (G3) need not be taken in that strong sense. It can be understood to claim only that we have implicit intentions, which we need not have made explicit and of which we need not be aware. Rather, as a speaker I might have to be credited with such intentions, since that would be the only way of rationalizing my action. For would I really make the utterance u , in order to assert that p , if I didn’t intend the audience to believe that p , or didn’t intend the audience to believe that I intended the audience to believe that p ? Wouldn’t a sufficient amount of reflection on my part reveal to me that those intentions or desires were really part of my reason for making the utterance? Perhaps they would. Perhaps in the normal case speakers are to be credited with such implicit but rationalizing intentions.

But even if that is true, there might still be counterexamples, even if they be abnormal. It has been suggested that in very formalized social contexts, such as military or judicial, one can give

1. For instance, Evans and McDowell (1976: xviii-xxi) appeal to phenomenology for rejecting it.

a command without any interest in obedience, or ask a question without interest in an answer, or make an assertion without interest in the beliefs of the audience. In response to this, it has been questioned whether such acts are genuine speech acts at all.²

The suggested type of counterexample concerns normal speakers in abnormal situations. A quite different type of example concerns abnormal speakers. One can imagine various kinds of abnormality, like speakers with highly deviant desires or highly deviant beliefs. The abnormality that will be in focus in this paper is that of a partial or total lack of the ability to think higher order thoughts. If a speaker is incapable of thinking any higher order thoughts at all but still is capable of performing speech acts like assertions, then that speaker is a counterexample to (G3). And if a sufficiently able speaker is capable of thinking, say, second order thoughts, but not thoughts of third order or higher, then that speaker too is a counterexample to (G3). It may even be that some able speakers are capable of thinking fourth order thoughts, but that their ability to do that is so restricted that they cannot be credited with having such thoughts, whether explicit or implicit, in connection with the performance of speech acts. These speakers, too, would be counterexamples.

Are there such speakers? Could there be such speakers? A possible stance to this latter question is to say no, there simply could not be any such speakers. First of all, one might deny that a speaker incapable of fourth order thoughts is capable of meaning_{NN} something with an utterance. That denial is perhaps not so controversial, since one can take Grice's analysis of what it is to mean_{NN} as a (stipulative) definition of the term. This can make (G1) immune to counterexamples, but immunity would not automatically extend to (G3), since the term 'assertion' isn't defined by any of the listed claims. Nonetheless, it would still be open to claim that a speaker incapable of fourth order thoughts could not assert anything at all, in "the full sense", or in "the

2. For an overview of the discussion, see Avramides (1989: 58-67).

strict sense”, of assertion. Although it is open to a philosopher to take this attitude, doing so would mean to relinquish the claim for philosophy to offer an account of the empirical phenomenon in question, in this case assertion, or speech acts, or the use of language in general. It would mean to give up theorizing for the sake of some kind of “conceptual analysis” that would differ only marginally from mere stipulation.

Therefore, the only reasonable stance is to say that, as far as conceptual possibility is concerned, there could indeed be speakers of the kind envisaged. The much more interesting question is whether there in fact are. That question is empirical. It is not straightforwardly empirical, since some conceptual care is needed in the evaluation of the evidence. Nonetheless it is basically empirical, and we are going to argue that the answer is yes: there are such speakers. More specifically, some autistic speakers are of this kind. Some autistic speakers are counterexamples to certain accounts of linguistic meaning, or of what it is to speak or mean something.³ We do not claim that there are no other counterexamples. Rather, because of experiments with autistic children carried out over the past 15 years or so, there is a good empirical basis for our claim about autistic speakers, and that is the reason for the choice of focus.⁴

Clearly, not all persons with autism are counterexamples. Some autistic patients don’t speak at all, or show such rudimentary speech behaviour that one would hesitate to regard them as language users. Others have a very well developed linguistic ability, but it is less clear that they

3. Only recently, and thanks to an anonymous referee, have we learned that the main point of our paper has been made before, by Stephen Laurence (1998:209-10). Since Laurence does not argue the matter at any length, his remarks do not make the present discussion superfluous. Still, had we known about Laurence’s paper at an early stage, our paper would probably not have been written.

4. One might think that small children, say with a mental age of around three years, could provide further counterexamples. That may be the case, but it is somewhat less clear that they have the required linguistic ability. A more recent example in the literature is speakers with RHD (right hemisphere brain damage), e.g. after stroke. Some of these speakers also seem to have an impaired capacity for higher order thoughts, but it is less clear that it is impaired enough. See Happé et al. (1999).

cannot be credited with the relevant capacity for higher order thoughts. In between those two groups, however, there is a group that, apparently, combine a sufficient degree of linguistic ability to be regarded as language users with a seriously impaired capacity for higher order thoughts, and speakers in this group are, we argue, counterexamples of the kind mentioned.⁵

The rest of the paper will be organized as follows. In section 2 we present the so-called mind-blindness hypothesis about autism. In section 3 we give a general characteristic of linguistic abilities and disabilities of autistic speakers, and in sections 4 and 5 we discuss and defend the counterexample claim. In section 6 we discuss the consequences of that claim for some theories of meaning. And in the final section, we briefly touch on the general issue of communication on primitive and on sophisticated levels.

2. Autism and mind-blindness

Research on autism has been very intense especially during the last 15-20 years. Nevertheless, autism has proved difficult both to describe and to explain. Definitions of autism have developed substantially since the term was first introduced by Leo Kanner in 1943 and a wide variety of explanations have been suggested. Today, it is widely agreed that subjects with autism have the following “triad” of impairments:

(1) Specific abnormalities of social behaviour, affecting in particular reciprocal relating and empathy,

(2) communication difficulties affecting non-verbal communication, conversational skills (pragmatics) and prosody,

5. Incidentally, some are also counterexamples to claims made by Donald Davidson (e.g. 1975: 170; 1982: 326/7; 1991:156) and others that one cannot have beliefs at all unless one has the concept of objective truth, which is to say of something being the fact irrespective of what anyone believes. According to Davidson, then, no one believes anything who does not grasp the concept of belief.

(3) lack of creativity and imagination accompanied by a characteristic rigidity and repetitiveness of behaviour (cf. Wing and Gould 1979; DSM IV; Boucher 1996).⁶

These conditions allow for an immense variety among those fulfilling them. Persons diagnosed autistic range between “people with profound mental handicap and university graduates, adults who barely have a word of expressive vocabulary and adults who read encyclopedias for recreation and speak with pedantic exactitude” (Tager-Flusberg et al. 1993). Generally, it is distinguished between low and high functioning subjects with autism. High functioning subjects with autism have both IQs in the normal range and language. Over and above the mere range of possible manifestations, it has also to be kept in mind that some of these change or even vanish with age. However, while most subjects with autism do develop some social skills these are invariably highly deviant and both quantitatively and qualitatively abnormal, even in the highest functioning individuals. These may exhibit a desire for social contact, but are typically incapable of it. Often, they are aware of their disability and develop a number of coping strategies, typically revolving around learning concrete rules for social interaction (Volkmar and Klin 1993: 43, 47).

Today, it is well-established that autism has biological causes. With the increasing evidence for organic brain damage, psychodynamic or learning theories have come to be widely rejected as *psychological* explanations of autism. They have been mainly replaced by theories seeing either cognitive or socio-affective deficits as primary. Here, we shall be concerned with a spe-

6. These conditions are also satisfied by individuals diagnosed as having “Asperger’s syndrome” (Asperger 1944; Frith 1991). However, while approximately 75% of people with the triad also have general learning difficulties (mental retardation), Asperger’s is usually not diagnosed in cases with mental retardation (Boucher 1996). Individuals with only one or two of the triad of impairments may be diagnosed as “showing autistic features” (Baron-Cohen and Bolton 1993: 16; Boucher 1996: 225) and it has been shown that non-autistic relatives of people with autism, especially their identical twins, often have either (1) or (2) (Boucher 1996: 224).

cific variant of cognitive explanation, with what is now known as the “mindblindness” or the “theory of mind” theory of autism. This was first introduced in 1985 when Baron-Cohen, Leslie, and Frith suggested that children with autism lack a theory of mind.⁷

To have a theory of mind means to be able to ascribe psychological or mental states such as beliefs and desires to oneself and others and thereby to explain and predict actions. To have these abilities, an understanding of *false belief* is essential, as perhaps first stressed by Daniel Dennett (1978). To understand the difference between *being* true and being *believed* to be true, one must understand that a belief can be false, and this understanding is manifested by means of the ability to ascribe beliefs one takes to be false. A number of well-known so-called false-belief tests have been developed in order to test for them. Since the first of these were carried out by Wimmer and Perner (1983), research has shown that normal children develop a theory of mind very early on; it is now clear that by the age of 3, children see people as possessing beliefs as well as desires (desires by the age of 2), and at 4, they are clear about beliefs explaining actions (Wellman 1993). Baron-Cohen et. al. adapted these tests to answer the question “Does the autistic child have a theory of mind?” in 1985. Their study showed that children with autism typically fail false belief tests; the following pattern emerged: 85% of normals and 86% of children with Down’s syndrome pass tests like the Sally-Anne or the smarties-test.⁸ However, only 20% of children with autism do. This pattern has been replicated numerous times, excluding practically all possibility of interpreting the results differently.⁹ Given that it had been argued in the philosophy of mind and language that a theory of mind is necessary both for social understanding and communication, it was suggested that a deficit in this area might account for at

