External Signs of Internal Representations: Developments in Processing Utterances and Beliefs.

Sarah Parsons

University of Nottingham

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy, May, 2000.

Abstract

Three factors in Referential Communication are worthy of special consideration: the utterance, the speaker's internal representation, and reality. These relationships form the 'referential triangle' of communication. This thesis explores how children and adults evaluate utterances when all three elements of the referential triangle need to be considered. The main aim was to investigate why utterances might be more difficult to understand than other externalisations of internal representations, such as pictorial representations of belief.

Chapter 2 investigated the usefulness of presenting an internal representation as a cartoon thought bubble. Children with autism performed significantly better on false belief tasks when they saw the protagonist's belief encapsulated in a thought bubble, compared to a false belief task without a bubble. This suggests that thought bubbles can be easily understood as representations of mental states. Given this facilitation, the use of thought bubbles was extended to the referential communication paradigm in Chapter 3. Presenting speech and thought bubbles alongside the array allowed the referential triangle to be depicted as separate, substantive elements. Children aged 6-10 years tended to overlook the pragmatic adequacy of unambiguous utterances when they could see the speaker's meaning depicted in a thought bubble. In Chapter 4, the speaker's meaning was not shown directly, but had to be inferred from the story context. Under these circumstances, children and adults tended to focus more on the relationship between the utterance and the array when deciding whether a

message was adequate or not. Chapter 5 explored whether adults inappropriately overextended their focus on the utterance-array relationship. In some cases, adults seemed to be influenced by their own knowledge of utterance-array link when making evaluations from the perspective of a naïve listener protagonist.

The general pattern of results suggests that listeners are particularly attuned to discrepancies between elements in the referential triangle. It is possible that utterances as externalisations of internal representations are difficult to understand because children have to learn when it is appropriate to accord the discrepancy prominence versus situations when isomorphism between other elements in the triangle might be more important for utterance evaluation.

Acknowledgments

The person who deserves the greatest thanks for his help with this thesis, is Peter Mitchell. He has been a true friend and great supervisor throughout this period and nothing would have been accomplished without his unerring optimism, support, comfort and advice (plus the occasional beer)!
Thanks must also go to all the schools and colleges, pupils and staff, who participated in this research. Without their patience and cooperation the research could not take place and I am grateful for their generosity with time and resources. Thank you to: Hope Lodge School, Hampshire; Portfield School, Dorset; Purbeck View, Dorset; St. Thomas' CE J/I School, Birmingham; Hallmoor School, Birmingham; Uplands School, West Midlands; St. Gregory's School, West Midlands; Bournville College of Further Education, Birmingham; Grange Primary, Derbyshire; St.Laurence Primary, Derbyshire; Firfield Primary, Derbyshire; Undergraduate students at the Universities of Birmingham and Nottingham.

Big thanks also to:
Martin King for proofreading most of this thesis; a thankless task, for which I am eternally grateful. (Any mistakes are his fault, not mine!) Thanks also for providing weekly check-ups on my mental health.
Claire "Showbiz" Shahbazian for help with proofreading in addition to the unenviable task of coding the children's replies. Mostly, though, thanks because you have always listened to me and supported me in everything that I do. Your belief in me has helped me to get this far and I don't know how I would have done it without your help, humour and honesty.
All my friends who have kept me sane with kind words and gin: Hix, Anna Pet, Jenni "Rajah" Rogers, Paul and Corinne, Hulmey, and James.
My family who have always provided support of an emotional as well as a financial nature (B & D – thank you for everything. I wouldn't have, or be, anything if not for your love and constant belief in me).

Steve – for encouraging me to go on when I felt like giving up and for encouraging me to stop when going on would have been ruinous. I love you very much.

Finally, to Margot Fonteyn, whose words should never be forgotten:

"The one important thing I have learned over the years is the difference between taking one's work seriously and taking one's self seriously. The first is imperative and the second is disastrous"

Table of Contents

CHAPTER 3

Ambiguity and thought bubbles: What do children understand about speaker meaning as an internal representation?

CHAPTER 4

When do children evaluate utterances according to the speaker's belief? Developments from middle childhood to adulthood.

CHAPTER 5

Does the realist bias in adult reasoning extend to evaluations of ambiguous utterances?

CHAPTER 6

General Discussion

APPENDIX

Examples of experimental stimuli and story scripts

Chapter 2

Chapter 3

Chapter 4

Experiment 4.1.

Experiment 4.2.

Experiment 4.3.

Experiment 4.4.

Chapter 5

Experiment 5.1.

Experiment 5.2.

List of Figures

List of Figures in Chapter 1

List of Figures in Chapter 3

List of Tables

List of Tables in Chapter 2

List of Tables in Chapter 3

List of Tables in Chapter 4

List of Tables in Chapter 5

CHAPTER 1

Externalisations of internal representations: Beliefs, utterances and thought bubbles.

1.1. Externalising the representational mind.

The development of a conception of mind is fundamental to our ability to function as social beings. It enables us to exist in a social environment in which we interact in groups and depend on others for many of our needs (Mitchell, 1997). There are two key benefits of having a conception of mind. Firstly, we are able to consider self and others as organisms that think and believe, and who behave in accordance with those beliefs about the world. Secondly, a conception of mind allows us to communicate with others and interpret what they say. It is not sufficient in conversation simply to attend to the words that are contained in people's utterances. Instead, the context of the communication and the speaker's motives, intentions or beliefs must be considered in order for the interaction to be sensible and meaningful.

In both cases it is crucial to understand that beliefs and utterances are representations: a belief is a representation of reality, whilst an utterance is an aural representation of a belief (which, in turn, is a representation of reality). Beliefs, therefore, have a direct relationship with reality since the former is simply

a representation of the latter. By contrast, utterances have a more complicated relationship with reality, since they can only be interpreted via the mediating mind of the speaker.

Normally developing children typically find beliefs as representations easier to understand than utterances as representations. Usually, an understanding of the representational nature of beliefs develops around the age of four years (e.g. Wimmer & Perner, 1983), whilst a consistent understanding of utterances as representations emerges at around the age of six years (e.g. Robinson & Whittaker, 1987). The fact that the speaker's belief about reality is made explicit via their utterance, does not seem to help children to treat the utterance as a clue to the speaker's representation. This stands in contrast to literature which suggests that an understanding of beliefs as representations can be facilitated if the belief is externalised in some way, usually through the use of pictorial media. It might be that the more complicated representational relationship that characterises utterances, compared to beliefs, is responsible for the extra difficulty experienced in treating utterances as externalisations of representations. Not only must the utterance's relationship with reality be considered, but so too must the speaker's belief and how it corresponds with reality.

This thesis aims to investigate how the relationships between utterance, reality and belief are handled by children and adults in an attempt to find out why utterances might be especially hard to process correctly. The remainder of this chapter provides some detail about how children are helped to understand the

mind by seeing externalisations of beliefs via pictures, and seemingly not helped to understand the mind by hearing externalisations of beliefs via utterances.

1.2. Background literature: Using externalisations of representations to facilitate understanding of beliefs.

Since the publication of Wimmer and Perner's seminal paper in 1983, there has been a wealth of research conducted into the developing child's understanding of the mind. This research has served to focus our attention on when and how 3- and 4- year-old children come to understand that beliefs are representations of reality. Most children aged about 4 years are able to make the distinction between belief and reality, as evidenced by their success in a standard false belief task.

In the task, a story protagonist's belief is discrepant with the current state of reality. For example, Paul thinks the chocolate is in the drawer where he left it, but the chocolate has been moved to the refrigerator in Paul's absence. Once children come to understand that the belief Paul holds is a representation of reality, they realise that he will respond on the basis of his outdated belief (he will look for the chocolate in the drawer, where he thinks it is). By contrast, if children did not understand the representational nature of beliefs, then they would predict Paul's behaviour based on what they (the participant) knew about reality, rather than on what Paul (the protagonist) knows about reality.

Although there is considerable debate about how children ultimately arrive at an understanding that the mind is representational (see Lewis & Mitchell, 1994,

for a review), it is, nevertheless, widely accepted that being able to predict behaviour from inferences of others' belief states is a crucial aspect of social cognition. Without such insight, the social world would be a confusing and unpredictable place, since there would be no understanding that another's perspective or beliefs might differ from one's own. Consequently, other people's actions would often be difficult to fathom.

Reputedly, this is the unfortunate state experienced by people with Autism and Asperger's Syndrome, who have often been demonstrated to experience great difficulty in treating the mind as a representational entity. For example, they are significantly more likely to fail false belief tasks compared to normally developing children, or children with Down's Syndrome (e.g. Baron-Cohen, Leslie & Frith, 1985). They also often have significant difficulties treating speech as emanating from the mind, favouring overly literal interpretations instead (e.g. Mitchell, Saltmarsh & Russell, 1997; and Frith, 1989, for anecdotal evidence). Aside from experimental observations, people with autism who have talked or written about their experiences, state that the social world can be a strange and frightening place. In Oliver Sacks' book 'An Anthropologist on Mars' (1995), Temple Grandin – an American academic with Asperger's Syndrome – says that she avoided dating, finding '...such interactions completely baffling and too complex to deal with...[because]... she was never sure what was being said, or implied, or asked, or expected' (p.272).

Given the importance of understanding the mind as representational, there have been many attempts to facilitate understanding, or provide conditions in

which nascent insight into the representational nature of beliefs can be demonstrated, both within the normally developing and autistic populations. Focusing on normal development, Mitchell and Lacohee (1991) demonstrated false belief understanding in children younger than 4 years. In a standard deceptive box procedure (Perner, Leekam and Wimmer, 1987), the child is shown a Smarties tube and asked what is inside. After the child responds 'Smarties' the tube is opened and revealed to contain pencils instead. The lid is then replaced and the child asked to anticipate what another – uninformed – person would think was inside the tube. 3-year-olds typically said that the other person would think there were pencils inside the box, whilst 4-year-olds correctly replied that the other person would think the tube contained Smarties.

In Mitchell and Lacohee's 'posting' procedure, 3-year-olds were allowed to post a picture of Smarties into a postbox at the time they first stated what was inside the tube (before their belief was exposed as false). Children in this condition were significantly more likely to give the correct response (Smarties) compared to children who received a standard version of the task. Freeman and Lacohee (1995) replicated and extended this finding, reporting that children were significantly more likely to be aided by a pictorial representation of the belief item compared to an actual sample of the item (e.g. a picture of an egg compared to a real egg). Mitchell (1996) argues that the reason why pictorial representations serve as such useful facilitators is because they embody an intangible entity (belief) with a tangible counterpart in reality (a picture). This allows the belief to exist at an equivalent salience to reality. Mitchell (1996) suggests that the child

has an evolutionary bias towards reality since, at least to start with, reality is much more useful and advantageous for the child to negotiate than the mental realm. Therefore, there might be a bias towards reality which 'masks' children's understanding of beliefs as mental representations. It is only as the child matures that the salience of reality diminishes, allowing the child to think about the mental realm. Consequently, by elevating a belief on to an equivalent par with reality, the salience of reality is diminished, resulting in the child being able to think about mental states as representations which are separate from reality.

Video evidence of children's beliefs has also been shown to promote more correct acknowledgments of false belief compared to a standard procedure. Saltmarsh and Mitchell (1999) videoed children when they first saw the Smarties tube and said that they thought there were Smarties inside. After children were shown that the tube contained pencils, and just before they were asked what they thought was inside the tube when they first saw it, children saw a playback of themselves reporting their initial belief. They were significantly more likely to acknowledge their earlier false belief in the video condition, compared to a standard deceptive box task. Children were also significantly more likely to report their earlier false belief in a condition where the videotape was paused just before the child's initial belief was articulated. This suggests that the facilitation was not simply due to children repeating what they had heard themselves say on the tape without understanding why this answer was the 'correct' one. Once again, by providing belief with a physical counterpart in reality, children were much more likely to make correct judgments about belief.

In both of the above approaches, children benefited from seeing an externalisation of an internal representation. Both paradigms relied on the indirect route of finding out what children understand about mental representations by providing an analogue of the belief in a different format. Wellman, Hollander and Schult (1996) developed a more direct way of assessing what children understand about mental representations. Rather than using a picture as a clue to an earlier belief, Wellman et al., presented mental contents *as* a picture in the form of cartoon thought bubbles. By using pictures to represent mental content directly, thought bubbles (and hence, thoughts) can be interpreted as having their own representational content. Consequently, thought bubbles might be useful devices for facilitating understanding of the mind, since they perform the crucial function of raising mental content to the same level as reality.

Wellman et al., presented 3- to 4-year-olds with a number of different tasks involving thought bubbles. At the most basic level, even the youngest children understood that thought bubbles represented entities that could not be seen or touched, unlike 'public' representations such as photographs. Children also knew that thought content was person-specific, and was different to physical actions that the protagonist was shown performing. Moreover, children used the thought bubbles as a source of information about what the protagonist had seen inside a box. The child participant could not see inside the box and so their use of the thought bubble content suggests that they understood the thought as being specifically *about* a particular state of affairs in the world. Importantly, children were able to judge correctly about a protagonist's false belief as depicted in a

thought bubble. However, they were not tested on a standard false belief task, so evidence of facilitation was only tentative. Nevertheless, young children's understanding of thought bubbles as devices that depict mental phenomena was impressive. They seemed to understand the basic representational nature of thought bubbles from an early age, suggesting that thought bubbles may serve as especially useful externalisations of internal representations.

One way of assessing the usefulness of thought bubbles as clues to mental representations would be to present them to people with autism. Wellman et al., (1996) predicted that children with autism would find thought bubbles extremely difficult to understand because they lack any insight into the representational nature of the mind. Children with autism seem to have difficulties understanding mental representations specifically, and not with representations in general (Charman & Baron-Cohen, 1992; Leslie & Thaiss, 1992; Leekam & Perner, 1991). Consequently, the use of alternative, non-mental, representations to augment weaknesses in the mental domain within the autistic population has proved to be a popular approach amongst researchers; albeit with varied success.

For example, McGregor, Whiten and Blackburn (1998a) adapted the posting procedure (Mitchell & Lacohee, 1991) for use with people with autism. Participants were instructed to post a photograph of the original location of some chocolate, into a mannequin doll's head. Compared to pre-instruction performance, people with autism were significantly better at giving correct false belief judgments once they had been taught that the photograph could show them what the doll was thinking. However, unlike normally developing three-year-olds,

the children with autism failed to generalise this understanding to novel false belief tasks. In a separate study by the same authors (McGregor, Whiten & Blackburn, 1998b), people with autism were helped by the 'picture-in-the-head' procedure to judge correctly about false belief scenarios presented on video. However, in both studies a period of repetitive teaching was necessary before any facilitation could be seen, and in the latter study, only a small subset of the individuals were able to apply their understanding of the picture-in-the-head procedure to the video scenarios.