7. However, theory of mind theories of autism have tended to move back from their original claim that lack of a theory of mind is the primary cause of autism to identifying precursors of a theory of mind as the primary cause. Thus, the difference between theory of mind theories and socio-affective theories has diminished (Boucher 1996: 228).

least two of the core symptoms of autism. Thus, Baron-Cohen concludes: “The robustness of this finding suggests that in autism there is a general inability to understand other people’s different beliefs” (Baron-Cohen 1993: 61). And this result seems to generalize to the autistic person’s understanding of their own beliefs.¹⁰

The explanatory power of the theory of mind-hypothesis is considerable. Thus, even one of its most lucid critics writes: “it goes a significant way towards explaining the specificity of characteristically autistic behaviour” (Boucher 1996: 240). Nevertheless, there are also considerable problems with this hypothesis as an explanation of autism.¹¹ It seems doubtful, for instance, that it really can explain the full range of autistic behaviours as delineated by the autistic triad.¹² And it has been argued that theory of mind deficits are on too high a developmental level to be at work when autism normally starts. Both Leslie and Baron-Cohen have therefore gone on to suggest more primitive mechanisms as the primary causes. Criticisms like these, important as they are for the theory of mind hypothesis as the supposed *explanation* of autism, do not concern us here. For our purposes it is sufficient if it provides a *true description* of autistic behaviour and psychology. It does not have to be explanatory, much less provide the full

8. In the Sally-Anne test, the child is presented with the following scenario (acted out with puppets, for instance): Sally has a basket, Anne has a box. Sally puts a marble into her basket. Then, she leaves the room. While Sally is out, Anne takes the marble from the basket and puts it into her box. Sally comes back soon after and wants to play with her marble. The question then is: where does Sally think her marble is? Or: Where will she look for it? The correct answer, “in her basket”, requires ascribing a false belief to Sally and predicting her action accordingly. 85% of normal 4-year-olds have no problems with this test.

In the Smarties-test, the child is shown a smarties tube and asked what he expects or believes to be in there. The normal and correct answer is “smarties”, of course. The experimenter then shows the child that the tube in fact contains a small pencil. He closes it again, and a third person enters the room. Now, the task consists in answering the following two questions: What will this third person think is in the tube? And: What did you yourself first think was in the tube? The correct answer, “smarties”, requires ascribing a false belief to the third person or one’s own former self.

9. See section 5, below.

explanation.¹³

More critical, it would seem, is the much discussed existence of those persons with autism that do pass the false-belief tests. But again, while clearly posing a problem for a theory of mind explanation of autism, these cases do not present one for the point we are trying to make. Autism seems to provide empirical evidence for the possibility of linguistic communication without possession (or use) of the intentional concepts, and this point is not touched by showing that some subjects with autism do eventually acquire these concepts. For our purposes, it is not necessary that no subject with autism does; it is perfectly sufficient if there are some that don't (or even some that don't have them at some stage in their development at which they are nevertheless able to communicate). And such cases are provided by those high-functioning subjects with autism that are perfectly able to be tested with theory of mind tasks but nevertheless fail.¹⁴

One last critique of the theory of mind hypothesis that should be mentioned here is the following: Boucher has argued that there is a fallacious underlying inference from the fact that subjects with autism do not show or use certain abilities to that they don't have them (Boucher 1989).¹⁵ However, whether subjects with autism have the ability of second order thought or not,

10. Both Baron-Cohen and Leslie have gone on and coupled this hypothesis with ideas from cognitive science and developed (different) modular models of the underlying cognitive architecture designed to understand and explain autism as a failure of a so-called theory of mind-mechanism (see e.g. Baron-Cohen 1995, Leslie and Roth 1993).

11. For a review, see Boucher (1996).

12. The most recent trend seems to be negative on this point. See e.g. Tager-Flusberg (1999).

13. Two current alternative psychological explanations are the *executive function* hypothesis and the *central coherence* hypothesis. See Joseph (1999) for an overview. The compatibility of the executive function hypothesis with the *descriptive* correctness of the theory of mind hypothesis will be discussed in section 4.

14. Those very high-functioning persons that do not fail, however, are interesting for us too, namely, with regard to language acquisition. It seems clear that in their case language acquisition maybe is hampered, but it is not made impossible by an initial lack of intentional concepts. Rather than posing a problem, these cases therefore even seem to strengthen our point.

if they don't use it, they would still be counterexamples in the sense we need.

In the next section, we shall give a short description of the ways in which autistic language is different or impaired. These impairments, especially as they affect the ability of engaging in normal communication, naturally form the focus of most of the psychological literature on the topic. In these contexts, the fundamentality of a theory of mind for communication is pointed out again and again, but what has received very little attention so far is the, to our minds, amazing fact that so much of the higher functioning persons with autism (literal and first-order) language is nevertheless intact. The next section deals mostly with their deficits, but our main concern is nonetheless with what they can do.

3. Linguistic abilities and disabilities of autistic speakers

In general, linguistic ability varies greatly between subjects with autism. About half of all autistic children don't develop language at all, and only half of those that do have a language rich enough to be suitable for autistic language studies. It is not completely easy to draw the line between oddities in autistic language use, such as the tendency for echolalia and prosodic deviances, and instances of lack in linguistic competence. Contrary to what was assumed earlier, however, there does not seem to be a specifically linguistic problem associated with autism. Rather, although there is a language problem, it appears to depend on more general deficits in cognition and communication. Studies carried out in the late seventies and thereafter show that high performing autistic children have little or no problem with syntax and phonology. For

15. Boucher has shown that this is actually false in the case of the absent pretend play in autistic infants. Children with autism do pretend play if prompted, they just don't do it spontaneously. The same seems to be true with regard to much of autistic language; the problems are with use. And the same might be true with regard to abilities such as mental state ascriptions; they might be "latent skills", abilities that for some deeper reason are not used.

instance, they master devices such as active-passive transformations, negation, clausal complement constructions and past tense. They clearly show generative capacity, i.e. the capacity to construct and understand new sentences formed from familiar constituents, and are overall comparable to normal control children in respect of using syntactic rules and strategies, like word order strategy, both for sentence production and for sentence comprehension (Tager-Flusberg 1985).

Primarily, high-functioning autistic speakers have problems in the areas of semantics and pragmatics. Semantic problems are typically manifested in delayed learning rather than in permanent inability (Lord 1985), and include difficulties with categorizing according to functional similarity, generalizing the use of an expression to arbitrary contexts, mastering shifts in point of view, and lexical ambiguity. But the most severe and persisting problems of the autistic speaker can be found in pragmatics.¹⁶ In general, the pattern is that the subject with autism runs into problems when some flexibility is required in interpretation, while at the same time himself often speaking in such idiosyncratic ways as to leave an ordinary interpreter unable to follow.

Even apart from questions of content, however, the subject with autism is a rather unusual communicator. Typically, the autistic speaker is less willing, or even unwilling, to initiate conversation and often does not respond at all to attempts of others. Autistic speakers are inept at maintaining a conversation. They are typically insensitive to the interlocutor's emotional state

16. Paradis (1998) provides an overview of twenty years of research about right hemisphere brain damage, concluding that there is overwhelming evidence that pragmatic competence, as opposed to semantic, syntactic, morphologic and phonologic competence, is subserved by the right brain hemisphere. Damage here result in impairments in pragmatic competence of the kind that are typical of high functioning subjects with autism. Such patients may have difficulties with figurative speech, metaphors, idioms, indirect speech acts, conveying emotional states by prosody, punchline and main point of stories, organizing discourse and using contextual information. Paradis suggests the term 'dyshyponia' - as an analogue of 'aphasia' or 'dysphasia' - for the lack of competence at inferring what is non-literally meant from contextual information.

and prone to making rude or embarrassing remarks (Frith 1989). The autistic speaker is also characterized by his lack of understanding or concern for the interpretational needs and abilities of the interlocutor. Some uses of language are purely idiosyncratic. In normal children the conversational abilities increase together with the increase in general linguistic ability, but with the autistic child this is *not* the case (Tager-Flusberg 1993).

As an interpreter, the subject with autism is inflexible and pretty much tied to comprehension of literal meaning. This naturally renders the autistic speaker a poor interpreter when a literal interpretation is misleading. Persons with autism, even adults, have marked difficulties with indirect speech acts, like e.g. indirect requests ('can you pass the salt?') (Tager-Flusberg 1993). They are in general bad or incapable at understanding non-literal speech such as sarcasm and jokes (Happé 1991), as well as simile, metaphor and irony (Happé 1991, 1993). Unsurprisingly, they have problems with idioms (Frith and Happé 1994). They are disadvantaged when it comes to using background or contextual information for understanding the relevance of what has been said, or understanding what has been indirectly conveyed (Happé 1991, Mitchell et al 1997). Children with autism seem to prefer a so-called word-order strategy over a probable-event strategy for interpreting utterances which can lead to dramatic misinterpretations of ambiguous or elliptical utterances.

Summing up, autistic speakers not only show characteristic especially prosodic oddities in their speech but most importantly have difficulties both with semantics and pragmatics. While the semantic difficulties take the form of delayed learning, and thus are typically overcome, their pragmatic difficulties stay. Even the autistic university graduate suffers from them.

4. Autistic speakers as counterexamples: linguistic capacity

With the aforementioned difficulties in mind, let's turn to the abilities of autistic speakers. What

we are interested in is the possibility that autistic speakers are counterexamples to such theories of meaning which require speakers to have higher order thoughts (HOT theories). From the material assembled so far, it seems clear to us that the first part of our case is well supported: higher functioning subjects with autism, while undeniably odd and especially pragmatically handicapped communicators, clearly are speakers, i.e. do communicate by language. However, in order to determine whether among them our counterexamples are to be found, it is necessary to relate linguistic abilities to theory of mind abilities. We have to find individuals with sufficient linguistic capacity but insufficient theory of mind ability.

Given the experimental basis for assessing theory of mind abilities, it would appear that such individuals must exist, simply because anyone who can be tested with theory of mind tasks must already possess a substantial linguistic ability. As Frith and Happé (1994: 3) point out, the verbal demands on test subjects are relatively high, which is why such tests can only be used with children from age 3 or 4 onwards. The test subjects are required to communicate their beliefs about what play figures will do or think or believe, or what they themselves—the test subjects—previously thought or believed. Communication of such beliefs is already a fairly sophisticated linguistic achievement, whether the test is passed or not. So it seems that if someone can be found that does fail the test, that person would meet our requirements.

To us, this conclusion seems largely correct. Still, it might be disputed, and if we can come up with clearer counterexamples, our case is stronger. Thanks to fairly recent literature on autism, we can. In the rest of this section we shall therefore discuss in more detail, first, the relation between linguistic abilities and theory of mind test performance, and, secondly, the relation between test performance and capacity for higher order thoughts.

During the last decade, a number of papers appeared which explicitly linked linguistic ability with theory of mind ability. Two examples, dealing with belief terms, or other cognitive expres-

sions, are Ziatas et al. (1998) and Nuyts and De Roeck (1997). Ziatas et al. indeed found that belief term ability is correlated with theory of mind ability. This is not surprising given the similarity of the capacities, and even less surprising given the similarity of the testing of the capacities. To be of more value in the present context, the linguistic capacities in question should be somewhat more general. Moreover, we would want to know something about the capacities of those who fail the tests.

Two papers by Francesca Happé, (1993) and (1995), are more valuable in this context. In (1993) Happé reports results of testing subjects both on theory of mind tasks and on linguistic tasks. The linguistic abilities tested were those of understanding simile, understanding metaphor and understanding irony. She also tested the subjects both for ordinary, first order, theory of mind ability, and for *second order* theory of mind ability. Persons with second order theory of mind ability are able to understand false beliefs about beliefs. Comparing the test results, Happé found a correlation between the two kinds of capacities. Subjects who failed the first order theory of mind task were, on average, capable of understanding simile but not of understanding metaphor (nor irony), subjects who passed first order theory of mind tests but failed second order tests did understand metaphor, but not irony, while those who passed the second order test were able to understand irony.

The subjects in this study were autistic children and a control group of children with mild learning difficulties (MLD). The tests were repeated with normal children of an average age just over 5. All of the latter passed the first order theory of mind test, and about a third passed the second order test. Here, too, there was a good correlation with understanding of metaphor and irony.

Happé presents the material as an empirical test of *Relevance Theory*, as developed by Sperber and Wilson (1986).¹⁷ But whatever the virtues of Relevance Theory, it is clear, on condition

that we can rely on the experimental results and the consequences for higher order thought capacity of the theory of mind tests, that some children who are advanced enough as speakers to understand the linguistic device of simile still are not equipped to understand the concept of *thought*. Moreover, it holds, on the same condition, that some children, including normal five year olds, capable of understanding metaphor still are not able to understand *thought about thought*.

In Happé 1995, theory of mind test results are correlated with results of a general linguistic ability test, the British Picture Vocabulary Scale (BPVS). BPVS delivers a standardized score (verbal IQ) and an age equivalent (verbal mental age, VMA) by way of applying a test procedure for receptive vocabulary (the comprehension vocabulary actually used by a person in silent reading or listening). Happé's test used a larger pool of subjects (70 autistic, 34 mentally handicapped, 70 normal) than usual, and two different first order theory of mind tests (Sally-Anne and Smarties). The result demonstrates a substantial delay in subjects with autism of developing the capacity to pass such tests, not just in chronological age, but in VMA as well. According to Happé's result normally developing children had a 50% chance of passing both tests at the VMA of 4 years, while the corresponding 50% chance among autistic children isn't reached until the VMA of 9 years and 2 months (9-2) (Happé 1995: 849). Among those normally developing children who failed the test the mean VMA was 3-8 and the highest individual VMA 6-9. Corresponding data for the autistic group was a mean VMA of 5-5 and a highest individual

17. Sperber and Wilson hold that metaphor, but not simile, requires an understanding of speaker intention. Irony is analysed as a comment on a possible (possibly instantiated) thought, which means that understanding irony would involve having a thought about a thought (the comment-intention) about a thought (the commented possible thought), i.e. having a third order thought. Happé's view is that Relevance Theory is empirically confirmed by the results. The theme of connecting the mind-blindness theory of autism with Relevance Theory of communication is also found in Happé 1991 and Frith 1989.

VMA of 11-7 (Happé 1995: Table 4). So the extreme case is an autistic speaker of verbal mental age of 11 years 7 months, who fails at least one of the tests. Moreover, all autistic subjects with a VMA below 5-6 failed at least one of the tests and a majority of the autistic children failed both.

The BPVS seems to give VMA results that agree with other tests on autistic subjects. Some results of this kind are reported by Happé (1995: 847) and later, more comprehensive study of the language profiles of autistic speakers by Jarrold et al. (1997). Here, four different standard tests are compared: a test for receptive vocabulary (BPSV), a test for comprehension of morphology and syntax, a test for active vocabulary, and tests for productive use of substantive words and of morphology/syntax. It is noted that the profiles were very “flat”. Results of the BPSV differed minimally from tests of comprehension of morphology and syntax and of productive use of morphology and syntax. Higher results were obtained in a test for productive use of single words (the Renfrew Word Finding Vocabulary Scale), but that test assesses only object naming, while the BPSV and others assess knowledge of a wider range of vocabulary items. All in all, the BPSV appears to give an accurate estimate of the verbal mental age of children with autism.

In Happé 1995, subjects were tested only for first order theory of mind. Given the outcome in Happé 1993, it obviously is to be expected that an even higher VMA is needed for passing tests of second or higher order theory of mind. The results obtained so far, however, by themselves already should make it quite clear that speaking at a pretty sophisticated linguistic level, as regards morphology, syntax and semantics (but not pragmatics) is compatible with lacking the ability for higher order thoughts.

This conclusion is further supported by results reported in Baron-Cohen 1997. Normal, mentally retarded and autistic children were tested for the ability to explain false predications by

appeal to intentions of the speaker. Typically, an explanation would be that the false statement (like saying of a cup that it is a shoe) was made as a joke or as pretend play. Only three of 15 autistic subjects were able to provide such an explanation, again indicating the lack of understanding of mental states. However, all of them passed the control tests, indicating that they did have a certain meta-linguistic understanding of linguistic utterances. They displayed understanding of the difference between what is said and what is the case, as well as of the ‘says that’ locution. Moreover, they correctly judged utterances as true or false, or correct and incorrect, thus indicating an understanding of the representational character of the utterances (Baron-Cohen 1997:177). The mean verbal mental age, by the BPSV, was 6.2 years, with a range of 4.5-8.0 years.¹⁸ Failure at the false statement explanation test at a verbal mental age between 6 and 8 is perfectly consistent with Happé’s results of correlating failure at false belief tests at these and even higher VMA:s, assuming that the underlying deficit explaining the failure is the same.

In this connection it should be emphasized that subjects with autism who fail theory of mind tests still manifest understanding of other kinds of representation. So called “false photograph” tests have been devised that parallel the false belief tests. A picture is taken, and the scene subsequently altered. The subject is then asked about the content of the picture. Not only does the autistic child typically pass tests of this kind, i.e. gives a description that fits the picture and does not fit the simultaneous real state of affairs, but even performs better than normal four year old children (Perner 1993, Leslie and Roth 1993).

Thus, there is strong evidence that some autistic subjects who apparently lack the capacity

18. The paper does not provide the mean VMA of those who failed the test, nor the highest VMA among them. However, given the data cited, the mean VMA of the failers cannot be lower than 5.75 (in which case the three passers all had a VMA of 8.0). Further, assuming that the standard deviation really was as high as 1.7, as the text has it, there must at least be one failer with a VMA as high as 7.8.

for higher order thoughts still are quite good speakers. Significantly, they have generative capacity, i.e. they can correctly make, and understand, utterances made by new sentences, and they have an understanding of truth and falsity of utterances. Denying them the title of language users is, assuming the correctness of the material, simply absurd. If the theory of mind tests warrant the conclusion that such speakers lack the capacity for higher order thoughts we do have counterexamples to theories requiring that capacity.