Swettenham, Baron-Cohen, Gomez and Walsh (1996) also used photographs as an analogue to mental states, in order to teach children with autism that a picture-in-the-head depicting a particular state of affairs can be useful in predicting behaviour. Whilst this method was successful in promoting more correct predictions about the protagonist's behaviour, none of the children with autism used the photographs to infer mental states. This suggests that the children may have learned a strategy for predicting behaviour, but seemed to have no insight that this was based on a mental representation. The authors suggest that a good way to make mental states more concrete (and by implication, help children to understand mental representation rather than simply behaviour) would be to present children with "cartoon think-bubbles" (p.86). The authors cite personal experience of a single case study in which a person with autism was helped to pass false belief tasks with the aid of thought bubbles. However, there has been no systematic exploration of whether children with autism might benefit from seeing thought content depicted in such a direct and tangible way.

Such an approach could have dual value. Firstly, autism could be a good assay for assessing the strength of thought bubbles as a representational clue. If children with autism showed an understanding of thought bubbles as representational entities, it would suggest that thought bubbles might be a particularly helpful way of externalising internal representations such that even one of the populations most impaired in this domain could derive some benefit. Secondly, if children with autism did understand, and derive benefit from, thought bubbles, this could provide evidence that there might not be a deficit in understanding the representational nature of the mind to the extent that some authors suggest (e.g. Leslie, 1987; 1994; Leslie & Thaiss, 1992). Indeed, Wellman et al., (1996) state that "At least when children develop a representational understanding of mind, [thought bubbles] might prove a sensible and easy to understand way to describe thoughts" (p.769). The implication is that only with some understanding of mental states as representational entities can children understand thought bubbles in any sensible way. These issues are tackled in Chapter 2.

1.3. The special case of utterances as externalisations of internal representations.

It would seem that externalisations of internal representations can be extremely valuable for revealing underlying competence in understanding the mind. However, utterances seem to provide children with more difficulty than the above statement would suggest. Utterances are aural representations of what a

speaker thinks, knows or believes. As such, utterances are externalisations of the speaker's meaning; but these two things might not always be isomorphic. Indeed, a speaker may choose to reveal as much or as little about their private thoughts and beliefs, with the implication that what is said may not always be exactly the same as what is meant. Therefore, a crucial insight into communication concerns the understanding that utterances are not merely direct reflections of reality, but are, instead, products of the speaker's mind. As a result, it is not sufficient in everyday communication to rely solely on the specific words that people say. Rather, communication is an active experience in which the listener may often infer underlying meaning from otherwise incomplete or inadequate messages.

Children below the age of 6 years seem to have difficulty treating utterances as products of the mind. Mitchell, Robinson and Thompson (1999) presented children aged between 3 and 6 years with a task which required interpreting utterances as representational cues. The task involved seeing a story protagonist, David, playing with some toy cars. He puts one of the cars in a toy garage and another on the racetrack and then leaves the room. In David's absence, Sarah plays with the cars and swaps their locations. David then calls to Sarah from another room, asking for one of the cars. In the message-desire discrepant condition, David states 'I would like the one in the garage please.' The child participant was asked to point to the car that David wanted.

If children considered the speaker's belief framework, they would know that David hasn't seen the swap and judge that David wants the car that *was* in the garage, but is *now* on the racetrack. This requires making a nonliteral

interpretation of the utterance. By contrast, children who do not interpret the utterance according to David's knowledge would respond literally – and pick the car in the garage. The children who interpret the utterance nonliterally can be said to understand that utterances are products of a speaker's mind. They realise that it is necessary to consider the speaker's internal representation of a particular state of affairs in order to make a 'correct' response. It was not until the age of about 6 years that children performed at ceiling on making a nonliteral interpretation in this condition.

The children who make a nonliteral interpretation of the utterance are able to discount the literal interpretation and respond according to the speaker's meaning instead. They understand that utterances are dual representations, consisting of a literal representation (the actual words of the utterance) and a non-literal representation (what the speaker means). The findings of Mitchell et al., (1999) correspond closely to the literature on referential communication which suggests that it isn't until around the age of about six years that children easily handle the distinction between what is said and what is meant (e.g. Robinson, Goelman & Olson, 1983).

Before this age, children tend to conflate literal and speaker meaning (Bonitatibus, 1988a). Young children often treat ambiguous utterances as adequate (e.g. Flavell, Speer, Green & August), presumably because they have little understanding of literal meaning aside from their own assumptions about what the speaker means (Beal & Belgrad, 1990). In addition, children younger than six years have difficulty understanding that it is the speaker's responsibility

to provide a good enough utterance such that a particular item can be picked out of an array of similar items. They tend to blame the listener, rather than the speaker, when the wrong item is chosen (Robinson, 1981). Seemingly, children have little understanding that the words of the utterance carry a specific meaning which determine communicative success.

Based on the analysis contained in the previous section (1.2) in this chapter, one of the problems with understanding utterances as externalisations of internal representations could stem from the fact that utterances are not substantive (Robinson, 1994). Unlike pictures, utterances are not permanent, visible representations and so they might not provide children with strong enough cues about the speaker's belief or meaning. In order to fortify utterances with substantive qualities, Bonitatibus and Flavell (1985) showed children the words of an utterance written on a board to see whether this might help them to detect ambiguous utterances. This manipulation did promote a significant improvement amongst 6-year-old children; they were much more likely to detect ambiguity when they could see the words of the utterance, compared to a control condition in which the message was illegible.

However, the key point is that the children who benefited from the manipulation were aged around six years. Younger children were not tested and the 6-year-olds were not at ceiling on the task. This suggests that even younger children would experience significant difficulties in detecting ambiguity, despite the words of the utterance being clearly presented. Consequently, the difficulty with treating utterances as representational cues does not simply seem to be a

function of them being ephemeral entities. Instead, there seems to be something special about utterances that makes them particularly hard to understand. This sentiment is echoed in Mitchell et al., (1999) who state that 'perhaps the interpretation of utterances should be viewed as a special kind of inferential problem' (p.63).

The question, then, is why are utterances a special case? Why do children seem to find it more difficult to treat utterances as externalisations of internal representations, compared to pictorial representations of belief? The central theme of this thesis is that utterances are one element of a tripartite relationship; the remaining two elements being the speaker's meaning and reality. To understand utterances as externalisations of internal representations, it is not sufficient simply to understand the dual representational nature of utterances (i.e. that what is said is a representation of what is meant). It also requires an understanding of how what is meant and what is said are related to a particular reality or context (cf. Robinson & Whittaker, 1987). This relationship is depicted in Figure 1.1.

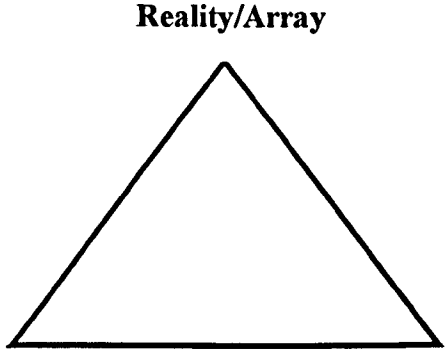


Figure 1.1: The triangular relation between reality, utterances and speaker meaning.

One of the most important qualities of this triangular relationship, is that all three elements could be different from each other. What the speaker means might be different from what the speaker says, which, in turn, could be different from the state of reality. For example, the speaker could say 'I really like the green house on the hill', but secretly mean 'I really hate the green house on the hill'. However, the speaker could be mistaken (for whatever reason: colour-blindness, reduced visibility, out-of-date knowledge etc.), and the house is actually red. By contrast, the relationship between an externalisation of a belief and reality is much more straightforward. Although the belief can be discrepant with reality, as in a false belief scenario, the external representation (e.g. picture) and the belief have to be isomorphic because the former is a direct 'stand in' for the latter. Thus, whilst it is crucial for the child to understand that the picture or photo is *identical* to the belief in false belief tasks, it is equally crucial for the child to understand that the external representation (i.e. the utterance) in communication tasks may *differ* from the speaker's meaning.

The notion of a three-way relationship characterising referential speech is by no means a new one (e.g. Robinson & Whittaker, 1987; Robinson, 1994). However, attempts to explore what children understand about the representational nature of utterances have tended to focus on only two aspects of the relationship rather than three. For example, Beal and Flavell (1984) chose to focus on knowledge of speaker meaning to see how this influenced judgments of message adequacy. They pointed to the intended referent before asking 6-year-olds whether the words of the utterance could refer to any of the other items in the

array. Children who were shown the intended referent in this way, were significantly less likely to detect ambiguous messages compared to children who were not shown which object the speaker 'was thinking about.' The authors argued that knowing the speaker's meaning led children to overlook the problems with the actual words of the utterance, because children have difficulty keeping the literal and non-literal representations separate. Consequently, a conflation of literal and non-literal meaning was interpreted as a sign that children have difficulty treating utterances as clues to the speaker's internal representation.

Beal and Flavell assumed that their methodology could stand to be informative about how children understand utterances in relation to the speaker's meaning and the array. They also assumed that, in referential communication, the speaker's meaning and the array are isomorphic; hence it is possible to infer the former from the latter. However, the problem with the first assumption in terms of the referential triangle (Figure 1.1), is that reality and meaning were confounded, since children were informed about the speaker's meaning by seeing a particular item in the array. Therefore, it is not possible to tell whether children were more concerned with utterance-array relations or utterance-meaning relations. The problem with the second assumption is that it is not inevitable that the speaker's internal representation and the array will be perfectly matched. Therefore, inferring the speaker's meaning from the array may not be the best way to access the speaker's internal representation. As a result, Beal and Flavell's procedure tells us little about how knowledge of the speaker's internal representation might influence judgments of utterance adequacy.

The difficulty in knowing which aspects of the referential triangle children are more concerned with, is a significant problem in ambiguity research (Robinson & Whittaker, 1987; Robinson & Mitchell, 1992). This is because the speaker's meaning is always assumed to be an exact match with the intended referent in the real world. Hence, when a problematical utterance leads to a failure to uniquely identify an item from a set of alternatives, it is not clear whether children judge the message to be bad because they notice that the utterance underspecifies the array or because they think there is a discrepancy between what was said and what was meant. The reliance on using reality as a means of inferring speaker meaning is perhaps a reflection of the difficulty in presenting mental states directly. However, the previous section (1.2) suggests that thought bubbles might provide an ideal way of informing children about a speaker's internal representation.

This would mean that, for the first time, the three components of the referential triangle could be depicted as separate, substantive elements. The referential array can be depicted alongside the utterance (contained in a speech bubble) and the speaker's meaning (contained in a thought bubble). This could allow a direct investigation into what makes utterances, in the context of referential communication, especially difficult to understand. Chapter 3 tries to elucidate where this difficulty might lie. Of course, in the real world, utterance and meaning are not usually substantive. Therefore, the direct approach outlined above needs to be supplemented with a paradigm in which the three components of the referential triangle are manipulated as separate elements, but in the absence

of concrete markers. It is possible that children might respond very differently to situations in which the speaker's meaning is clearly depicted via a thought bubble, versus situations in which the speaker's meaning is known to the participant, and differs from the array, but is not augmented by a pictorial representation. This approach forms the content of Chapter 4.

Finally, just as knowledge of the speaker's internal representation might influence judgments of how the message relates to reality, so too might knowledge of reality influence judgments about how a message relates to a speaker's belief. This possibility has not been investigated directly, but there is some indirect evidence that adults, at least, seem unable to ignore privileged knowledge about reality when they are required to reason about false beliefs. Mitchell, Robinson, Isaacs and Nye (1995) presented adults, and 5- and 9-year-old children, with video scenarios in which a story protagonist heard a message that supposedly informed them about a particular state of reality; for example, they were told that there was milk in a jug in the kitchen. The protagonist who heard the message had previously seen the contents of the jug. In some instances, the message contradicted the protagonist's belief (the protagonist had seen orange juice in the jug). In all cases, the observing participant had privileged knowledge about the truth or falsity of the speaker's message (e.g. they had seen the speaker swap the contents of the jug), and had to judge whether the listener would believe or disbelieve the message.

Both groups of children answered according to a 'seeing-leads-to-knowing' rule. That is, they judged that if the protagonist had seen inside the jug,

he would believe what he had seen rather than what he had been told. These judgments were made irrespective of children's knowledge about the truth of the message. By contrast, adult participants judged the believability of the message according to what they themselves knew about reality; they indicated that the protagonist would believe a message that they (the participant) knew to be true and disbelieve a message that they knew to be false.

It is possible that adults may be similarly 'contaminated' by their own knowledge if they were asked to make a judgment about the adequacy of an ambiguous message from the perspective of the listener. If they could be presented with scenarios in which they (the participant) knew whether a speaker's utterance was ambiguous or not, would they tend to judge the adequacy of the utterance on the basis of what they themselves know, or what the uninformed listener knows? In other words, adults might be more influenced by their knowledge of the utterance-reality relationship, even though this should be irrelevant to the judgment about the *listener's* perspective. Alternatively, they might consider the utterance-meaning relationship to be more important, since the *listener's* perception of utterance adequacy should be based on how what is said is related to what the speaker means. Whatever the outcome, this kind of approach could help to illustrate which elements of the referential triangle are deemed to be of particular significance in judgments about whether an utterance is adequate or not and is tackled briefly in Chapter 5.

1.4. Summary

This thesis is concerned with finding out more about how children and adults process externalisations of internal representations. The first step is to see whether thought bubbles might be useful externalisations for depicting internal representations. If thought bubbles are useful, then the depiction of the referential triangle as clearly separable, substantive elements becomes a possibility. The next step is to try to manipulate the elements that form the referential triangle – utterance, belief and reality – so that it becomes possible to speculate about which aspect(s) might be responsible for the apparent difficulty in treating utterances as externalisations of internal representations.

CHAPTER 2

Can thought bubbles facilitate understanding of mental representations in people with autism?

The following chapter forms part of the publication:

Parsons, S. & Mitchell, P. (1999). What children with autism understand about thoughts and thought bubbles. Autism, 3, 17-38.

2.1. Introduction

Thought bubbles are a well-known pictorial convention for informing people about what someone might be thinking. They provide an observable and concrete way of presenting thoughts and even very young children seem to understand them easily. Since a thought bubble is used to depict thoughts directly, it might be that they are an especially useful way of providing children with clues that thoughts are representational. Given that people with autism often struggle with this type of understanding, thought bubbles might equip them with enough representational support to be able to improve their performance on false belief tasks. If people with autism derive some benefit from seeing thoughts depicted in this way, thought bubbles could be interpreted as effective externalisations of beliefs.

2.2. Background literature

Within the cognitive domain, children with autism have been widely reported to suffer profound deficits in understanding the representational character of the mind (e.g. Baron-Cohen, Leslie & Frith, 1985; Baron-Cohen, 1989; Frith, 1989; Perner, 1991; Leekam & Perner, 1991; Leslie & Thaiss, 1992; Happe, 1994; Yirmiya, Solomonica-Levi, Shulman & Pilowsky, 1996). Specifically, it seems that children with autism have difficulty understanding beliefs as special kinds of mental representations '...intended to capture something real' (Wellman, 1990). That is, beliefs are conceptually distinct from other mental states such as dreams or fantasy in the sense that they purport to be about a real, external state of affairs. Although it is possible to have thoughts about non-existent entities (e.g. a thought about a unicorn), a belief about something entails a conviction that it is true (Wellman, 1990). The finding that many children with autism fail to acknowledge false belief suggests they might lack a 'Theory of Mind'.

In the test of false belief, a story protagonist is denied perceptual access to a change in reality (e.g. some chocolate is moved from a drawer to a fridge without the protagonist's knowledge), and the child participant is asked to infer where the ignorant protagonist will look for the chocolate. Normally developing 3- and 4-year-olds and many children with autism, even with verbal mental ages (VMAs) greater than four years, fail this task by reporting that the protagonist will look for the chocolate in its present location (Wimmer & Perner, 1983; Baron-Cohen et al., 1985). Supposedly, people with autism do not understand that beliefs depict

reality as being a certain way (that is, beliefs are only *representations* of reality), and consequently just report reality when questioned about belief (Perner, 1991).

Further evidence suggested that children with autism might have a very specific impairment in understanding the representational character of the mind. Leslie and Thaiss (1992; also Leekam & Perner, 1991) tested children with autism on Zaitchik's (1990) 'false-photo' task. Photographs, like beliefs, are 'reality-oriented' representations (Wellman, 1990) in the sense that they depict, or capture, a state of affairs in the real world. In this task, children watch as a Polaroid photo is taken of, for example, a doll in a green dress. Whilst the photo develops, the doll's dress is changed for a red one and children are asked to say what colour the doll's dress is in the picture. The photograph contains a representation of an outdated state of reality in much the same way as the protagonist's belief in the false belief task. However, children with autism found it much easier to anticipate the image in an outdated photograph than to anticipate the content of a person's outdated belief, suggesting that not all 'reality-oriented' representations pose a problem for children with autism. Furthermore, Charman and Baron-Cohen (1992) report that children with autism found 'false drawings', which depicted an obsolete state of reality, were much easier to understand than false beliefs. In all three of the above studies, normally developing preschoolers found the false photo/drawing tasks and the tests of false belief to be equally difficult. Thus it seems that people with autism experience great difficulty in understanding mental phenomena as representations, compared to their virtually

unimpaired understanding of non-mental representations such as drawings and photographs.

It is difficult to predict how children with autism would interpret a thought bubble. Whilst the bubble is essentially a drawing (a non-mental representation), it is nevertheless designed to represent the mental content of someone's mind. As such it seems to be a hybrid form of representation, neither exclusively mental or non-mental. Focusing on the mental aspect of thought bubbles, Wellman, Hollander and Schult (1996) predicted that children with autism would fail to understand bubbles as representing the content of someone else's mind, viewing them instead as a '...devilishly strange device' (p.787). In a series of experiments, Wellman et al. (1996) tested normal 3- and 4-year- olds' understanding of thoughts as simple representational entities, in the sense of '...capturing, ..depicting, ..or showing' something else (p.776). Even the youngest children understood thought bubbles to be private, intangible entities which depicted something in the world, in contrast to photographs which were judged as public and observable representations (see also Flavell, Green & Flavell, 1995). Indeed, they easily contrasted what a protagonist was thinking with a physical action they were involved in, suggesting that children as young as three understand thoughts as mental, rather than physical, activities.

Children were then asked to judge about the unseen contents of a box. A story protagonist was depicted looking into a box and in the next scene there appeared suspended above his head a picture of some scissors encapsulated in a thought bubble. Under another condition, a protagonist was shown taking a photo

of the content of a box, and in the next scene she held up the resulting photo of a ball. Under a third condition, a protagonist was shown reaching into a box and in the next scene was holding a book with a large picture of a teddy bear on the cover. The interior of the box could not be seen in any of the conditions. Children were asked to judge what was inside the box. If participants understood thought (bubbles) to be about reality, then they could utilise it as a basis for inferring the content of the box. If they did not, then various clues would be available to suggest as much. It might be that participants did not regard the thought bubble as revealing what the protagonist thought about reality in which case they would be confined either to saying that they did not know, or guessing.[1]

If they guessed, they might base this on any conspicuous image, such as the information contained within the thought bubble or photograph. If that was the basis of their judgment, then a tell-tale error should become apparent in their performance on the 'book' control task. In this, the protagonist lifted a book from the box which displayed a picture of a teddy on the cover. Hence, the conspicuously framed image associated with the protagonist was a teddy bear. If participants were judging correctly for the wrong reason in the thought bubble task, then presumably this would become apparent in their wrongly reporting

1: If participants were simply interpreting the thought bubble as something detached from reality, like daydreams [similar to Perner's (1991) notion of 'thinking of'] there would not be much evidence from the children of linking what the protagonist had seen with what they were thinking. Instead, children would be no more likely to answer the test question with the content of the thought bubble than with any other salient image in the scene, such as the cars on the floor. By contrast, if a child made correct judgments about the content of the box from the information in the thought bubble, they would be showing an understanding that seeing a certain thing conjures up an image of that thing in the mind, with the implication that this image is about reality ('thinking that'; Perner, 1991).

'teddy' as the content of the box in the book control task, when they should be reporting 'book'.

Wellman et al's (1996) youngest participants had few problems inferring the contents of the box from the thought bubble and the photograph and importantly, almost never referred to the teddy in the 'book' task. They, therefore, demonstrated a sensible understanding of thought bubbles as simple representational entities which could be used to inform them about reality. Also, contrary to Wellman's prediction that children with autism would find it difficult to understand thought bubbles, Parsons (1996) found that children with autism performed almost at ceiling on Wellman et al's thought bubble tasks outlined above. Moreover, the same group of children with autism performed significantly better on a test of false belief when a thought bubble depicted the protagonist's false belief compared to a standard task. Given the previous demonstrations of autistic children's impaired understanding of mental representations, it was expected that children with autism would perform poorly on tasks involving thought bubbles, but in line with normally developing four-year-olds on tasks requiring an understanding of photographic representations. Instead, however, it seemed to be the case that investing an intangible, mental representation (which autistic children have difficulty understanding) with a physical, observable embodiment (which they do understand), helped children with autism to attune to the representational character of thoughts.

McGregor, Whiten and Blackburn (1998a) obtained results consistent with such an hypothesis using a 'picture-in-the-head' procedure. They found that

people with autism were helped to give correct judgments on false belief tasks if they had previously posted a photograph of the original location of the chocolate into a slot in a specially modified doll's head. Creating a tangible image of the doll's belief seemed to enable people with autism to contrast belief and reality when these were discrepant. Importantly, McGregor et al. emphasised the notion of the posted picture being an indicator that the doll *thinks* something is a particular way and would subsequently act on the basis of this belief. By contrast, Swettenham, Baron-Cohen, Gomez and Walsh (1996) did not explicitly teach children about the link between a picture-in-the-head and a thought. Instead, they attempted to provide autistic participants with an alternative reasoning route that avoided 'thoughts' altogether; telling children that seeing something would lead to a picture-in-the-head which would then guide the protagonist's behaviour. This difference in emphasis could account for the finding that none of the children in Swettenham et al. (mean VMA 6:0) were able to use photos as a basis to infer mental states.

It would appear that emphasising 'thoughts' in tasks designed for children with autism does not seem to promote confusion and this is supported by the finding that children with autism can benefit from the inclusion of thought bubbles in tests of false belief. However, one limitation of the results from Parsons (1996) was that there was a general failure to find any differences between a group of normally developing 5-6-year-olds and the children with autism. Usually, it would be expected that the 5-6-year-olds would perform significantly better than the children with autism. In addition, there was a strong

positive correlation between VMA and success on the false belief tasks. The mean VMA of the group of children with autism was relatively high (7:6) and the mean CA of the group was much higher than 5-6 years (12:0), leading to the possibility that it is only the more high functioning individuals with autism that might benefit from thought bubbles. Other studies, too, have noted the ability of high functioning autistic participants to pass false belief tasks and even second (higher) order belief tasks (e.g. Sparrevohn & Howie, 1995; Holroyd & Baron-Cohen, 1993; Leekam & Perner, 1991). It was therefore decided to conduct a further study to investigate whether thought bubbles could help autistic children with lower VMAs to pass false belief tasks. In addition, it was important to establish the general usefulness of thought bubbles in helping different populations to attune to the representational content of the mind.

Experiment 2.1

2.1.1. Introduction

A clarification of the relative influence of verbal mental age to an understanding of thought bubbles could be achieved by testing children with autism with lower VMAs on a partial replication of Parsons (1996). In particular, the main point of interest was a comparison between a standard false belief task and a false belief task that also included a thought bubble. In Parsons (1996) the Standard task asked the children about a change of *location*, whilst the Bubble task required children to recall the *contents* of a box. Perhaps the observed improvement in performance on the Bubble task was attributable to the much less

interesting possibility that judging about contents is simply easier than judging about location. In the present experiment, children with autism, with lower VMAs, (between four and seven years) were tested on false belief tasks controlling for this possibility.

There were also additional improvements to the methodology. First, it could be that only presenting one test under each condition of false belief (Standard and Bubble), either over- or under- estimates children's ability. Participants could be credited with mentalistic understanding when correct responses have simply been fortuitous; or they could answer incorrectly due to extraneous task variables such as nervousness in the testing situation, and therefore be erroneously described as lacking any insight into mind. This concern echoes that of Happe (1995) who suggests that in some cases '...success on a single [false belief] task was due to chance' (p.851). Two versions of each false belief task were, therefore, included.

Second, a group of children with non-specific learning disabilities, also with VMAs between the ages of four and seven years, were included. Their participation allowed an investigation into the issue of the specificity of a Theory of Mind deficit in people with autism compared to learning disabled groups more generally. It also provided an opportunity to investigate whether children with learning disabilities can benefit from the addition of a thought bubble in false belief tasks.

2.1.2. Method

2.1.2.1. Participants. Four groups of children were included: normally developing 3-4-year-olds and 5-6-year-olds; children with autism and children with learning disabilities of mixed or unknown aetiologies. Both groups of normally developing children attended a primary school in a predominantly working class area of Birmingham. All of the 19 children with autism attended schools specifically catering for children with disorders along the autistic continuum.[2] Every child in this group had a statement of Special Educational Needs which explicitly included a diagnosis of 'Autism' (and in one case only, a diagnosis of Asperger's Syndrome). Teachers of the children with autism were asked to identify possible candidates with verbal mental ages greater than four years. Parents of identified candidates were then contacted to grant permission for their children to participate in the Study. Nine children attended the same school in Hampshire visited in Parsons (1996). Four children attended a day school for children with autism in Dorset and the remaining six attended an independently funded boarding school, also in Dorset. All nineteen children were tested for receptive language comprehension on the British Picture Vocabulary Scale (BPVS; Dunn, Dunn, Whetton & Pintilie, 1982). Participants in the learning

2. 5 out of the 9 children were participants in the first study 18 months previously. These children tended to give incorrect judgments on the standard false belief task in Parsons (1996) and correct judgments on the Bubble task. Their responses on the false belief tasks in the present study were less uniform; ranging from 0-4 correct responses, therefore suggesting little evidence of carry-over from Parsons (1996).

disability group attended a special school for mild/moderate learning disabilities in the Birmingham area; none had received a diagnosis of autism. Background details of participants are supplied in Table 2.1.

Table 2.1. Participant details of Chronological age (CA) and Verbal Mental Age (VMA).

	Mean CA (range in years:months)	Mean BPVS VMA (range in years:months)
Autistic (19 male)	13:5 (8:7 to 18:3)	5:7 (3:5 to 7:10)
Learning Disability (13 male, 6 female)	8:11 (7:2 to 10:8)	4:9 (3:4 to 7:5)
Normal 3-4 years (15 male, 13 female)	4:2 (3:8 to 4:11)	- -
Normal 5-6 years (15 male, 15 female)	5:10 (5:1 to 6:11)	- -

2.1.2.2. Materials. All introductory tasks and the four story scenarios were illustrated with coloured cartoon pictures and presented on A4 laminated sheets of card (examples of experimental stimuli and story scripts for all experiments in Chapters 2-5 are included in the Appendix. See pp. ix-xi for a summary). Each false belief story was presented on a total of three cards, despite small differences in number of illustrative pictures within each scenario.

2.1.2.3. Introduction to thought bubbles. This was based on the shortened version of Wellman et al's (1996) thought bubble tasks. Children were first shown a picture of a girl with a large thought bubble above her head

containing a bunch of flowers. The experimenter told the child that this was Susie and then pointed to Susie's jeans and T- shirt asking 'What's this?' each time. Next, the experimenter gestured broadly to the thought bubble and its contents and asked children what it was. If children made reference to mental states such as thinking or dreaming the experimenter confirmed this was correct, and if the participant gave a different answer or did not respond, the experimenter announced that the thought bubble '...shows us what Susie is thinking about.' Children were then asked 'So what is Susie thinking about?' The second picture showed a different girl, Jane, with a thought bubble encapsulating a picture of a birthday cake. The experimenter asked 'What is Jane thinking about?' The next two pictures required children to explicitly contrast mental activity with physical action. The first showed Peter, kneeling down, holding a dog by its lead. Above Peter's head was a thought bubble containing a picture of a car. Children were asked 'What is Peter thinking?' and 'What is Peter doing?', in fixed order. Next, children saw a picture of Ann, pushing a toy pram, and above her head was a thought bubble depicting a drink. Children were asked 'What is Ann doing?' and 'What is Ann thinking?' in fixed order.

2.1.2.4. Main Procedure. Each participant was then presented with four false belief tasks. Children were shown four story scenarios involving the protagonists Susie, Claire, Paul and John. The Susie and Claire scenarios were 'contents' stories as in 'What does x think is inside the box/bag?', and the Paul and John scenarios were 'location' stories as in 'Where does x think the (item) is?' Each story had two versions; one containing a thought bubble encapsulating

the protagonist's false belief in the final picture (Bubble condition) and one which did not (Standard condition). Children in each diagnostic group were split into two experimental groups. Group 1 received the Bubble followed by the Standard version for the content and location stories: They saw the bubble version of Susie and the Standard version of Claire, followed by Paul (Bubble) and John (Standard). Group 2 received the story narratives in exactly the same order, but the Bubble and Standard conditions were reversed.