5. Autistic speakers as counterexamples: higher order thought capacity

Can we draw the required conclusion from the theory of mind tests?¹⁹ The inference from theory of mind test results to conclusions about higher order thought capacities is unjustified if there is at least one sufficiently plausible alternative explanation of the test results. In principle, such alternative explanations could be of two kinds: Either, there might be other factors than the lack of theory of mind capacity that explain the results. Or it might be the case that a *partial* lack of theory of mind capacity indeed explains, or is part of an explanation of, the test results, but that the remaining intact part is sufficient for the meaning theories we are concerned with.²⁰ Neither of these alternatives can be given an exhaustive treatment, perhaps not even in principle. Possible examples of alternatives of the first kind would be hypotheses like these: there are disturbing

19. As was pointed out at the end of section 2, what we need for a counterexample is not really speakers who are absolutely incapable of higher order thought. Rather, what is needed is speakers who either do lack this ability altogether, or else are incapable of making use of it in ordinary contexts of linguistic communication. They might have a theory of mind capacity that is accessible in certain circumstances but not others. What is of interest is whether or not having this capacity explains their conversational performance. In principle it would be enough that some speakers do not in fact make use of their theory of mind capacity in communicative situations, although they very well can, but at present we don't know of any way of testing such an hypothesis.

20. Which of these kinds a suggested explanation belongs to depends on where we draw the line between the main theory of mind capacity and extra abilities needed in addition for specific tasks.

peculiarities of the test situations that block otherwise intact capacities, or that induce false collateral beliefs which together with an intact theory of mind yield the wrong answer, or that reduce or distort the motivation of the test subjects.

One such alternative explanation concerns the deficiency of autistic subjects with so-called *executive function*. Roughly, executive function comprises a set of faculties and abilities, such as planning, working memory, cognitive flexibility (shifting of mental set) and inhibition of prepotent responses (an ability exercised in blocking one impulse for the sake of another). It is experimentally established that subjects with autism have executive function impairments. They perform badly at planning tasks requiring both response inhibition and working memory, such as the Tower of Hanoi disk transfer task, and similarly at cognitive flexibility tasks such as the Wisconsin Card Sorting Test (Joseph 1999). The former deficit has been suggested as the correct explanation of the failure in the false belief tests.

The reason is that “in the false belief task it is necessary to suppress one’s own true belief whilst simultaneously holding in mind the requirement to answer a question about what the protagonist will do. ... it can plausibly be argued that failing a task like false belief and its cognates may be due, at least in part, to executive difficulties” (Russell et al. 1999: 859; see also Joseph 1999). In extreme form the idea would be as follows. The autistic subject has full understanding of what it is for others to have beliefs, including false beliefs, but because of difficulties with executive function he is unable to *ascribe* a belief he knows is false. He cannot suppress his own true belief. The full explanation of the ascription failure is then provided by the executive function impairment. A less extreme alternative is that there are difficulties both with understanding of the mental and with executive function, and that it is the combination of these factors that result in the failure of the false belief tests.

No doubt this is an interesting alternative explanation. It might also be the correct explana-

tion. However, testing this would require the possibility to filter out the interference by executive dysfunction in alternative experimental setups. Finding such setups is, given the nature of the competing hypotheses, extremely difficult and seems to call for very special scenarios. However, if having a theory of mind requires understanding false belief, we cannot simply say that someone *has* a theory of mind, but cannot access or employ it whenever the belief in question is false (or false by *x*'s own light). For having a theory of mind must make at least a possible empirical difference. If "access" is permanently blocked, the subject just does not have the required theory of mind capacity.²¹

It might well be that lack of theory of mind capacity *results from* an underlying executive function deficit. In that case the theory of mind impairment does not properly *explain* the false belief test results, since the correct explanation refers to executive dysfunction, but nevertheless obtains: it remains *descriptively* true to say that there is a theory of mind impairment. And to the extent that executive dysfunction normally blocks theory of mind, this of course holds even if it were possible to filter out executive dysfunction in very special experimental setups. Thus, again, even if executive dysfunction turned out to be the right explanation of autism, this would not make a difference to our claim.²²

The executive function alternative is special in singling out the false belief aspect itself as the reason why the subject's theory of mind capacity is blocked in the test situations. Other alterna-

21. Note that this is different from the fallacious inference from simply not (normally) *using* a capacity to not having it, mentioned in section 2. Here we are concerned with a capacity (that of understanding false belief) that *cannot* be used.

Module-talk might be misleading here. The dedicated theory of mind module (if any) might be unimpaired, and access to it blocked by executive dysfunction, but that does *not* mean that the subject has a theory of mind. It only has one if the module can be properly used, not if it is irreparably damaged or blocked. (Similarly, you wouldn't say that a pocket calculator can add if the basic addition circuits are intact, but because of malfunction in other parts of the circuitry the display nonetheless gives random returns to addition task inputs.)

tive explanations would highlight other, non-central features of these situations. This is not promising, however, for saving the HOT theories of meaning. In order to save these theories by means of features that block access to theory of mind capacities, those capacities must be blocked in the test situations, but *not* in ordinary communication. This assumption is far less plausible, and also easier to deal with, than the more general and evasive assumption that access is always or in general blocked. For instance, it is rendered implausible by the fact that test results are stable across rather substantial variation in test conditions. Baron-Cohen (1993) provides a review of different theory of mind experiments, reporting that the results persist despite the following variations: using real people instead of puppets, using a computer graphic representation rather than a laboratory setting, using the child's mother as experimenter, using prompting techniques, and using different stories (Sally-Anne and Smarties). In addition,

22. The further question whether the executive function deficit can be filtered out is not easily answered. It seems to be unavoidable to have some relevant difference between what the test subject believes and what role figures believe, however hypothetical the imagined situation. Otherwise the understanding of false belief cannot be manifested. But where there is such a difference, the executive function factor, as it has been defined, is relevant. If this is right, then what remains is to assess the relative strength of the preponent impulse, or impulses, and try to establish that subjects with autism perform at the same level at tests of the same executive difficulty.

Russell et al. (1999) pursued this line in devising an alternative conflicting desire test and an alternative false photograph test. In both cases the authors believe that the executive function demands were higher, because of stronger preponent impulses. The autistic children performed worse compared with results from corresponding standard tests. The authors conclude that this is plausibly explained by the executive function deficit.

On the other hand, Mitchell et al. (1997) conclude from a modified false belief test that executive function deficits cannot explain their results (consistent with earlier results of standard tests), but here the preponent impulse is assumed to be that of selecting the articulated alternative at the expense of the unarticulated one.

It may be inferred that at present, results about the role of executive function are not fully reliable, due to the fact that pronouncements about what impulses subjects have and how strong they are remain somewhat speculative.

Baron-Cohen et al. (1986) used 15 picture stories (four frames long) for testing understanding of both mental and non-mental states and relations, with the result that subjects with autism perform well on the non-mental part but badly (at chance level) on the mental part (the differences in results at different kinds of test indicates, among other things, that the experimental situation in itself does not reduce motivation). Moreover, the results are consistent with different kinds of test, e.g. of pragmatic abilities. All this makes it extremely implausible to assume that subjects with autism are employing a theory of mind capacity in non-experimental communication that they cannot access in experimental communication settings.²³

Of course, the Gricean account does not require the speaker to attribute a false belief to the hearer, only intend the hearer to have a (possibly so far unpossessed) belief. But, to repeat, the problem is *not* that there is only a problem with attributing *false* beliefs. The problem is that understanding what it is to have a false belief is an essential part of understanding what it is to have a belief at all. It makes little sense to assume someone to understand only what it is to have true beliefs, for that does not involve an understanding of the difference between what is believed to be true and what is true.

The second kind of alternative explanation of the test results requires a different kind of discussion. Consider the following objection: the subject with autism does understand what it is to have a false belief, but has an impaired ability to use knowledge of belief states for predicting and explaining behaviour. For instance, although he understands that Sally believes that the marble is in the basket, he is unable to use this information for predicting that she will look for

23. To this can be added considerations about deception. Autistic children have been tested for their ability to deceive by means both of outright lying and of deceptive pointing. Their deceptive performance in such tasks is highly correlated with their theory of mind performance in false belief tasks. Moreover, it is consistent with records of everyday, non-experimental situations, for the absence of lying is a striking feature the behaviour of autistic children (Sodian and Frith 1993).

the marble in the basket and not in the box, where it in fact is. His prediction is based on information about the marble rather than on the information about Sally's belief. Or, it might be that the subject with autism understands what false beliefs are but not the connection between seeing and believing. So, although he sees Sally put the marble into the basket (with her eyes open), he does not infer that Sally believes that the marble is in the basket. Had he inferred that, he would, according to this view, predict that she would look in the basket rather than in the box.