The narrative from the John story provides an example of the style and length of stories used. Where necessary the experimenter pointed to the relevant details to make sure each participant was following the correct sequence of pictures:

'Here's John. John is playing with his toy cars on the table (picture 1), but John wants to watch TV in the other room. So John packs his cars away into a drawer (picture 2) and then leaves the room (picture 3).'

Children were then asked a prompt question: Where did John put his cars?

The narrative continues:

'Whilst John is watching the TV, his brother Mark is looking for the toy cars. He finds them in the drawer (picture 4) plays with them for a while (picture 5) and then puts the cars in the toy-box instead of back in the drawer (picture 6).'

Check question: Did John see Mark do that? No, he didn't did he...../That's right.....

..... 'John was watching the TV so he didn't see Mark put the cars in the toy box. John has come back into the room now (picture 7).'

Children were then asked a fixed-order series of questions:

Belief question: Where does John think the toy cars are?

Reality question: Where are the toy cars now?

Memory check: Where did John put the toy cars at the beginning of the story?

'Content' stories were identical in format except the protagonist was depicted putting something into (e.g.) a bag which was subsequently removed and replaced with a different item in the protagonist's absence. Children were then asked the belief question in the following form: 'What does Susie/Claire think is inside her bag/box?'; followed by the reality question: 'What is inside Susie's bag now?', and the memory check: 'What was in Susie's bag at the beginning of the story?'

Each child completed the Introductory tasks and the four false belief scenarios in two separate sessions. During the first session, children completed the 'Introduction to thought bubbles' procedure, followed by the Susie and Claire scenarios. On a different (usually consecutive) day, children were reminded of the Introductory tasks and were then presented with the two remaining false belief tasks (Paul and John). The children with autism and the children with learning disabilities were seen on a previous occasion where they were assessed for receptive VMA on the BPVS. Children were randomly assigned to experimental groups. The majority of children were tested individually in a quiet room, but a few of those with autism who reacted badly to environmental change were tested in their classrooms.

2.1.3. Results[3]

2.1.3.1. Introduction to thought bubbles. The majority of children did not refer to mental phenomena when the experimenter asked 'What's this?' whilst gesturing to the thought bubble on the first occasion it was presented. The most common response was to report the contents of the bubble (flowers); twenty 5-6-year-olds (66%) gave this response, as did 25 (89%) 3-4-year-olds, 17 (89%) children with learning disability, and 15 (78%) children with autism. Of the remaining children, seven 5-6-year-olds, and one child in each clinical group used mental vocabulary to describe the bubble, such as 'thinking', 'dreaming' or 'wishing', but none of the clinically normal preschoolers did so. Other responses were either 'don't know', no response or mention of 'cloud', 'bubble' or a colour in the picture (9 responses in total).

After being told by the experimenter that the bubble 'shows us what Susie is thinking', all in the total sample, except for one 3-4 year old, answered correctly that Susie was thinking about 'flowers'. An identical pattern of responding was found for the second picture of Jane; only one 3-4 year old gave an answer other than the correct one of 'cake'. Next, contrasting the actions of protagonists with their depicted thoughts, all but three children with autism said correctly that Peter was thinking about a car. The majority in each group also

3. To avoid the loss of potentially useful data, all children were included in these analyses. Some analyses of the data from the reality question only are also included, to assess directly whether the presence of the thought bubbles causes a decrement in correct reality judgments.

correctly referred to Peter's depicted action (holding a dog): 25 (89%) 3-4-year-olds, 17 (89%) children with autism, 28 (93%) 5-6-year-olds and 18 (94%) children with learning disability. Children performed similarly well on the final introductory picture of Ann. All of the 5-6-year-olds, 24 (86%) 3-4-year-olds, 15 (80%) children with autism and 17 (89%) children with learning disability correctly referred to the pram that Ann was pushing when asked what she was doing. Also, 18 (94%) children with autism and children with learning disability; 92% (26) of 3-4-year-olds and 100% (30) of 5-6-year-olds, mentioned the contents of the thought bubble ('drink') when asked what Ann was thinking. Generally, the majority of children had few difficulties understanding a thought bubble as a basis for judging what someone may be thinking.

2.1.3.2. Main Procedure: Comparing groups and conditions on the belief and reality questions. Each child was credited with a score of 0-2 to record how many (out of 2) stories they passed for the Standard and Bubble conditions. A score of 0 was awarded for incorrect answers on both story scenarios within each condition, a score of 1 for one correct answer, and a score of 2 was given for correct answers on both story scenarios. Table 2.2 shows the number of children in each group scoring 0,1 or 2 correct responses for the Standard and Bubble conditions. A belief judgment was only counted as correct if the accompanying reality question was also answered correctly; if the reality question was answered incorrectly the child automatically received a score of zero for that particular story. Diagnostic group was entered into an ANOVA as the between group variable while type of task (Standard or Bubble) was entered as

the repeated measure. The ANOVA revealed a significant main effect for diagnostic group [$F(3,92) = 15.89$, $p<0.001$] and type of task [$F(1,92) = 13.23$, $p<0.001$]. The interaction between task and group was not significant. Table 2.3 contains a summary of mean scores for each group.

To determine where the main effect between groups lies, data was combined across conditions to yield a score of 0-4 for each child, indicating correct judgments on the belief and reality questions for the four false belief tasks. Post-hoc Tukey's multiple comparisons revealed that 5-6-year-olds,

Table 2.2: Numbers of children in each group scoring 0, 1 or 2 correct responses for the Standard and Bubble tasks.

Group	Tasks Correct					
	Standard			Bubble		
	0	1	2	0	1	2
Autistic	13	1	5	7	7	5
Learning Disability	4	7	8	3	6	10
Normal 3-4 years	18	6	4	13	7	8
Normal 5-6 years	2	7	21	1	3	26

children with autism and the learning disability group achieved significantly higher scores on average than the youngest normal children. In addition, the learning disability group performed significantly better than the children with autism but significantly worse than the normal 5-6-year-olds ($p< 0.05$ in all

cases). There were no significant order effects detected in the data. That is, correct responses to the belief questions in the Standard and Bubble conditions were not dependent on which version children saw first.

2.1.3.3. Main Procedure: Comparing groups and conditions on the reality question only. The following analyses are based on the data from the reality questions only, irrespective of whether the belief question was answered correctly. Table 2.3 contains a summary of the mean scores for each diagnostic group for the reality question. None of the groups scored lower on the Bubble condition than on the Standard condition. Children were assigned a score of 0-2 for correct judgments on the reality question for the Standard and Bubble

Table 2.3: Mean scores (and standard deviations) on the belief and reality questions, and the reality question only, for the Standard and Bubble conditions.

Group	Mean scores (from a total of 2)			
	Test and reality questions		Reality question only	
	Standard	Bubble	Standard	Bubble
Autistic	0.58 (0.90)	0.89 (0.81)	0.95 (0.85)	1.11 (0.81)
Learning Disability	1.21 (0.79)	1.37 (0.76)	1.53 (0.61)	1.58 (0.61)
Normal 3-4 years	0.50 (0.75)	0.82 (0.86)	0.89 (0.83)	1.14 (0.85)
Normal 5-6 years	1.67 (0.61)	1.83 (0.46)	1.80 (0.48)	1.93 (0.25)

tasks, and the data were entered into a repeated measure ANOVA design as before. There was a significant main effect for group [$F(3,92) = 12.4$, $p<0.001$]

and for task [$F(1,92) = 5.44$, $p<0.05$], but no significant interaction between the two factors. Combining data across conditions (yielding scores of 0-4) for Tukey's post-hoc analyses, revealed that the 5-6-year-olds and the children with learning disabilities performed significantly better than the 3-4-year-olds and the children with autism. There were no significant differences between the former or the latter two groups. A two-tailed pairwise comparison of performance on Standard versus Bubble tasks combining across groups revealed that children made more correct reality judgments in the presence of a thought bubble than without this prop [$t(95) = 2.53$, $p<0.05$]. Therefore, there seemed to be no decrement in correct judgments of reality in the presence of the thought bubble.

2.1.3.4. Main Procedure: The influence of VMA. The following analyses are based on the combined theory of mind score from all four tasks (providing a score of 0-4 for each child). Separate analyses from the Standard and Bubble tasks are not included because they did not differ significantly from the combined score. In the autism group there was only a weak and non-significant association between VMA and total theory of mind score ($r = 0.26$). However, in the learning disability group there was a significant correlation between verbal mental age and theory of mind score $r = 0.51$, $p<0.05$. Additional analyses examined whether children passing all four tasks tended to register higher VMAs on average than those failing all four, or those who gave 'inconsistent' responses (scoring 1-3 correct answers). Amongst the children with autism, the mean age of the 'passers' was 6:5 ($n=4$; range 4:7 to 7:1) compared to a mean of 6:2 for the 'failers' ($n=6$; range 4:6 to 7:10). The average age of the 'inconsistent' children

was lower than both of these groups at 5:6 (n=9; range 3:5 to 7:2). The children with learning disability showed a clearer grouping of responding according to VMA, as would be expected from the strong correlations between VMA and Theory of Mind task success reported above. Seven children were correct on all four tasks and had a mean VMA of 6:1 (range 4:8 to 6:11). There were only two 'failers' (mean VMA 4:7), but 10 'inconsistent' children (mean = 4:9; range 3:4 to 5:9). Overall, VMA seemed to be an influential factor on false belief task performance for the children with non-specific learning disabilities, but not for the children with autism.

2.1.4. Discussion

Contrary to Wellman et.al's (1996) prediction that children with autism would fail to understand thought bubbles, at least some children in the autism group seemed to understand thought bubbles as representational devices that can be used to talk about the contents of thoughts and beliefs. Whilst the 5-6-year-olds and the children with learning disability generally performed better than the children with autism and the 3-4-year-olds, the evidence in the latter two groups was still largely consistent with their being able to derive some benefit from having a pictorial representation of a false belief. The only group not to show a significant benefit from the bubble consisted of the children with learning disability. Nevertheless, the results from the children with learning disability were in the same direction as the other groups, suggesting that there was a similar, but weaker, effect in operation. However, the lack of a significant effect amongst the

children with learning disability makes the significant difference in the group of children with autism all the more striking. In short, many children were helped to attune to the content of a belief when this was encapsulated in a thought bubble, and when this contrasted with the current state of reality.

Alternative explanations for this fledgling competence, such as 'hacked-out' solutions (Frith, Morton & Leslie, 1991; Eisenmajer & Prior, 1991), possible cueing effects or guessing, seem unlikely due to children's ability to judge appropriately about reality. If the bubble simply served as a general memory marker (or promoted low-level 'associative' responses such as linking the word 'think' with whatever was contained in the bubble), thereby prompting superficially correct belief judgments, then children would be led to make errors when asked about reality. That is, children could interpret the contents of the thought bubble *as* reality rather than as a *representation* of reality and, therefore, judge that what is contained in the thought bubble is really the contents of the box or the location of the object. This pattern of responding was apparent in the 3-year-olds in Parsons (1996), but not in those with autism in either study. Therefore, using correct judgments about belief and reality as the criterion for 'passing' provides persuasive evidence that children's understanding of false beliefs in the context of a thought bubble is appropriately representational.

These findings are particularly encouraging in the light of other research which suggests children with autism can only be helped to pass false belief tasks after periods of explicit teaching (e.g. Hadwin, Baron-Cohen, Howlin & Hill, 1996; Swettenham et al., 1996). The instruction in those studies was not always

successful and, as mentioned earlier, none of the autistic children in Swettenham et al. (1996) were able to use photos as a basis to infer mental states. Why should thought bubbles serve as a potentially more powerful aid to understanding mental phenomena?

Wellman (1990) makes the distinction between reality-oriented representations, such as beliefs and photographs, and fictional representations, such as drawings and other mental states like dreaming and fantasy. Since beliefs and photographs are both reality-oriented representations, it makes sense to predict that a photograph could serve as a useful 'metaphor' of belief (Swettenham et. al., 1996). However, drawings (so-called 'fictional' representations) seem to provide a more intuitively plausible link with mental phenomena: Both can refer to reality, can depict hypothetical states of affairs, and can accommodate non-existent entities (Wellman, 1990; Freeman, 1994). By contrast, a normal photograph can only represent what was in front of it at the time the button was pressed. It is not an active and constructive representational device like a mind, and cannot depict hypothetical or non-existent states as drawings can.

This is not to say that children hold an explicit *concept* of the similarities between drawings and beliefs, but rather that they understand both as representations and, at the very least, recognise a difference between the fixed, observable and veridical nature of photographic representations and the more accommodative and changeable nature of beliefs and drawings (cf. Zaitchik, 1990). In short, thought bubbles seem to make some kind of 'intuitive sense' to

children, enabling them to overcome difficulties in the department of representation by investing an intangible belief with a tangible counterpart in reality. Therefore, if a normal preschooler or a child with autism has even the rudimentary cognitive foundations in place for basic representational understanding (see Grant, Riggs & Boucher, in press, for additional evidence), perhaps such nascent insight can be augmented by a tangible suggestion of thoughts as representations (see also McGregor et al., 1998a).

This is consistent with Mitchell & Lacohee (1991; also Freeman & Lacohee, 1995) who found that 3-year-old children were significantly more likely to acknowledge their own earlier false belief, if they had previously posted a picture of their belief into a toy postbox. Indeed, Freeman (1994) suggests that children are able to understand the similarity between representations (e.g. between beliefs and pictures) at a "primordial level". That is, children possess some conscious understanding of the symbolic nature of representations before an understanding of representations becomes available to verbal report (the 'secondary' level). This results in the child being significantly aided by symbolic representations of beliefs in various tasks because a drawing (for example) serves as a symbolic 'tracer' which allows the child to contrast the referent of the belief with the belief itself. Therefore, the thought bubble might make 'intuitive sense' because it serves as a symbol which targets children's understanding of representations at an earlier stage than that required for a standard task.

The finding that upon initial presentation of the thought bubble, very few children spontaneously mentioned thinking or thoughts does not threaten the

'intuition' hypothesis. The important point is not whether children immediately recognised thought bubbles as such, but rather, whether they demonstrated a *rapid* understanding of thought bubbles once they had been told what the bubbles were. Indeed Wellman et al. (1996) emphasise this point: '...our data.....suggest that the introduction and pretraining needed for young children to understand and use thought-bubble depictions appropriately can be minimal'(p.786).

What remains a puzzle, however, is why children with autism typically go on to fail 'standard' tests of false belief (and other tasks requiring a representational understanding of mental states) if they possess some kind of rudimentary notion of thoughts as representational entities. Indeed, in treating the thought bubble effectively as a clue to reality rather than as reality itself, the children with autism seemingly related to the thought bubble differently than they routinely relate to speech. Referential speech can be regarded as an aural representation of reality. In other words, we allow ourselves to be informed by speech even in the absence of first-hand direct experience. Despite this, we do not regard speech as being reality itself, but only as a clue to reality that comes to us via another mind (Mitchell & Robinson, 1994; Robinson & Mitchell, 1992, 1994). However, children with autism are different, and they do tend to interpret speech literally, as though it were reality itself and seemingly do not apprehend the mind behind the message (Mitchell & Isaacs, 1994; Mitchell et al., 1997).

Perhaps children with autism understand enough about representation to be able to recognise a plausible suggestion, but not enough to be able to formulate the content of a representation unaided. Evidence in support of that view arises

from Mitchell and Isaacs (1994, Experiment 2), who report that children with autism succeeded in accepting a suggestion of verbal misrepresentation, whilst correctly rejecting untrue distracter suggestions. These same children seemingly were unable to acknowledge verbal misrepresentations without assistance in the form of suggestion. It might be that *sensible, intuitively plausible* suggestions as to the correct content of a representation, can be used to help children with autism understand instances of 'mental misrepresentation'.

2.1.5. Conclusion

The facilitatory effect of the bubble in the present study coupled with the competence of higher functioning children with autism reported in Parsons (1996), suggests that children with autism can be helped to answer appropriately about belief with the aid of thought bubbles. Moreover, at least for the children with autism, this benefit was not restricted to those with higher VMAs since there were no significant correlations between success on false belief tasks and verbal ability in the present study. Consequently, thought bubbles appear to be a powerful method for externalising internal representations, such that even members of a population documented to be significantly impaired in understanding the mental realm, can be helped by them.

In addition, it seems that normally developing children can also derive significant benefit from seeing the protagonist's false belief depicted in a thought bubble. By endowing a (usually) intangible belief with a physical embodiment, children can be helped to answer correctly about a protagonist's false belief.

Generally, it seems that competence in understanding the mind, both amongst normally developing children and those with autism, can be revealed by the use of thought bubbles. This suggests that the use of thought bubbles could be extended to alternative contexts in which seeing a protagonist's mental representation directly could provide a new perspective on children's understanding of the mind.

CHAPTER 3

Ambiguity and thought bubbles: What do children understand about speaker meaning as an internal representation?

3.1. Introduction

The results from Chapter 2 suggest that thought bubbles can serve as useful clues to the content of mental representations. In this sense, there is a parallel between thought bubbles and speech, since speech also provides information about the content of a speaker's beliefs. However, whilst normally developing children aged around 4 years can be significantly aided in understanding a thought bubble as an externalisation of a belief, it is not until around the age of 6 years that children typically begin to understand how an utterance functions as an externalisation of a speaker's internal representation or meaning. Beal (1988) states that '...an important component of the child's theory of mind is the understanding that the communicative quality of messages can determine the mental states of the self and others' (p.315). A substantial body of research carried out in the 1970's and 1980's lead to the general conclusion that children have difficulty with this type of understanding until the age of around six or seven years. Therefore, a thought bubble (as a clue to an internal

representation) does not seem to have the same effect as speech since there is at least a two-year gap between understanding the former compared to the latter. An important question, then, is why might there be extra difficulty in understanding speech as an externalisation of a representation?

The following three experiments were designed to explore this question using a procedure based on the Referential Communication literature and utilising the more contemporary thought bubble paradigm. The usefulness of thought bubbles in (a) helping children to treat beliefs and reality as separate representations and (b) informing children directly about the content of a protagonist's belief, might also help us to determine where the source of children's difficulty lies.

3.2. Background literature.

Referential communication is defined as "...communication about an object, location, etc., that would enable a listener to select a referent from nonreferential alternatives" (Sonnenschein, 1984; p.251). The referential communication paradigm was widely used to investigate children's understanding and production of utterances that *underspecify* the referent; that is, utterances that are ambiguous with respect to the alternative referents in a display. The typical approach required children to produce or respond to messages based on a small selection of items in an array (for example, pictures, building blocks or dolls). The items in the array were manipulated so that they differed from each other in terms of specific attributes such as colour, size, adornments and shape. Using this basic

set-up, researchers could investigate children's ability to successfully produce utterances so that a listener would be able to pick the referent that the speaker had in mind. Also, the paradigm provided a controlled situation in which children could interpret and evaluate utterances from a speaker who may, or may not, have done a very good job of describing a particular referent.

The general finding from this approach was that it is not until around age six or seven that children begin to understand the special role of language as a representation of the speaker's beliefs, thoughts or knowledge (e.g. Beal, 1988; Robinson & Whittaker, 1987). This understanding requires the child to consider the dual representational nature of verbal messages: (1) as a literal representation comprising the actual words spoken and (2) as a (sometimes incomplete and approximate) representation of the speaker's meaning (Bonitatibus, 1988a).

Developmentally, it is well known that children tend to focus on speaker meaning before they come to appreciate that an utterance has a literal meaning also. For example, children can recall the 'gist' of an utterance much more easily than the actual words of the message (Olson & Hildyard, 1983) and tend to respond to the underlying intention or 'illocutionary force' of the utterance before being able to respond only to the actual words spoken (Ackerman, 1981). Moreover, Beal and colleagues (Beal & Flavell, 1984; Beal & Belgrad, 1990; Beal, 1988) explicitly varied the accessibility of the speaker's meaning, by pointing to the intended referent, in message evaluation tasks. They found that 5 and 6-year-olds were significantly worse at detecting ambiguous utterances (or pictorial 'clues' in Beal & Belgrad, 1990) when they knew the speaker's intended

referent, compared to an 'uninformed' group who did not know the speaker's meaning. Beal (1988) argues that these difficulties arise because children younger than about six years are unable to '...maintain a distinction between the literal meaning and intended meaning of the speaker' (p.317) and, therefore, tend to conflate the two (Bonitatibus, 1988a, 1988b; Robinson, Goelman & Olson, 1983).

The failure to treat utterances as dual representations has been manifested in various ways. For example, children younger than age six will often judge an ambiguous or uninformative utterance to be adequate (e.g. Flavell, Speer, Green & August, 1981; Markman, 1977, 1979) and will tend to blame the listener, rather than the speaker, when communication fails and the wrong item is chosen (Robinson & Robinson, 1977, 1983). Moreover, the children that fail to detect referential ambiguity tend to be the same children who accept disambiguated versions of the original message as what was actually said (Robinson, Goelman & Olson, 1983). Correct performance on these tasks requires the child to understand that the actual words in the utterance carry a specific meaning, which is preserved independently of what the underlying intention of the speaker might be. These findings therefore suggest that, prior to the age of about six years, children seem to have difficulty discriminating the literal representation of the utterance from what may have been meant by it.

Research by Bonitatibus (1988b; Bonitatibus & Flavell, 1985) also supports the idea that problems in detecting ambiguity in utterances arise from a difficulty viewing the literal and speaker meanings as separable representations. Bonitatibus and Flavell (1985) presented 6-year-olds with a task in which a hand

puppet "Mr.Lion" wrote two-word messages (ambiguous and unambiguous) on a board, and the children were asked to judge whether they had been told enough to enable them to choose the referent. Children were significantly more likely to detect ambiguous utterances in this condition, compared to a condition in which the message was written in indecipherable scrawl. Using a slightly different procedure, Reid (1996) obtained similar results. It seems that drawing children's attention to the 'very words' of the utterance helped them to consider the role of the literal representation of the utterance in communication tasks.

By contrast, informing children of the speaker's meaning can lead them to overlook the role of literal meaning and, hence, cause them to judge inadequate messages to be informative. Beal and Flavell (1984) pointed out the intended referent to the child before asking them to evaluate the adequacy of the utterance. Their procedure involved placing a photograph of a girl 'Sheri' beneath one of the referents in an array so that it looked like Sheri was pointing to the object above her head. The children were then told that this was the picture/object that Sheri wanted them to find. The words of the utterance were also presented on message cards when children were asked to make an evaluation ("Do these words 'the blue one', let you pick this picture?"). Six-year-olds informed about Sheri's meaning were significantly less likely to judge the message to be ambiguous (i.e. say that the message would not let them to pick another item), compared to children ignorant of Sheri's meaning. The authors concluded that this was because 6-year-olds find it difficult to separate the speaker's meaning from the '...referential possibilities of the literal meaning of her message' (p.924). In other words,

knowing the speaker's meaning led children to overlook the referential problems with the message; they were less likely to attend to the pragmatic relationship between the actual words of the utterance and the items in the array. As a result, they judged the message to be adequate because they overlooked the fact that the utterance did not allow them to identify the referent uniquely.

In Beal's approach, children's detection of ambiguity was impaired by knowing the *referent* rather than by knowing the speaker's *representation of the referent.* This is an important distinction since the former focuses on meaning as the intended outcome, whilst the latter considers meaning to be an internal representation of the speaker (Robinson & Whittaker, 1987). The distinction between these two characterisations of meaning can be made clearer with reference to Figures 3.1 and 3.2. Figure 3.1 depicts the linear relationship assumed in characterising meaning as intended outcome (as in Beal's work), whilst Figure 3.2 represents the triangular relationship necessary for characterising meaning as an internal representation.

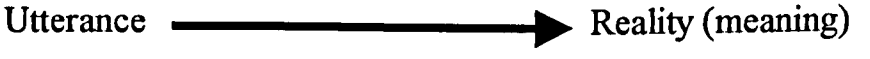


Figure 3.1: The linear relationship between the utterance and reality.

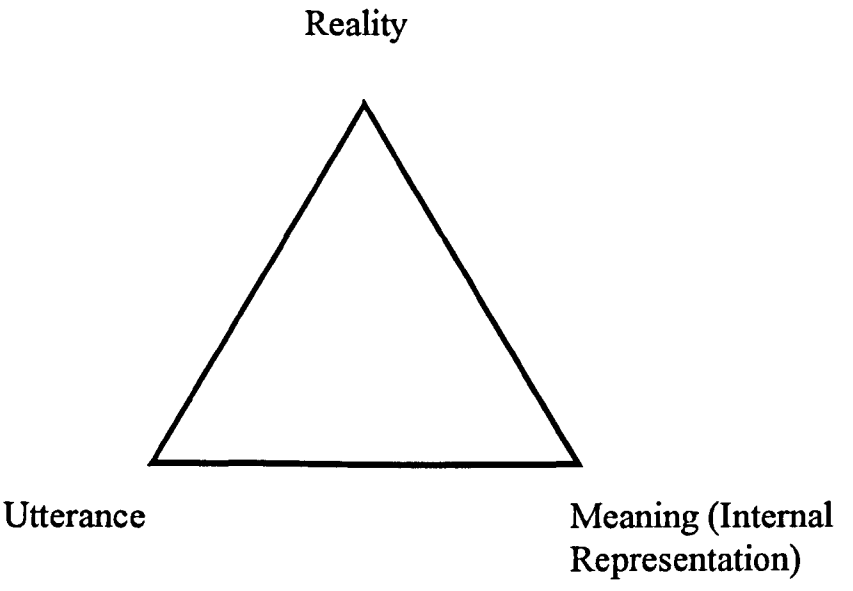


Figure 3.2: The triangular relation between reality, utterances and speaker meaning.

Initially, children tend to be outcome focused in their message evaluations. If a message results in the correct referent being chosen, even when a correct choice is merely fortuitous, children aged around 5 years will tend to judge that the message was adequate (Robinson & Whittaker, 1986; Ackerman, 1981). It is only when children come to understand the importance of the literal representation of the message that they are able to evaluate according to the informational value of the utterance rather than the outcome (Robinson & Whittaker, 1987). However, Beal's procedure might simply encourage children to retain focus on meaning as intended outcome, rather than making the next step to

considering meaning as an internal representation. Beal assumes that pointing to the intended referent is tantamount to revealing the speaker's internal representation. However, the problem with procedures that are covered by Figure 3.1 is that reality and meaning are confounded, such that it is impossible to tell whether the child is more concerned with utterance-array (referent in the real world) relations or utterance-meaning relations. Hence, it is not possible to make any conclusions about whether it is knowing the speaker's meaning (as an internal representation) that is responsible for children's difficulties in assessing the pragmatic value of the utterance. In addition, in Beal and Flavell (1984; also Beal & Belgrad, 1990) the experimenter points to a referent in the display, thereby ostensively providing an interpretation of the utterance. This could unwittingly supply the child with a cue on how the utterance *should* be interpreted, thus diverting attention away from the utterance such that they fail to fully consider all of its 'referential possibilities.' A linear relationship between the utterance and a particular referent, as depicted in Figure 3.1, has been pre-specified by the experimenter, perhaps making it difficult for the child to say that the words of the utterance could refer to any other items in the array.

A procedure that allows the three components of referential communication, Figure 3.2, to be separated out would provide a more direct test of how children handle knowledge of the speaker's meaning. In the three experiments that follow, children were informed about a speaker's representation directly by seeing a thought bubble suspended above the protagonist's head in a cartoon story. Thought bubbles are ideal vehicles for explicitly informing children

about internal representations, and for dissociating meaning (as depicted in a thought bubble) from the utterance (presented in a speech bubble) and from the referent array. In addition, revealing the speaker's meaning to children via a thought bubble removes the problem of making an ostensive interpretation of the utterance. Since the speaker's meaning has not been explicitly represented before, it is unclear how children should be expected to behave in situations where the speaker's internal representation is shown directly. However, it is possible to make a two-tailed prediction based on some of the findings outlined above. On the one hand, given that many researchers (e.g. Beal, 1988; Bonitatibus, 1988a; Robinson et al., 1983; Robinson & Whittaker, 1987) argue that difficulties with ambiguous utterances arise because young children conflate literal and intended meaning, thought bubbles (combined with speech bubbles) might help children to make correct judgments because the two types of meaning are visually separated. By contrast, Beal and Flavell (1984) might argue that children would continue to experience difficulty on ambiguity tasks because being informed about the speaker's meaning – in whatever form – is problematic for younger children.

Whether or not the utterance was represented in a speech bubble was also manipulated in Experiment 3.1. As noted above, Bonitatibus and Flavell (1985) and Reid (1996) demonstrated that children were significantly better at detecting ambiguity when utterances were written on cards rather than spoken. Therefore, it might be the case that any facilitatory effect of the thought bubble is especially noticeable when a speech bubble is also present because the literal and speaker meanings are clearly separated. On a more general level, Ackerman (1993)

showed that children were much more likely to make correct judgments about ambiguous utterances if they were presented with information that aided access to relevant details in memory. Providing the utterance in a speech bubble means that children do not have to struggle to remember exactly what was said. Therefore, at least some of the time, children were presented with stimuli that provided the best chance of giving correct evaluations.