Both of these explanations are far-fetched and implausible. One would need very strong reasons to sever connections like these – between belief and action or between belief and perception – which clearly are an integral part both of folk psychology and of standard philosophical conception of belief. Recent philosophical tradition has especially stressed the role of the concept of belief in belief-desire explanation of action, and thus that this role needs to be understood for a grasp of the concept of belief.

But beyond this a further point can be made. How would a speaker function who, by assumption, indeed grasps the concept of belief, and does understand what it is to have a false belief, but cannot predict or explain action on the basis of information about belief? Such a speaker cannot intend an addressee to act in a particular way *because* of being asked or told to do so, since that would involve the ability to explain the subsequent action of the addressee on the basis of his – the addressee's – belief about what has been said. Analogously, if a speaker doesn't understand the connection between belief and perception, he cannot even conceive of the addressee having a particular belief *because* of having perceived the speaker's utterance. Moreover, he certainly would be unable to understand that, or why, a speaker would intend the interpreter to have a belief because of perceiving an utterance. It is, after all, HOT theories that conceive of the grasp of connections between belief on the one hand, and action and perception on the other, as an integral part of communicative ability. Therefore, these theories cannot be

saved by separating grasp of such connections from what is essential to a grasp of the concept of belief. Even though it cannot in principle be excluded that there is some cognitive factor allowing of such separation without detriment to HOT theories of meaning, nothing has emerged to make this at all plausible.

There is, however, a final possibility to consider.²⁴ A HOT theory of meaning is unaffected by the test results if the higher order thought capacity required for passing the false belief test is *greater* than higher order thought capacity required by the particular meaning theory. More accurately: if the theory itself *predicts* that a greater capacity is needed for passing the test than for engaging in ordinary linguistic communication, then a failure to pass the test does not falsify the theory. For it is then possible that the subject has enough to manage ordinary communication, but not enough for passing the test. That this is actually the case is not implausible, for the false belief tests are administered verbally, and so the meaning theory has a bearing on interpreting the result.

Suppose theory T claims that an n th order thought is in general required for speaking and for understanding. Then consider the question ‘Where does Sally think her marble is?’. Because the cognitive term ‘think’ occurs in the sentence, introducing a propositional attitude context, it may be that the difficulty is raised one level, requiring a thought of order $n+1$ to understand the question. At least on Grice’s theory, the propositional content of an utterance is embedded in the communicative intentions and beliefs of the participants. So if the expressed proposition itself has a propositional attitude content, then that should add to the order of thought required for understanding it.

This might have been a way out, but in fact the test results for the question ‘Where does Sally think her marble is?’ match those for the question ‘Where will Sally *look for* her marble?’. And

24. This possibility was suggested to us by an anonymous referee.

even though the action verb 'look for' does indeed have an intentional component, it also has a behavioral element. Therefore, this concept can to some extent be mastered without a sophisticated understanding of its intentional component (besides, desire and intention appear better understood by persons with autism anyway). Understanding the second question should *not*, then, require more than a thought of *n*th order. This means that the subject with autism fails the test because of an incorrect prediction, not because of failing to understand the question. This false prediction does depend on a theory of mind failure, a failure which, however, is independent of the comprehension performance. Because of this independence, this false prediction, like any simple prediction of this kind, is of *second* order, and, thus, by assumption lower than (or equal to) *n*. This, in turn, is inconsistent with the claim that the test subject masters thoughts of *n*th order in ordinary linguistic communication. So this way out appears blocked. As far as we can see, then, there is no alternative to the conclusion that the false belief tests do show that there are speakers that do not have the theory of mind capacities required by HOT theories.

6. Which theories?

The paradigmatic example of a higher order thought theory of meaning is Grice's. As was noted in section 1, Grice's original account, in the case of producing belief, requires the speaker to have an intention about a belief about an intention about a belief, i.e. a fourth order thought. If it is possible to be a sufficiently competent speaker without the effective ability of thinking even second order thoughts, then clearly Grice's account is empirically refuted.

There are a couple of objections to consider. One might object, for example, that Grice does not in fact speak of an audience's belief about an intention, but of its *recognizing* an intention. This, or so the objection might run, does not concern belief, and hence there is no need for a fourth order thought.²⁵ Grice does, of course, speak of the intention to produce a belief, so at

least a second order thought is involved. On the next level, however, the Gricean formulation is indeed in terms of recognition: ‘A must intend to induce by x a belief in an audience, and he must intend his utterance to be recognised as so intended’ (Grice 1957:45). Such an objector would hold that recognition, like knowledge, or as a kind of knowledge, is not a kind of belief. Since ‘recognize’ and ‘know’ are factives, there is no such thing as understanding false recognition or false knowledge. Hence, a person with autism that fails false belief tests and therefore lacks understanding of belief can still understand recognition. So there is a discrepancy between what false belief tests detect and what is needed in the Gricean analysis.

We have three things to say in response. First, even if the objector were right, a second order thought would be required to meet the Gricean conditions, because of the requirement of an intention to produce a belief, which anyway would make it impossible for some people, and so the counterexample remains. Secondly, we don’t agree that knowledge isn’t a kind of belief, but a discussion of this issue goes far beyond the scope of the present paper.

Thirdly, however, it deserves to be pointed out that the false belief tests do have consequences for the understanding of knowledge, too. To understand knowledge, the subject needs to understand the difference between something’s *being* true and it’s being *known* to be true. Part of this difference consists in the fact that something can be true without being known to be true. Thus, part of the understanding of knowledge would have to be manifested in *withholding* some ascriptions of knowledge to others (or past self) of what is known by the subject himself to *be* true. But grasp of this difference is precisely what the test failer lacks. For when the subject makes an incorrect prediction in response to the question ‘Where will Sally look for her marble?’, this would (on the assumption that the subject understands knowledge and makes use of this understanding in the prediction) be based on the belief that Sally *knows* that the marble is in

25. This objection was pressed by a referee.

the basket. But the only basis for this ascription is the test subject's own knowledge that the marble is *in fact* in the basket. Hence, if the test can detect a failure of understanding belief, it can detect a failure of understanding knowledge, too.²⁶

It might be further objected that the false belief tasks test for beliefs about beliefs rather than intentions to get others to form beliefs, and that, therefore, the result does not immediately apply. However, the deception tasks mentioned above do test for intentions concerning beliefs and have given similar results as the false belief tests. It is true that autistic children find it much easier to understand desire than to understand belief, and likewise easier to understand intention (Russell and Hill 2001), so one might suspect that the difference between understanding belief and understanding intention be crucial in this context. But even if that is so, it doesn't help, since the speaker on Grice's account isn't simply required to understand an intention. Rather, he is required to understand a belief about an intention, which in turn is an intention to produce a belief, and this is beyond the ability of the subject in question even if he does understand intentions. So Grice's account cannot really be saved.

David Lewis and linguistic conventions

Other theories that explicitly require the existence of higher order thoughts are similarly affected. This holds for instance for David Lewis's account of meaning in terms of linguistic conventions (conventions of truthfulness – or truthfulness and trust – in abstract language *L*).

Lewis gives the following definition of *convention*:

26. One related objection here might be that we could replace the full-fledged intentional concept of belief by something thinner and more primitive, something that could be ascribed even to the person with autism that fails the false belief tests. Then plug this concept into Grice's schema, and we have an account that can survive the present objections. The only proper response to this objection is to admit the possibility. *If* such a thinner concept can be rigorously specified, and *if* it makes sense to use it in the Grice schema, and *if* the resulting theory is substantially different from Grice's original theory, then that is certainly a theory that we haven't considered. We are prepared to do so when presented with one.

A regularity R in the behaviour of members of a population P when they are agents in a recurrent situation S is a *convention* if and only if it is true that, and it is common knowledge in P that, in almost any instance of S among members of P that

- (1) almost everyone conforms to R ;
- (2) almost everyone expects almost everyone else to conform to R ;
- (3) almost everyone has approximately the same preferences regarding all possible combinations of actions;
- (4) almost everyone prefers that any one more conform to R , on condition that almost everyone conform to R ;
- (5) almost everyone would prefer that any one more conform to R' , on condition that almost everyone conform to R' ,

where R' is some possible regularity in the behaviour of members of P in S , such that almost no one in almost any instance of S among members of P could conform both to R' and to R (Lewis 1969: 78).

The ingredient in this definition that is crucial in our context is the requirement of common knowledge, defined as follows:

Let us say that it is *common knowledge* in a population P that ___ if and only if some state of affairs A holds such that:

- (1) Everyone in P has reason to believe that A holds,
- (2) A indicates to everyone in P that everyone in P has reason to believe that A holds,
- (3) A indicates to everyone in P that ___ . (Lewis 1969:56)

Again, the crucial clause in this definition is (2). Lewis explains what it is to indicate something to someone: “Let us say that A indicates to someone x that ___ if and only if, if x had reason to believe that A held, x would thereby have reason to believe that ___ . What A indicates to x will depend, therefore, on x ’s inductive standards and background information” (Lewis 1969: 53).