Children aged between 7 and 9 years were included in Experiment 3.1 and children aged 5 to 10 years were included in Experiments 3.2 and 3.3. This was so that any developmental trend in the understanding of meaning as an internal representation could be assessed. Robinson and Whittaker (1987) suggest that understanding meaning as an internal representation is a later development than understanding meaning as an intended outcome. Therefore, perhaps only the older children would derive any benefit from the depiction of meaning in a thought bubble.

Experiment 3.1.

3.1.1. Introduction

In the following experiment, one of the main interests was in the effect of presenting the utterance in a speech bubble in addition to the effect of displaying the speaker's internal representation in a thought bubble. To maintain consistency of media between thought and speech, the information in the thought bubbles was presented in written form.

3.1.2. Method

3.1.2.1. Participants. 19 Year 3 children (11 males and 8 females; mean age, 7:5; range 7:2 to 8:1) and 20 Year 5 children (13 males and 7 females; mean age, 9:9; range 9:6 to 10:1) were included. All attended a LEA maintained Primary School in Derbyshire.

3.1.2.2. Materials. With the exception of the 'Batman and Superman' comic, stories were illustrated with coloured cartoon pictures and presented on A4 card. Pictures were placed in clear plastic wallets and mounted in a Lever Arch File. The comic was no.41 in the 'I love to Read' series published by The Redan Company Ltd (1999).

3.1.2.3. Design and Procedure. All children received a series of Introductory tasks, completed in fixed order, before the experimental stories were presented. They then received a total of eight story scenarios, four of which contained a speech bubble and four which did not. Half the children in each age group received all eight stories with a thought bubble present and the remaining children received all stories without a thought bubble. 6 stories (3 speech bubble absent and 3 speech bubble present) were Ambiguous and 2 stories (1 speech bubble absent and 1 speech bubble present) were Unambiguous. Thought Bubble present/absent was, therefore, a between-subjects classification and Speech Bubble present/absent was a within-subjects classification. Story scenarios were divided into two blocks and the order in which the stories were presented in each block was fixed across all subjects. The order of Unambiguous and Ambiguous

trials was rotated within each block and was identical for both speech bubble absent and speech bubble present trials. Presence or absence of the speech bubble was mapped onto each block, so children received all four stories (3 ambiguous and 1 Unambiguous) with speech bubble absent followed by all four stories with speech bubble present (for example). The order in which children received speech bubble absent and speech bubble present stories was alternated between subjects.

3.1.2.4. Introduction to Thought and Speech Bubbles. Children were shown a copy of a 'Batman and Superman' comic, with a large picture of Superman on the front cover. The experimenter then turned to the first page and started to read the story and the dialogue in the speech bubbles (there were no thought bubbles on the first page). At the end of the first page, the experimenter pointed to the speech bubbles and asked the children 'So, what is special about these bubbles? What are these bubbles showing us?'. The majority of children correctly responded that they were speech bubbles. In the rare situation of a child not knowing about speech bubbles, the experimenter explained that the bubbles 'show us that someone is saying something.'

On the second page, there was a combination of speech and thought bubbles and the experimenter continued to read these aloud. At the end of the second page, the experimenter pointed to the thought bubbles and said 'These bubbles are a bit different to the others aren't they? What are these bubbles showing us?' If the children were not sure what the bubbles were, the experimenter explained that 'they show us what someone is thinking'. The experimenter clarified the difference between the bubbles once again by pointing

to a speech bubble and saying 'So this bubble shows us what someone is *saying*' and, pointing to a thought bubble, 'this bubble shows us what someone is *thinking*'.

Children were then presented with two Introductory pictures, both of which depicted a protagonist with a speech bubble and a thought bubble. The first picture showed Paul kneeling on the floor, holding a dog by its lead. The words 'good dog' were contained in the speech bubble and the word 'football' in the thought bubble. Children were asked 'Can you show me which bubble shows us what Paul is saying?' The majority of children were able to point to the correct bubble and were then asked to read what the words said. Children were asked 'Can you show me which bubble shows us what Paul is thinking?' and asked to read the content of the thought bubble. If children pointed to the wrong bubble initially, they were prompted by the experimenter 'Is that his speaking bubble or his thinking bubble?'. This was sufficient for children to change their response and point to the correct bubble for 'saying' and 'thinking'. This procedure was repeated for the second picture of Sheila who was shown pushing a pram, saying 'hello' and thinking 'drink'. The only difference was that children were asked about thinking before they were asked about speaking.

Children were told: 'Now we are going to look at some stories together and in each of the stories, someone tells us something that they like or something that they want. Some of the people are quite good at doing this and some of the people are not so good, and I need you to help me figure out who's doing a good job and who's not doing such a good job, OK?'.

3.1.2.5. Main Procedure. Each story scenario included a protagonist who saw a display of four items (e.g. bags). In the Ambiguous trials, all four were the same colour but differed in additional details (shape/stripes/spots etc.). In the Unambiguous trials, the four items were exactly the same as the ones in the Ambiguous trials except that each was a different colour. So, for example, in an Ambiguous scenario where Rosie is getting ready to go swimming the first picture shows Rosie standing in the changing rooms where there are four yellow bags. The observing child participant was told 'Rosie is getting ready to go swimming'. They were then shown the second picture of Rosie in the swimming pool with her friend Megan and told 'In the pool, her friend Megan wants to borrow Rosie's swimming goggles and asks Rosie which bag is hers. Rosie says "the yellow one"'. The only difference in the Unambiguous condition was that only one of the bags was yellow.

If the trial included a speech bubble, the experimenter pointed to each word as Rosie's utterance was read aloud. Very often, children spontaneously read the utterance aloud with the experimenter. The utterance was read twice by the experimenter. If a thought bubble was included, the experimenter then pointed to the bubble and said '...and Rosie *thinks* 'the yellow one with the flowers''; making sure to point to each word as it was read. The content of the thought bubble was repeated by the experimenter. The protagonist's thought was always read after the utterance. Note, in the Ambiguous trials, the content of the thought bubble allowed the utterance to be disambiguated by the child participant because only one of the yellow bags had flowers on it. By contrast, the information

contained in the thought bubble on Unambiguous trials was redundant since there was only one yellow bag and it also had flowers on it.

Children were then asked the test question 'Did (speaker) tell listener exactly which (object) was his/hers? Or 'Did (speaker) tell listener exactly which (object) he/she wanted?' (depending on story context; from Sodian, 1988). The experimenter recorded the child's response on pre-prepared data sheets. Each child was seen individually in a quiet room at the school. At the end of the procedure, children were given a sticker for their participation and thanked for their help.

3.1.3. Results.

Children received a score of 1 every time they said 'no' to the test question; indicating that the speaker did not say exactly which item they wanted. A summary of overall means can be seen in Table 3.1.

Table 3.1: Summary of overall mean scores - saying 'no'- (and standard deviations) for the Ambiguous and Unambiguous conditions for Years 3 and 5.

	Ambiguous (max. = 6)	Unambiguous (max. = 2)
Year 3	3.58 (2.67)	0.68 (0.88)
Year 5	5.90 (0.31)	0.90 (0.79)

3.1.3.1. Ambiguous condition. The data were analysed in a 2 (Year Group) x 2 (Thought Bubble: present or absent) x 2 (Order: Speech Bubble present or absent first) x 2 (Speech Bubble: present or absent) mixed design

ANOVA. Year Group membership, Order and presence or absence of Thought Bubble were between group factors, and presence or absence of Speech Bubble was a within subjects factor.

There was a significant main effect of Year Group [$F(1,31) = 15.38$, $p<0.001$]. The 9-year-olds were significantly more likely to say that the speaker had not said exactly which item they wanted (mean = 2.95, from a total of 3, SD = 0.82) compared to the 7-year-olds (mean = 1.91, SD = 0.82). There was also a main effect for Order [$F(1,31) = 6.47$, $p<0.05$], which was qualified by a Year Group x Order interaction [$F(1,31) = 6.47$, $p<0.05$]. Figure 3.3 illustrates this interaction and shows that the Year 3 children were much more likely to say 'no' if the first block of trials they received contained a speech bubble compared to the group who received speech bubble absent trials first.

There was no order effect for the Year 5 children, who were performing virtually at ceiling on this task. There was no significant main effect for Speech Bubble and no main effect for Thought Bubble (although the latter only narrowly missed significance: $F(1,31) = 3.89$, $p = 0.058$). The trend was in the direction of higher 'no' scores in the presence of a thought bubble (mean = 2.69, SD = 0.82) compared to when it was absent (mean = 2.17, SD = 0.82). There were no further significant interactions between any of the variables.

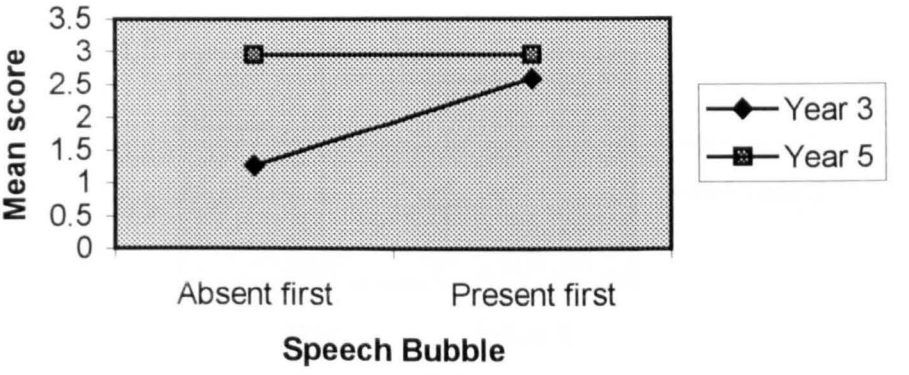


Figure 3.3: Mean scores for the Evaluation question (saying 'no') for both Year Groups according to the order of presentation of Speech Bubble present or absent trials.

3.1.3.2. Unambiguous condition. The data were submitted to the same ANOVA design as before. There was a significant main effect of Thought Bubble [$F(1,31) = 58.43$, $p<0.001$]. Children were significantly more likely to report that the speaker did not say exactly what he/she wanted in the presence of a thought bubble (mean = 0.81, from a total of 2, SD = 0.28) than in its absence (mean =0.13, SD = 0.28). There were no main effects for Speech Bubble or Year Group, but there was a significant Thought Bubble x Year Group interaction [$F(1,31) = 9.96$, $p<0.005$], which is illustrated in Figure 3.4. This shows that whilst both groups were more likely to say 'no' in Thought Bubble present compared to Thought Bubble absent trials, this effect was more pronounced amongst the 7-8-year-olds. The difference between thought bubble absent and present trials maintained for each year group independently [Year 3: $F(1,18) = 46.15$, $p<0.001$; Year 5: $F(1,19) = 13.23$, $p<0.005$].

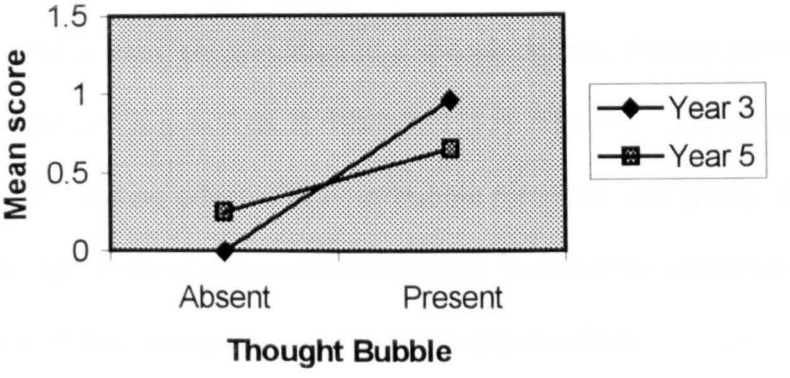


Figure 3.4: Mean scores for the Evaluation question (saying 'no') for both Year Groups according to presence or absence of Thought Bubble.

There was also a significant difference between the year groups in the Thought Bubble absent trials independently of Thought Bubble Present trials [$F(1,19) = 6.0$, $p<0.05$]. However, this was probably due to a floor effect because the children in Year 3 never said 'no' in this condition. There was no significant difference between year groups when the Thought Bubble Present trials were analysed independently of Thought Bubble Absent trials.

3.1.4. Discussion

Overall, the 9-year-olds demonstrated an almost ceiling performance on the Ambiguous trials; reporting that the speaker had not said exactly which item he/she wanted. Although the 7-year-olds demonstrated a fairly good understanding of ambiguity, they were still significantly poorer than the 9-year-olds in this respect. That is, they were significantly less likely to acknowledge that the speaker did not say exactly what he/she wanted. The average score for the 7-

year-olds on the Ambiguous trials was 1.89 (from a total of 3, SD = 0.88), suggesting that there was room for improvement if the presentation of the 'very words' of the utterance were encapsulated in a speech bubble, thereby possibly facilitating performance (Bonitatibus & Flavell, 1985). This was not the case, however, since there was no effect of speech bubble for either year group. This suggests that drawing children's attention to the literal words of the utterance did not help children to detect ambiguity using the present procedure.

Perhaps more importantly though, knowledge of the speaker's meaning in the Ambiguous trials did not significantly reduce children's ability to answer the test question appropriately, in contrast to the findings of Beal and colleagues (Beal & Flavell, 1984; Beal & Belgrad, 1990). If this had been the case we would expect to find a significant reduction in mean 'no' responses in the thought bubble present condition compared to thought bubble absent. However, there was no significant main effect for thought bubble when the speaker's utterance was Ambiguous. In fact, the direction of mean scores according to whether the thought bubble was present or absent was in the opposite direction to that suggested by Beal's hypothesis. It might be that children do not have difficulty making judgments about utterance adequacy when speaker meaning is made known to them (as an internal representation). Rather the problem arises because, in previous studies, speaker meaning has been construed as intended outcome rather than as an internal representation (but maybe not – see below).

Unlike the Ambiguous trials, children performed similarly on the Unambiguous trials, irrespective of Year Group. Also, somewhat surprisingly,

both groups of children demonstrated a significant tendency to say that the speaker did not say exactly what he/she wanted when there was a thought bubble present compared to when it was absent. This resulted in children seemingly judging adequate messages to be insufficient or faulty in some way. Perhaps this is not so surprising if we consider what children were actually asked to do. That is, the information contained in the thought bubbles was always in the form of 'the referent *plus x*' (e.g. 'the yellow one *with the flowers*'), whereas the utterance was always in the form of 'the referent' only (e.g. the yellow one). Therefore, it was always the case that the information in the utterance was not *exactly* the same as that which appeared in the thought bubble, hence it is likely that children noticed the discrepancy between utterance and thought and judged in accordance with this. In addition, children saying 'no' in the Ambiguous condition could also have been judging negatively because they noticed the discrepancy between thought and speech rather than noticing that the utterance was ambiguous with respect to the referent display. In other words, children could be credited with especially good performance in the Ambiguous condition for the wrong reason. It might be that seeing the speaker's meaning does result in children overlooking the pragmatic 'fit' between the utterance and the array, as suggested by the findings of Beal and Flavell (1984) and Beal and Belgrad (1990), but it is not possible to tell using the present procedure. One way to address this issue would be to ask children to justify their evaluations. This would provide some insight into whether children are judging according to the utterance-thought discrepancy or whether

they are aware that it is the utterance-array relationship that is important for determining ambiguity of reference.