Lewis's idea here is that when the three clauses of the common knowledge definition are satisfied, then, given "ancillary premises regarding our rationality, inductive standards, and background information" (1969: 53), higher order expectations will be justified. Higher order expectations (i.e. beliefs) are expectations about other people's expectations, including expectations about other people's higher order expectations. The existence of such higher order expectations is, on Lewis's view, what maintains a convention: I expect you to act in a particular way, i.e. conform to regularity *R* (like keeping to the right hand side of the road), because I expect you to expect me to conform and expect you to prefer to conform if I conform. Moreover, I expect you to expect me to conform because you expect me to expect you to conform and believe that I prefer to conform if you conform. And so on. Higher order expectations are the heart of conventions. And higher order expectations are what our proposed autistic speakers lack.²⁷

As these examples illustrate, given the evidence from autism, the demand for higher order thoughts poses a problem for meaning theories.²⁸ However, it is in principle possible to distinguish between the following two features of a philosophical account of meaning: what goes into the explanation of what language or linguistic meaning is, and what is actually required of speakers. Thus, it could in principle be the case that an appeal to higher order thoughts is an essential part of the right account of meaning, even though it is not required that speakers actually have a capacity for higher order thoughts.

Still, it should be obvious that such a distinction is not available for theories that define meaning something in terms of actually having higher order thoughts, such as Grice's. Neither is it available to theories like Lewis's that make some other phenomenon (convention) constitutive of meaning, where that phenomenon requires people to actually have higher order thoughts. On Lewis's account, a community of individuals lacking the capacity of higher order thought

cannot have any conventions. Saying that speakers do not actually have to have the capacity for higher order thought thus means allowing conventionless linguistic communities. Given that convention is supposed to be what constitutes the connection between the population and the language itself (construed as an abstract entity), i.e. the actual language relation, this means giving up that account of the actual language relation.²⁹

Donald Davidson and the principle of charity

A different candidate for separating the account of meaning from the account of speaker psychology is Donald Davidson's theory of meaning in terms of radical interpretation.³⁰ In the present context we shall simply leave aside that part of Davidson's account which concerns the character of meaning theories for individual languages/speakers: that they be compositional, extensional, truth theoretic. Here we are concerned with the other part, which deals with the

27. Lewis's definition does not explicitly require the existence of higher order thoughts. It only requires that higher order expectations be justified. However, what I am justified to believe, i.e. have reason to believe, does, on Lewis's view, depend on my rationality, inductive standards and background information. Therefore, an autistic speaker will not have the same reasons as a more sophisticated speaker. In particular, if he is unable to form beliefs about beliefs, then, it seems, no state of affairs *A* will give him reason to believe, and thus will not *indicate* to him, that everyone in the population *P* has reason to believe that *A* holds. So the presence in the population of a sufficiently competent autistic speaker lacking higher order thought ability, would have the effect that condition (2) of the common knowledge definition would not be satisfied. In particular, the common knowledge clauses in the definition of *convention* cannot be satisfied under these conditions, for there the members of the population would need to have the ability to understand that other members have reason to believe that almost everyone expects almost everyone else to conform to *R*, etc.

In fact, the decisive factor for generating higher order expectations is the background belief that others share my rationality, inductive standards and background information. This would allow me to infer that *A* indicates to everyone in *P*, not only that everyone in *P* has a reason to believe that *A* holds, but also that everyone in *P* has a reason to believe that everyone in *P* has a reason to believe that *A* holds. And so on. If I don't have the background belief about sharing rationality, inductive standards and background belief, then it may well be that condition (2) is met, and yet at most second order expectations are generated.

question of the conditions under which such a theory is correct. Davidson's answer to this question is that a theory is correct just if it would come out as acceptable in a methodologically sound *radical interpretation* of a speaker of that language. An interpretation is radical in case it proceeds without any background knowledge or assumptions about what the speaker believes or desires or means by his words, and it is methodologically sound if it proceeds by applying the *principle of charity*.

The principle of charity was originally identical with the maxim of interpreting a speaker so as to maximize the number of *true beliefs* of the speaker. Since a belief is judged true by an interpreter insofar as it agrees with the interpreter's beliefs, in practice this maxim boils down to an endeavour to maximize agreement between speaker and interpreter. The original motivation (inspired by reflections of W.V. Quine on translating logical connectives (1960: 57-61)) was that if a speaker is attributed false beliefs beyond a certain point, it is more plausible that there is something wrong with the interpretation than that the speaker really has these beliefs. In order for me to have some false beliefs about something, say the moon, it is necessary that I have a sufficient number of true beliefs about the moon, or otherwise it isn't reasonable to say that my

28. Other accounts that explicitly appeal to higher order thoughts in the Gricean tradition are found in Shwayder (1965), Schiffer (1972) and Bennett (1976). Furthermore, appeal to higher order thoughts is a standard feature in theories of communicative intentions (a part of speech act theory). For a good presentation, see Recanati (1987). If we are right, this field virtually as a whole faces serious problems.

29. It might be objected that convention does not require that each and every member of a community has the required higher order thought capacities. Some members could be allowed a kind of parasitic membership, while core members still would need the capacities. Most probably, there are no empirical counterexamples against this proposal, since most probably there are no linguistic communities consisting only of subjects with autism (or toddlers, or right hemisphere brain damage patients). Maybe it is even empirically correct that a community of only persons with autism could not develop language from a pre-linguistic state. But if it is possible that only autistic members survive, and that they continue to use language by themselves, then the lack of counterexamples is a mere contingency.

30. See (Davidson 1984) for the classic collection of articles presenting his ideas on theories of meaning and radical interpretation.

beliefs are about the moon at all. It does not, for instance, make good sense to ascribe to me the false belief that the *moon* weighs less than a kilogram when my assumed background belief is that the moon is a part of my right shoe, not a heavenly body. Thus, part of what the principle of charity requires, is that beliefs not be attributed to a speaker unless they are minimally reasonable.

The idea that beliefs be reasonable led over from the simple idea of maximizing true beliefs to a more refined and complex notion of *optimizing* belief attributions. In optimizing account is taken of the fact that some beliefs, like observation beliefs, are more basic and thus more important to get right than others and that errors of judgement must be explainable. It is not only a question of number, or frequency, but also a matter of which false beliefs (and which true beliefs) can reasonably be attributed.

As the term has come to be used, moreover, the principle of charity does not only concern truth and falsity of beliefs, but a wider evaluation of speakers' attitudes and actions, and should be seen as denoting rather a collection of interpretational maxims. Speakers should be interpreted as having largely coherent belief systems, pretty stable and coherent preference orderings, and as acting in accordance with their beliefs and desires. There is no canonical list of what goes into that collection.

What makes the principle of charity relevant for interpretation is a fundamental but rather uncontroversial idea: if you have a particular attitude to a sentence, i.e. to a linguistic expression, and that sentence has a particular meaning relevant to that attitude, then you have a corresponding (possibly) non-linguistic attitude. For instance, if you *hold true* the sentence 'gold is yellow', and that sentence in your idiolect *means that* gold is yellow, then you *believe* that gold is yellow. Similarly for *desiring true*. Because of this connection we can conclude that a sentence *s* does not *mean that* the moon weighs less than a kilogram, in a speaker's idiolect, if that

speaker holds *s* true but cannot acceptably be attributed the belief that the moon weighs less than a kilogram.

Thus far the role of charity. Now, it seems clear that an autistic speaker who fails theory of mind tests, and presumably lack a capacity for higher order thoughts, could not engage in methodologically proper radical interpretation. Applying the principle of charity explicitly involves judging the reasonableness of beliefs, which does require higher order thought capacity. So that much seems out of the question.

We could try to bypass that problem by reformulating the principle of charity itself in terms only of attitudes to sentences. Thus we could require the interpreter to try to maximize, or optimize, truth among sentences held true by the speaker, without any explicit mention of belief. There is reason to think that the attitude of holding true, or at least an attitude of holding true that is revealed by means of an utterance, is easier to understand for the autistic speaker. At least it would seem so, since autistic speakers can evaluate other speakers' utterances. However, these evaluations are performed when the autistic interpreter already has mastered a scheme of literal interpretation. When the interpretation aims at *assigning* meaning to linguistic expressions, rather than just assigning meaning to utterances made by already understood expressions, the difference is less significant. The autistic interpreter is then supposed to pair an interpretation with an utterance on the basis of – among other things – what it is reasonable to attribute to the speaker in the context. Thus, in this case, just as with the false belief tests, the thought content of the speaker (or the Sally character) is something that must be inferred on indirect evidence. The interpreter does not know what the sentence means that the speaker holds true, just as he doesn't know what the speaker believes, but is asked to find out by means of having an idea of what it would be *reasonable to think* the speaker believes, or what the sentence held true means, given the circumstances. The task is equivalent, whether explicitly described in terms of

belief or not.

This conclusion is further supported by empirical evidence concerning the lack of interpretational abilities of autistic speakers, in particular their lack of capacity for flexible and creative interpretation. Interpreting by means of charity is simply beyond the limits of the autistic speakers we have drawn attention to.

The question is whether this matters. Neither Davidson's theory of the form of an explicit meaning theory about a speaker or a language, nor the principles he proposes for proper radical interpretation, is intended as providing a psychological model of actual interpreters. The principle of charity serves to determine which meaning theory (or theories) is (are) the right one(s), whatever the mechanisms involved in actual interpretation. Because of this role the principle of charity has within Davidson's account, the lack of higher order thought capacity of some autistic speakers does *not* constitute a counterexample to the account.

Still, the issue can and should be pushed further. Regarding empirical evidence, there is a more general question to be asked: if the interpretational mechanisms of actual speakers might be (and perhaps in general are) different from interpreting by way of charity, what reasons do we have for thinking that it is the principle of charity that determines what is the *right* meaning theory? After all, it is *prima facie* conceivable – whether or not plausible – that if I performed a methodologically proper radical interpretation of you, the result would differ from and rule out my habitual scheme of interpretation. Would there, then, be reason for saying that the error lies with my previous interpretation rather than with the principle of charity? What is at stake here, is the general status of the principle in its relation to empirical data: can there really be counter-evidence? It is often claimed that the principle of charity is *constitutive* of meaning, and Davidson himself has given hints in this direction (see, for instance, 1970: 220-3; 1974: 236/7; 1985: 92). Exactly what this amounts to is not so easy to say, but it appears at least to be part of the

common conception that constitutive principles do not have counterexamples.

And it is here that the autistic speakers do come in again. Suppose that the principle of charity is correct, either in providing an adequate model of the performance of normal interpreters, or in only determining meaning in accordance with the interpretations of normal interpreters (thus providing an adequate model of interpretational competence). Does the existence of autistic speakers provide any empirical counterevidence to the theory? The immediate answer is no. If normal speakers speak and interpret by or in accordance with charity, and if normal speakers can successfully interpret and be interpreted by autistic speakers, then autistic speakers too interpret in accordance with charity (as far as normal and literal language use is concerned), and speak so that they are successfully interpretable by way of charity. If charity applies to the rest of us, and we successfully communicate with autistic speakers, then it applies to the autistic speakers, too. There does not seem to be any way around that conclusion.

A further point can be added. The assumption that autistic speakers interpret in accordance with charity, together with the assumption that their interpretation is pretty mechanic and stereotypical, indicates that when it comes to language learning the autistic learner simply, automatically and permanently accepts everything the teacher says. In fact, in actual teaching of lexical meaning, say the meaning of the word *chair*, teachers hardly ever make mistakes, since teachers tend to use clear and obvious examples. So interpreting by way of maximizing the truth of what the teacher says evidently pays off. And if you mechanically interpret in accordance with that principle you get it right, too, even if you are unable to understand the principle itself.

Divergence is possible, though. If the teacher does make a mistake, the flexible learner is able to later revise his earlier interpretation in the light of new evidence and therefore able to get the meaning right by discounting or even explaining away the teacher's problematic application. The mechanic interpreter isn't, and will get it wrong, or simply be confused. But as long

application mistakes aren't made in teaching, this will not happen.

7. The simple and the sophisticated

We argued in the previous section that if normal speakers interpret by way of applying, explicitly or tacitly, the principle of charity, then autistic speakers interpret in accordance with that principle. That is, the resulting interpretations are just the ones that a speaker would arrive at who explicitly (and correctly) applied the principle. You need not have any inkling of the principle itself, or of why you get it right, nor even have any idea of what it is to make an interpretation or that interpretations can be correct and incorrect. What matters in very basic interpretation, especially in early language learning, is simply that you do interpret, and that you get it right.

We believe that this is the clue to understanding how it is possible for psychologically unsophisticated individuals to learn language at all, and perhaps to understanding how it is possible in general, at least if we are all initially unsophisticated. Consider the example of making and reacting to a pointing gesture. When I direct your attention to an object by pointing I am aware of using my own body as a signpost. I expect you to understand this, and to try to locate the object of my attention along an imagined axis going out from my shoulder and passing through my index finger. Since I expect you to attempt this, I try to position my arm and index finger so that the projected axis would indeed hit the object. Moreover, I expect you to assume that this is the case, and that this be your reason for using that method of locating the object. If I were even more sophisticated I would make this very assumption my own reason for using this attention directing method in the first place. And so on. There is no end to how sophisticated our understanding can be of even such a simple communicative device.

However, a high degree of sophistication is not necessary for the device to work. We don't

assume that the very young child is aware of using its own body as a signpost, and much less of the fact that the parent will search along a projected axis passing through its index finger. Nor do we expect the child to imagine a line in reacting to the pointing. Rather, we expect that the attention of the child is automatically directed, at least if the child has been through some minimal training by precedents. And if it is automatically directed to the intended object, the device has worked. The child doesn't have to understand how or why it worked. In fact, neither does the parent.

Sophistication is plausibly developed only when basic communicative functions are already in place. When you have it, you also have some understanding of how simple communicative devices work. From a practical point of view, however, sophistication also provides something of greater importance, viz. flexibility. If I know that you are a pragmatically sophisticated communicator, then I can make use of that. I can count on your recognizing my attempt to communicate, and I can count on your interpretational capacity for understanding a gesture that is new to us both. Since I know that you know that I know that you will not understand a new gesture unless it is perceptually similar to an action that is relevant in the context, and since I know that we share judgements about relevance and perceptual similarity, I count on your limiting your search space to possible interpretations that you think I think you would manage to find. Sometimes communication succeeds just because of such understanding, and this kind of success is reserved for the sophisticated. Sophistication, however, should not be projected back onto the elementary.

In a sense, this is precisely what Grice did. In an effort to separate out "non-natural" meaning from mere causal regularities Grice correctly turned to the intention to communicate, but his model of communication required too much sophistication of the participants. However, even if this is right, we seem to have a problem. For if the intention to communicate is what character-

izes, or part of what characterizes, meaning something in the “non-natural” way, including the linguistic way, how can the autistic speaker who lacks higher order thought capacity be counted among those who non-naturally mean something at all? How can the subject with autism be ascribed the intention to communicate? After all, the subject with autism apparently does not have the required concept of communication.

In the literature we do find testimonies to the *desire* of autistic speakers to communicate, including clear, if somewhat unusual or clumsy attempts to initiate conversation. Are these ascriptions of desire and intention to communicate merely erroneous?³¹ We don’t think so. True, it is overwhelmingly plausible to say that the subject with autism lacks the concept of communication that the sophisticated speaker has, not to speak of the philosopher’s concept. Still, as regards the range of communicative situations that the subject with autism can belong in, it is equally plausible to say that he is able to recognize them on the basis of perceptual similarity. He may lack the understanding of what functionally distinguishes communicative situations from others, but as long as he is able to distinguish them by their perceptual properties, he does have some conception of communication. And if that is right, he can surely desire and intend to create new situations of that perceptual kind. We use our term ‘communication’ in our attributions for specifying the *content* of the desire or the intention of the autistic subject, and that is perfectly appropriate, to the extent that the correct theory of the semantics of attitude contexts allows it. That is, it is appropriate if the attributer is allowed to use his own vocabulary within the attitude context to capture distinctions that the attributee can make in the relevant

31. The question “Why would people bother to talk if they weren’t also mind-readers, interested in informing others and learning about the information others have?” is rhetorically asked by Simon Baron-Cohen in (1995), p 132, just after having claimed (p 131) that “... by itself, unless it is linked up to mind-reading system, the language faculty may hardly be used—at least, not socially”. This view is surprising, considering the evidence presented in the literature, but perhaps less surprising considering the stress on deficits at the expense of abilities.

range of situations, even if he cannot make them in the same general and sophisticated way as the attributer. But that is a question for the semantics of attitude attributions.³²

Department of philosophy, Humboldt university of Berlin

Department of philosophy, Stockholm University

References

Asperger, H. 1944: “Autistic psychopathy” in childhood, translated from German in U Frith (ed), *Autism and Asperger syndrome*, Cambridge University Press, Cambridge 1991.

Avramides, A. 1989: *Meaning and Mind: An Examination of a Gricean Account of Language*, MIT Press, Cambridge.

32. The main ideas of this paper have been presented at the Swedish national conference of philosophy, Filosofidagarna, Gothenburg, June 1999, and at a seminar at the Department of philosophy at Stockholm University, March 2000. We are grateful for comments at these events, especially to Staffan Carlshamre, Per Martin-Löf and Dag Prawitz. We also received helpful comments on an earlier draft from Ingar Brinck and Jan Willner. We Above all, however, we are indebted to Francesca Happé at the Institute of Psychiatry at Kings College, London. She not only provided extremely valuable comments on an earlier draft, but also helped us along the way with hints about the literature.