There are additional improvements that could be made to the procedure. Firstly, children could have focused on the relationship between thought and speech because they were 'cued' to do so by the length of the phrases in the utterance and thought bubble. That is, the length of the phrasing in the thought bubble was always longer than the phrasing in the utterance. This may have made it difficult for children to overlook the utterance-thought discrepancy, thereby making it less likely that they would notice (in the Unambiguous trials) that the statements expressed were mutually equivalent with respect to reality (they both adequately specified only one referent). Secondly, children need to be provided with an equal number of Ambiguous and Unambiguous trials. This ensures that performance in one condition can be directly compared to performance in the other. Finally, younger children need to be included in order to test more closely the claim made by Bonitatibus and Flavell (1985) that presenting the actual words of the utterance in written form significantly improves children's ability to detect ambiguous utterances. It might be that an effect of the speech bubble would be found with slightly younger children who perhaps have a more fledgling and tentative understanding of the role of the literal representation of meaning in communication tasks.

Experiment 3.2.

3.2.1. Introduction

The main aims of this experiment were, firstly, to continue to explore the role of literal and speaker meaning in tasks of referential ambiguity and, secondly, to investigate the interesting result from Experiment 3.1 that children tended to focus on the discrepancy between thought and speech in the Unambiguous condition, rather than the relationship between speech and reality. To summarize the methodological improvements:

(1) Children were asked to justify their responses to the test question so that it was possible to investigate whether children used the utterance-reality or the utterance-thought relationship as the main criterion for their utterance evaluations.

(2) The written statements included in the thought bubble and the utterance were made equal in length.

(3) Children received an equal number of Ambiguous and Unambiguous trials.

(4) 5- and 6-year-olds were included in addition to 7- and 9-year-olds.

3.2.2. Method.

3.2.2.1. Participants. 35 Year 1 children (21 males, 14 females; mean age, 5:11; range 5:5 to 6:5), 30 Year 2 children (15 males, 15 females; mean age, 6:11; range 6:6 to 7:4), 36 Year 3 children (23 males, 13 females; mean age, 7:10;

range 7:5 to 8:5) and 33 Year 5 children (21 males, 12 females; mean age, 9:9; range 9:6 to 10:4) were included. Children in Years 2, 3 and 5 attended a LEA maintained Primary School in Derbyshire. The Year 1 children attended an LEA maintained Infant School, also in Derbyshire.

3.2.2.2. Materials. Exactly the same stimulus materials were used as in Experiment 3.1. The only difference was that extra words were added to the Speech Bubbles to make the phrases equal in length to those in the Thought Bubbles.

3.2.2.3. Design and Procedure. The procedure was identical to Experiment 3.1 except for the following alterations. Firstly, instead of receiving 2 Unambiguous and 6 Ambiguous trials in total, children were presented with 4 trials of each type. Secondly, Unambiguous and Ambiguous trials were alternated throughout the procedure and the order in which they appeared alternated between participants. Thirdly, to control for the number of words in the utterance compared to the thought bubble, a redundant phrase was added to the utterance. For example, an utterance which previously read 'the yellow one' was changed to 'the yellow one in the changing rooms'. The additional phrase (e.g.) '...in the changing rooms' did not alter the utterance-reality relationship in either condition, since, for all scenarios, the information was equally relevant to all items in the display (all bags were in the changing rooms). The information in the thought bubble remained the same as that presented in Experiment 3.1, for example ' the yellow one with the flowers'. Finally, after the test question was asked 'Did

(speaker) tell (listener) exactly which (item) he/she wanted?', children could justify their response by answering the question 'Why/why not?'

3.2.3. Results

3.2.3.1. Responses to the test question. As in Experiment 3.1, children received a score of 1 every time they said 'no' to the test question; indicating that the speaker did not say exactly which item they wanted. This was a correct judgment in the Ambiguous condition, but an incorrect judgment in the Unambiguous condition. Data were analysed in a 4 (Year Group) x 2 (Thought Bubble: present or absent) x 2 (Order: Ambiguous first or Unambiguous first) x 2 (Speech Bubble: present or absent) x 2 (Condition: Unambiguous or Ambiguous) mixed design ANOVA. The first three factors were between subject classifications and the remaining two were within subject classifications.

There was a significant main effect for Condition [$F(1,118) = 124.31$, $p<0.001$]. As expected, children were significantly more likely to acknowledge that the speaker did not say exactly which item they wanted in the Ambiguous condition (mean = 1.36, SD = 0.76) compared to the Unambiguous condition (mean = 0.64, SD = 0.66). However, this effect was confined to the children in Years 2, 3 and 5 as can be seen in Figure 3.5, which illustrates a significant Year Group x Condition interaction [$F(3,118) = 8.20$, $p<0.001$].

The children in the three older age groups showed a significant tendency to answer 'no' more often in the Ambiguous condition compared to the Unambiguous condition, and this was confirmed for each year group

independently [Year 2: $t(29) = 3.64$, $p<0.001$; Year 3: $t(35) = 7.28$, $p<0.001$; Year 5: $t(32) = 6.98$, $p<0.001$].

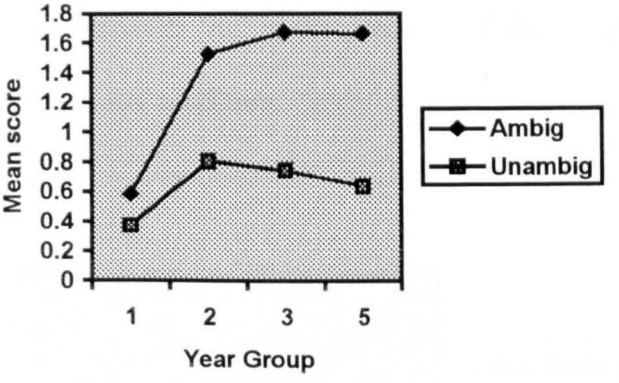


Figure 3.5: Mean scores for the Evaluation question (saying 'no') for each Year Group in the Unambiguous and Ambiguous conditions.

This trend was also found for the Year 1 children, but the difference between scores for the Ambiguous and Unambiguous conditions was not significant [$t(34) = 1.87$, $p = 0.07$]. Post hoc Tukey's comparisons showed that the Year 1 children differed significantly from the other three year groups in the Ambiguous condition (children in Years 2, 3 and 5 did not differ significantly from each other). That is, the Year 1 children were significantly less likely than the other three groups to evaluate negatively in the Ambiguous condition. A similar pattern of responding was found in the Unambiguous condition also, where Year 1 children were significantly less likely to say 'no' compared to the other three groups.

There was a significant main effect for Thought Bubble [$F(1,118) = 8.68$, $p<0.005$]. Children were much more likely to respond 'no' (that the speaker did

not say exactly what they wanted) in the presence of a thought bubble (mean = 1.15, SD = 0.61) than in its absence (mean = 0.84, SD = 0.61). A significant Thought Bubble x Condition interaction (Figure 3.6) [$F(1,118) = 20.09$, $p<0.001$] indicates that the tendency to answer 'no' in the presence of the thought bubble was confined to the Unambiguous condition.

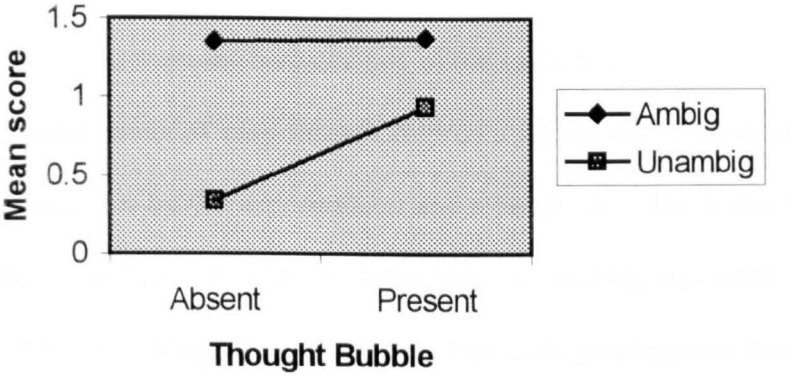


Figure 3.6: Mean scores for the Evaluation question (saying 'no') across all Year Groups for the Unambiguous and Ambiguous conditions, according to presence or absence of Thought Bubble.

To examine whether the difference between scores in the Unambiguous condition maintained independently of scores in the Ambiguous condition, data were submitted to a one-way ANOVA with Thought Bubble Present or Absent as the between subjects factor. The ANOVA confirmed that children were significantly more likely to say 'no' in the Unambiguous condition when a thought bubble was present rather than absent [$F(1,133) = 26.73$, $p<0.001$].

The only other significant result was a 3-way interaction of Condition x Year x Order [$F(3,118) = 4.37$, $p<0.01$]. In an attempt to interpret the 3-way interaction, pairwise comparisons were used to check the differences in mean

(5) Making reference to the listener's mental state e.g. the shopkeeper will know which one, Dad won't be able to tell which picture.

(6) No answer/don't know/not sure.

(7) Invoking hypotheticals e.g. there could have been more there, she might have looked somewhere else.

(8) Other e.g. he wanted that one, she told the truth, he forgot his money.

Table 3.2 contains a summary of children's justifications for the Ambiguous condition according to Thought Bubble Present or Absent. In the absence of a thought bubble, the majority of children answering that the speaker had not said exactly which item he/she wanted tended to give an array based justification. That is, a clear majority in all age groups in the Ambiguous condition mentioned multiple similar items in the referent display. Even though, in total, there was a low number of negative responses amongst the 5-6-year-olds, it is still noticeable that 70% of justifications from the youngest children who made correct evaluations mentioned the ambiguous array as the source of the problem. Generally though, the majority of the 5-6-year-olds evaluated positively in the Ambiguous condition with most (53%) being unable to provide a justification, and many (36%) providing uninformative or nonsensical explanations. There were only 5 justifications, across all four year groups, that were utterance based; that is, saying that the speaker did not provide enough information about the desired referent.

In the presence of a thought bubble, the 5-6-year-olds responded very similarly to those who did not see a thought bubble. Again, the majority (67%) of

Table 3.2: Total number of justifications (and percentages)[a] for the Ambiguous condition according to Thought Bubble present or absent.

	Number of responses (%ge)	1 Multiple objects	2 Single object	3 Not enough info	4 Thinking it/not saying it	5 Listener's mental state	6 No answer/ Don't know	7 Hypotheti-cals	8 Other	Totals
	Year 1									
	Yes[b]	0	2	0	0	2	28 (54)	1	19 (36)	52
	No[b]	14 (70)	0	1	0	3	1	0	1	20
TB Absent	*Year 2*									
	Yes	7	0	0	0	0	7	0	0	14
	No	42 (93)	0	1	1	0	1	0	0	45
	Year 3									
	Yes	0	0	0	0	4	7	0	1	12
	No	52 (87)	0	3	0	2	0	2	1	60
	Year 5									
	Yes	4	2	0	0	0	5	0	0	11
	No	55 (96)	1	0	0	0	1	0	0	57
	Year 1									
	Yes	1	6	0	0	0	23 (49)	0	17 (36)	47
	No	14 (67)	1	3 (14)	2	1	0	0	0	21
TB Present	*Year 2*									
	Yes	1	2	1	0	0	9	0	1	14
	No	14 (30)	1	24 (52)	5	0	0	1	1	46
	Year 3									
	Yes	1	0	0	0	0	11	0	0	12
	No	20 (33)	3	16 (27)	8	2	8	1	2	60
	Year 5									
	Yes	1	0	0	0	0	9	0	1	11
	No	27 (55)	0	7 (14)	1	2	10	0	2	49

[a] For the sake of brevity, only the modal percentages of high response categories are provided. [b] Yes or No indicates response to the Evaluation Question.

the children who evaluated negatively made reference to the multiple similar items in the array; whilst those who gave positive evaluations either did not answer (49%) or gave uninformative replies (36%). By contrast, there were markedly different patterns of responding within the three older groups of children. Around a third of the 6-7 and 7-8-year-olds, and just over half of the 9-10-year-olds mentioned the ambiguous array when justifying their negative evaluations. The 6-7-year-olds seemed to be especially heavily focused on the utterance with 52% of justifications falling into this category. Thereafter, there was a decrease in the tendency to focus on the utterance with only 27% and 14% of this type of justification amongst the 7-8 and 9-10-year-olds, respectively.

Table 3.3 contains a summary of children's justifications for the Unambiguous condition. Most of the children in Year 1 evaluated positively in the absence of a thought bubble, but could not give sensible reasons for their judgments; 42% could give no answer at all, whilst 43% gave uninformative replies. Amongst the older groups of children, there was an increasing tendency with age to single out the uniquely identified referent in the array when justifying positive evaluations. Even so, it is still surprising that amongst the 7-8 and 9-10-year-olds, around a third in each case could not provide explanations for their judgments.

In the presence of a thought bubble, responses were more mixed. Compared to the thought bubble absent trials, there were fewer children overall who answered positively and, amongst these children, relatively few mentioned the uniquely identified item in the array; only 39% and 38% of the Year 3 and

Table 3.3: Total number of justifications (and percentages) for the Unambiguous condition according to Thought Bubble present or absent.

	Number of responses (%ge)	1 Multiple objects	2 Single object	3 Not enough info	4 Thinking it/not saying it	5 Listener's mental state	6 No answer/ Don't know	7 Hypotheti-cals	8 Other	Totals
	Year 1									
	Yes[a]	1	5 (8)	0	0	3	26 (42)	0	27 (43)	62
	No[a]	2	1	2	0	2	0	0	3	10
TB Absent	*Year 2*									
	Yes	0	19 (43)	0	0	4	19 (43)	0	2	44
	No	2	1	5	2	1	3	1	0	15
	Year 3									
	Yes	0	32 (55)	0	0	3	20 (34)	0	3	58
	No	3	4	1	0	1	0	5	0	14
	Year 5									
	Yes	0	37 (61)	1	0	0	21 (34)	0	2	61
	No	1	0	4	0	0	0	2	0	7
	Year 1									
	Yes	0	16 (31)	0	0	7	19 (36)	0	10	52
	No	3	2	5 (31)	0	0	2	2	2	16
TB Present	*Year 2*									
	Yes	0	14 (54)	3	0	0	8 (31)	0	1	26
	No	1	0	15 (44)	10 (29)	2	3	0	3	34
	Year 3									
	Yes	0	13 (39)	0	0	1	18 (54)	0	1	33
	No	0	1	22 (56)	7 (18)	1	2	4	2	39
	Year 5									
	Yes	0	11 (38)	0	0	0	17 (59)	0	1	29
	No	1	1	16 (52)	3 (10)	0	9	1	0	31

[a]Yes or No indicates response to the Evaluation Question.