- Baltaxe, A. M. and Simmons III, J. Q. 1985: Prosodic development in normal and autistic children, in E Schopler and G B Mesibov (eds), *Communication problems in autism*, Plenum Press, New York.
- Baron-Cohen, S. 1993: From attention-goal psychology to belief-desire psychology: the development of a theory of mind, and its dysfunction, in S Baron-Cohen et al. (eds), *Understanding Other Minds: Perspectives from Autism*, Oxford University Press, Oxford.
- Baron-Cohen, S. 1995: *Mindblindness; An Essay on Autism and Theory of Mind*, MIT Press, Cambridge, Ma.
- Baron-Cohen, S. 1997: Hey! It was just a joke! Understanding propositions and propositional attitudes by normally developing children and children with autism, *Israel Journal of Psychiatry and Related Sciences* 34, 174-78.
- Baron-Cohen, S. and Bolton, P. 1993: *Autism: The Facts*, Oxford University Press, Oxford.
- Bennett, J. 1976: *Linguistic Behaviour*, Cambridge University Press, Cambridge.
- Bennett, J. 1978: Some remarks about concepts, *The Behavioural and Brain Sciences* 1, 557-560.
- Bishop, D V M. 1989: Autism, Asperger's syndrome and semantic-pragmatic disorder: where are the boundaries?, *The British Journal of Disorders of Communication*, 24, 107-121.
- Boucher, J. 1989: The theory of mind hypothesis of autism: explanation, evidence and assessment, *The British Journal of Disorders of Communication*, 24, 181-198.
- Boucher, J. 1996. What could possibly explain autism?, in P Carruthers and P K Smith (eds), *Theories of theories of mind*, Cambridge University Press, Cambridge.
- Bruner, J and Feldman, C. 1993: Theories of mind and the problem of autism, in S Baron-Cohen et al. (eds), *Understanding Other Minds: Perspectives from Autism*, Oxford University Press, Oxford.

- Churchill, D W. 1978: *Language of autistic children*, Winston and Sons, Washington.
- Davidson, D. 1970: Mental events, in *Essays on Actions and Events*, Clarendon Press, Oxford 1980.
- Davidson, D. 1974: Psychology as Philosophy, in *Essays on Actions and Events*, Clarendon Press, Oxford 1980.
- Davidson, D. 1975: Thought and Talk, in *Inquiries into Truth and Interpretation*, Clarendon Press, Oxford 1984.
- Davidson, D. 1982: Rational Animals, *Dialectica* 36, 317-327.
- Davidson, D. 1984: *Inquiries into Truth and Interpretation*, Clarendon Press, Oxford.
- Davidson, D. 1985: A new basis for decision theory, in *Theory and Decision* 18, 87-98.
- Davidson, D. 1991: Three varieties of knowledge, in A. J. Ayer; *Memorial Essays*, ed. Philip Griffiths, 153-166. Royal Institute of Philosophy, Suppl. 30.
- Dennett, D C. Beliefs about beliefs, *The Behavioural and Brain Sciences* 1, 568-570.
- DSM IV, *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition, American Psychiatric Association, 1994.
- Dunbar, R. 1996: *Grooming, Gossip, and the Evolution of Language*, Faber and Faber, London.
- Evans, G. and McDowell, J. 1976: Introduction, in G Evans and J McDowell, *Truth and Meaning; Essays in Semantics*, Clarendon Press, Oxford.
- Frith, U. 1989: A new look at language and communication in autism, *British Journal of Disorders of Communication*, 24, 123-150.
- Frith, U. and Happé, F. 1994: Language and communication in autistic disorders, *Philosophical Transactions of the Royal Society*, Series B, 346, 97-104.

- Frith, U and Happé, F. 1999: Theory of mind and self-consciousness: What is it like to be an autist?, *Mind & Language* 14, 1-22.
- Grice, P. 1957: Meaning, *Philosophical Review* 66, 377-388.
- Happé, F. 1991: The autobiographical writings of three Asperger syndrome adults: problems of interpretation and implications for theory, in U Frith (ed), *Autism and Asperger Syndrome*, Cambridge University Press, Cambridge.
- Happé, F. 1993: Communicative competence and theory of mind in autism: a test of relevance theory, *Cognition* 48, 101-119.
- Happé, F. 1995: The role of age and verbal ability in the theory of mind task performance of subjects with autism, *Child Development* 66, 843-855.
- Jarrold, C., Boucher, J and Russell, J. 1997: Language profiles in children with autism, *Autism* 1, 57-76.
- Jospeh, R. M. 1999: Neuropsychological frameworks for understanding autism, *International Review of Psychiatry* 11, 309-325.
- Jordan, R. 1989: An experimental comparison of the understanding and use of speaker-addressee personal pronouns in autistic children, *British Journal of Disorders of Communication*, 24, 169-179.
- Kanner, L. 1943: Autistic disturbances of affective contact, *Nervous Child* 2, 217-250.
- Laurence, S. 1998: Convention-based semantics and the development of language, in P Carruthers and J Boucher (eds), *Language and Thought. Interdisciplinary Themes*. Cambridge University Press, Cambridge.
- Leslie, A. and Roth, D. 1993: What autism teaches us about metarepresentation, in S Baron-Cohen et al. (eds), *Understanding Other Minds: Perspectives from Autism*, Oxford University Press, Oxford.

- Lewis, D. 1969: *Convention, A Philosophical Study*, Harvard University Press, Cambridge, Ma.
- Lord, C. 1985: Autism and the comprehension of language, in E Schopler and G B Mesibov (eds), *Communication problems in autism*, Plenum Press, New York.
- Menyuk, P. and Quill, K. 1985: Semantic problems in autistic children, in E Schopler and G B Mesibov (eds), *Communication problems in autism*, Plenum Press, New York.
- Mitchell, P., Saltmarsh, R. and Russell, H. 1997: Overly literal interpretations of speech in autism: understanding that messages arise from minds, *Journal of Child Psychology and Psychiatry* 38, 685-691.
- Paradis, M. 1998: The other side of language: pragmatic competence, *Journal of Neurolinguistics* 11, 1-10.
- Parisse, C. 1999: Cognition and language acquisition in normal and autistic children, *Journal of Neurolinguistics* 12, 247-269.
- Perner, J. 1993: The theory of mind deficit in autism: rethinking the metarepresentation theory, in S Baron-Cohen et al. (eds), *Understanding Other Minds: Perspectives from Autism*, Oxford University Press, Oxford.
- Premack, D. and Woodruff, G. 1978: Does the chimpanzee have a theory of mind?, *The Behavioural and Brain Sciences* 1, 515-526.
- Pylyshyn, Z. 1978: When is attribution of belief justified?, *The Behavioural and Brain Sciences* 1, 592-593.
- Quine, W.V. 1960: *Word and Object*, The MIT Press, Cambridge, Mass.
- Recanati, F. 1987: *Meaning and Force*, Cambridge University Press, Cambridge.

- Russell, J., Saltmarch, R. and Hill, E. 1999: What do executive factors contribute to the failure of false belief tasks by children with autism?, *Journal of Child Psychology and Psychiatry* 40, 859-868.
- Russell, J. and Hill, E, 2001: Action-monitoring and intention reporting in children with autism, *Journal of Child Psychology and Psychiatry* 42, 317-328.
- Rodier, P. M. 2000: The early origins of autism, *Scientific American*, February 2000.
- Schuler, A. L. and Prizant B. M. 1985: Echolalia, in E Schopler and G B Mesibov (eds), *Communication problems in autism*, Plenum Press, New York.
- Schiffer, S. 1972: *Meaning*, Clarendon Press, Oxford.
- Shwayder, D. 1965: *The Stratification of Behaviour*, Humanities Press, New York.
- Sodian, B. and Frith, U. 1993: The theory of mind deficit in autism: evidence from deception, in S Baron-Cohen et al.(eds)., *Understanding Other Minds: Perspectives from Autism*, Oxford University Press, Oxford.
- Sperber, D. and Wilson, D. 1986: *Relevance: Communication and Cognition*, Blackwell, Oxford.
- Tager-Flusberg, H. 1985: Psycholinguistic approaches to language and communication in autism, in E Schopler and G B Mesibov (eds), *Communication problems in autism*, Plenum Press, New York.
- Tager-Flusberg, H. 1993: What language reveals about the understanding of minds in children with autism, in S Baron-Cohen et al. (eds), *Understanding Other Minds: Perspectives from Autism*, Oxford University Press, Oxford.
- Tager-Flusberg, H. 1999: A psychological approach to understanding the social and language impairments in autism, *International Review of Psychiatry* 11, 325-334.

- Tager-Flusberg, H., Baron-Cohen, S. and Cohen, D. 1993: An introduction to the debate', in S Baron-Cohen et al.(eds), *Understanding Other Minds: Perspectives from Autism*, Oxford University Press, Oxford.
- Volkmar, F. R. and Klin, A. 1993: Social development in autism, in S Baron-Cohen et al.(eds), *Understanding Other Minds: Perspectives from Autism*, Oxford University Press, Oxford.
- Wellman, H M, 1993: Early understanding of mind, in S Baron-Cohen et al.(eds), *Understanding Other Minds: Perspectives from Autism*, Oxford University Press, Oxford.
- Wimmer, H. and Perner, J. 1983: Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception, *Cognition* 13, 103-128.
- Wing, L. 1988: The continuum of autistic characteristics, in E Schopler and G Mesibov (eds), *Diagnosis and Assessment in Autism*, Plenum Press, New York.
- Wing, L. and Gould, J. 1979: Severe impairments of social interaction and associated abnormalities in children: epidemiology and classification, *Journal of Autism and Developmental Disorders* 9, 11-29.
- Ziadas, K. Durkin, K. and Pratt, C. 1998: Belief term development in children with autism, Asperger syndrome, specific language impairment, and normal development: links to theory of mind development, *Journal of Child Psychology and Psychiatry* 39, 755-763.