Year 5 children (respectively) answered in this way. Also, there was still a substantial proportion of children in each year group who were not able to provide justifications for why they had evaluated positively. The highest percentage for no answer/don't know was found amongst the 9-10-year-olds (59%). Of those evaluating negatively, many focused on the utterance, saying that the speaker had not provided enough information about the referent. Also, a few of the children in Years 2, 3 and 5 reasoned that because there was no speech bubble visible on the page, the speaker had not said anything at all.

3.2.4. Discussion

All children, except the 5-6-year-olds, reliably discriminated between Ambiguous and Unambiguous trials, at least in some conditions. The three older groups were significantly more likely to acknowledge that the speaker did not say exactly what he/she wanted when the utterance was Ambiguous. The youngest children, in line with previous literature (e.g. Mitchell & Robinson, 1994), tended to respond positively (saying 'yes') on most trials – indicating that they experienced difficulty in acknowledging that the speaker's utterance was ambiguous. Moreover, the 5-6-year-olds often struggled, in both the Unambiguous and Ambiguous conditions, to give meaningful justifications for their answers. Often, their justifications included superfluous, fictional or meaningless information, such as 'because she said please', 'it was nice of her' or 'they are the same colour as his top.' This suggests that the youngest children had little understanding of the utterance as an object for scrutiny and evaluation

(Flavell, 1977; Beal & Flavell, 1984) since they rarely interpreted the justification question as being *about* the utterance.

At first glance, and in line with results from Experiment 3.1, there was no evidence to suggest that being informed of the speaker's meaning significantly reduced children's ability to detect ambiguity in utterances. Indeed, performance in Ambiguous trials, for all children, was almost identical whether a thought bubble was present or absent. Ostensibly, most children seemed to do a good job of judging ambiguous utterances negatively, even when they knew the speaker's meaning. This seems to stand in contrast to the findings of previous research (Beal & Flavell, 1984; Beal & Belgrad, 1990) which has suggested that children aged around 6 years are significantly impaired in their ability to detect ambiguous utterances if they are informed previously of the speaker's meaning.

Experiment 3.1 highlighted the fact that children could have been responding negatively (a correct response in the Ambiguous condition) for the wrong reason. Instead of noticing that the utterance was inadequate with respect to the referent array, children could have been evaluating negatively because they focused on the discrepancy between what was said and what was meant. Considering the types of justifications that children made in Experiment 3.2, it is clear that the criteria for evaluations differed markedly between Thought Bubble Absent and Thought Bubble Present trials. Generally, children almost never mentioned the utterance in their justifications when the thought bubble was absent. Instead, the majority of children in the older three groups made reference to the array in both Unambiguous and Ambiguous conditions.

However, in the presence of a thought bubble, children were much more likely to focus on the utterance and say that the speaker did not provide enough information about the item he/she wanted. That is, when children knew the speaker's meaning, they were more likely to overlook the relationship between the utterance and the array, choosing, instead, to focus on the discrepancy between thought and speech. This is consistent with Beal and Flavell's (1984) results, which demonstrated a significant tendency amongst 5- and 6-year-olds to focus on the speaker's meaning to the extent that the pragmatic value of Ambiguous utterances was ignored. The present approach suggests that this tendency diminishes gradually during the middle school years, with even some of the 9-10-year-olds still focusing on the utterance-thought relationship rather than on the utterance-array relationship.

The most surprising aspect of the results, though, is the extent to which children seem to focus on the utterance-thought link in the Unambiguous condition. The significant effect in Experiment 3.1, whereby children were more likely to judge an Unambiguous utterance to be inadequate in the presence of a thought bubble, compared to when a thought bubble was absent, was replicated across all age groups. Moreover, there was no diminishing trend to focus on the thought bubble less as children got older. Instead, similar numbers of children in Years 2, 3 and 5 evaluated negatively and approximately half of the relevant justifications in each group indicated that the speaker had not provided enough information about the referent. Also, somewhat surprisingly, many children – even amongst the oldest group – who evaluated negatively in the presence of a

thought bubble in the Unambiguous condition, were unable to provide justifications for their answers. This was not the case in the Ambiguous condition, where the vast majority of children (except in the youngest group) who evaluated negatively, were able to provide meaningful explanations. It could be that children experience some conflict when they say that an Unambiguous utterance is no good. They might know that the utterance *is* adequate in the sense that it uniquely specifies an item in the array, yet judge negatively because they focus on the difference between what is said and what is meant. As a result, they might be unwilling to give a justification that includes reference to the thought bubble because they know that this is 'wrong' in some way. Generally, it might be relatively easy for children to make an array based comparison when an utterance is Ambiguous, but this seems a harder process for Unambiguous utterances. In short, children can often see clear reasons why an utterance may be bad, but find it more difficult to realise why an utterance may be good. It seems they are not always satisfied that an utterance is sufficiently unambiguous.

Despite the finding that children seemed to focus on the speech-thought relationship, there was no evidence to suggest that the speech bubble had any effect on children's judgments. On the one hand, the effect of judging negatively (in both conditions) in the presence of a thought bubble could have been especially strong with a speech bubble also present, because the difference between what was said and what was meant was clearly depicted and available for scrutiny. On the other hand, children might have been more likely to judge positively in the Unambiguous condition or negatively in the Ambiguous

condition because the literal meaning of the utterance was emphasised, as in Bonitatibus and Flavell (1985). It was suggested that the lack of such a facilitative effect in Experiment 1 may have been due to the omission of a younger group of children, since the seven-year-olds in the first experiment may already have a fairly robust understanding of the role of literal meaning in referential communication (Olson, 1970). As a result, children may have been able to appreciate the 'very words' of the utterance irrespective of whether they appeared in written form or not. However, even the youngest group of children in Experiment 3.2 seemed to derive no benefit from seeing the words of the utterance written in the speech bubble. In addition, older children tended to focus on the utterance-meaning relationship irrespective of whether the utterance was depicted in a speech bubble or not. This suggests that the tendency to mention the inadequacy of the utterance in the presence of a thought bubble was not simply because children could see that the words on the page differed according to what was said or what was meant. The most pertinent question, therefore, is what leads children to focus on the thought bubble to the extent that they overlook the relationship between the utterance and the array?

Pedantically, one could argue that it is natural to attend to the utterance-thought relationship in these circumstances, rather than the relation between the utterance and the array. Maybe it is 'correct' to say that the speaker did not say *exactly* which item they wanted since the utterance did not contain exactly the same words as the thought bubble. To assess this possibility, a group of 15 mature students were presented with 2 Ambiguous and 2 Unambiguous stories (in

alternating order) with speech and thought bubbles present. They were exposed to the same story scenarios as the children and asked the same test questions. Whilst a significant majority of the children judged that the speaker had not said exactly which item they wanted in the Unambiguous conditions, the mature students rarely did so. Out of a total of 30 responses for the Unambiguous condition, only 5 (17%) indicated that the speaker did not say exactly which item he/she wanted. Therefore, the majority of adult responses (83%) indicated that the speaker had said exactly which item they wanted. It seems that whilst adults interpreted the test question pragmatically (focusing on the utterance-array relationship), children were much more likely to focus on the utterance-thought relationship.

A more theoretically interesting explanation could be that children only gradually relinquish their early tendency to attend initially to the speaker's underlying intent or meaning. This is in contrast to Olson (1981 – cited in Bonitatibus, 1988b) who argues that an understanding of ambiguity signals a revolutionary shift in the child's thinking in the early school years because they are able to "...differentiate what the speaker means from what the sentence means, and to pay attention to the latter" (p.7). Moreover, Bonitatibus (1988a), based on Olson's (1970) theory of reference explains how the sentence, or literal meaning, should be construed: 'the literal meaning [of the utterance] is instantiated by the set of all possible referents in the array... [therefore] children capable of detecting referential ambiguity and attending to the literal meaning of the message should be attending to the very words of the message, as well as their multiple instantiations in the referent array' (p.329). In other words, children

experience a conceptual leap in their understanding of referential communication around the age of six years because they learn that speaker meaning can be different from literal meaning. In addition, they learn that the relationship between the utterance and the array is centrally important for deciding whether an utterance is ambiguous or not. It seems that Olson's (and also Bonitatibus, 1988a) account may be correct when considering the way adults make their judgments, but children older than 6 years seem to respond in a markedly different way. They continue to be influenced by the speaker's meaning at a time when they should have come to understand the primary importance of the utterance-array relationship. In short, it may be that acquiring a grasp of the relative importance of literal meaning, speaker meaning and their relationship to the array does not occur in a stage-like shift once the child understands that utterances can be ambiguous.

It could still be argued, however, that children's understanding of the pragmatic importance of referential communication is being underestimated by the present procedure. It might be that the evaluation question encourages children to answer in a particular way - focusing on the utterance-thought discrepancy rather than on the utterance-array link. It is important to establish whether children can understand that although the speaker's utterance is not identical to their internal representation, the utterance is, nevertheless, sufficient in pragmatic terms. That is, the utterance still allows a unique referent to be identified. In Experiment 3.3, therefore, an additional question was included to see whether children would be able to make an accurate judgment about

knowledge sufficiency from the perspective of the listener ['will (listener) pick the right (referent)?]. Even children as young as five years (e.g. Ackerman, 1993) can discriminate reliably between Unambiguous and Ambiguous utterances when they are asked whether they 'knew enough to pick exactly the right object.' This judgment requires a straightforward comparison between utterance and the array to decide if the speaker has been told enough information to enable them to pick the referent. It does not require an appeal to the speaker's internal representation to work out what is meant. Therefore, even the youngest children should be able to answer this question correctly if they have a robust understanding of how utterances relate to reality. However, it could be that children's tendency to focus on the utterance-thought relationship results in them being less able to make accurate judgments about the listener's knowledge state. That is, children might think that if the speaker's internal representation is not identical to the utterance, the listener would not be able to pick the correct item from the array. This would suggest that children have a hitherto unknown difficulty in their ability to handle the relation between utterance and array, which is not confined to utterance evaluation.

Experiment 3.3.

3.3.1. Introduction

Experiment 3.3 focuses on the Unambiguous condition only since it provides an opportunity to see whether there is a tendency to (incorrectly) judge negatively about a listener's knowledge state when a thought bubble is present. If

they did, it would suggest they overly attend to discrepancies in speech and thought even when assessing the utterance's pragmatic adequacy.

In Experiments 3.1 and 3.2, children were presented with textual thought bubbles only. In these experiments, thought bubble content was presented in written, propositional form to be consistent with speech, whereas Wellman et al. (1996) assumed that thought content should be pictorial. Given that the speech bubble had no effect on children's judgments in Experiments 3.1 and 3.2, the requirement of consistency between thought and speech no longer existed. In Experiment 3.3, children, therefore, received thought bubbles in either written or pictorial format.

3.3.2. Method

3.3.2.1. Participants. Four groups of children were included: 24 Year 1 children (14 males, 10 females; mean age, 6:0, range 5:7 to 6:7), 24 Year 2 children (13 males, 11 females; mean age, 6:10, range 6:8 to 7:5), 26 Year 3 children (10 males, 16 females; mean age, 8:1, range 7:7 to 8:7) and 26 Year 5 children (8 males, 18 females; mean age 10:0, range 9:7 to 10:7). All attended a LEA maintained Primary School in Derbyshire.

3.3.2.2. Materials. These were identical to Experiment 3.2, except that each story scenario was adapted so that the final picture could either contain a textual or pictorial thought bubble.

3.3.2.3. Design and Procedure. All children received Introductory tasks, identical in format and procedure to Experiment 3.2, except that half of the

children saw pictorial thought bubbles and half saw textual thought bubbles. Children then received four story scenarios, all of which were Unambiguous and contained the speaker's utterance in a speech bubble. Children received two stories with a thought bubble and two without. The order of Thought Bubble Absent/Present trials was alternated for each participant and systematically varied between participants. Thought Bubble Present/ Absent was therefore a Within subjects factor and Bubble Type (words or pictures) was a Between subjects factor. A within-subjects design was employed in order to keep the sample size to a manageable level. The order of story scenarios was fixed across all children. After the presentation of each story, children were asked the following questions in fixed order[1]:

Evaluation: Did (speaker) tell (listener) exactly which object was his/he wanted?

Justification: Why/why not?

Knowledge: Will (listener) pick the right (object)?

3.3.3. Results

Children received a score of 1 every time they said 'no' to either test question; indicating that the speaker did not say exactly which item he/she wanted (Evaluation question) or that the listener would not be able to pick the correct referent (Knowledge question). Data for the Evaluation and Knowledge questions

1. Ackerman (1993) emphasizes the importance of presenting an evaluation question before children are asked to make a referent choice so that "...children's knowledge evaluation [is] unconfounded by response decision processes involved in choosing a referent" (p.71). Test questions were, therefore, asked in fixed order to gain a more reliable estimate of the way in which children evaluate the utterance.

were analysed separately in a 4 (Year Group) x 2(Bubble Order: present or absent first x 2(Bubble Type: Words or Pictures) x 2 (Thought Bubble: present or absent) mixed design ANOVA. The first three factors were between classifications and the final factor was a within classification.

3.3.3.1. Evaluation question. There were no significant main effects for Year Group, Thought Bubble (present or absent) or Bubble Type (words or pictures). A Thought Bubble x Year Group interaction narrowly missed significance [$F(3,84) = 2.56$, $p = 0.06$]. However, given the pattern of results from Experiments 3.1 and 3.2, it was important to examine whether the tendency to say 'no' more often in the presence of a thought bubble was evident in any of the year groups independently. Pairwise comparisons were performed between Thought Bubble Absent and Thought Bubble Present scores for each year group. There was a significant difference between scores for the 7-8-year-olds in the predicted direction [$t(25) = 2.13$, $p<0.05$]. That is, the Year 3 children were significantly more likely to say that the speaker did not say exactly which item they wanted in the presence of a thought bubble (mean = 0.61, from a total of 2, SD = 0.75) compared to when a thought bubble was absent (mean = 0.31, SD = 0.62). Mean scores for the other three Year Groups did not differ significantly according to whether the thought bubble was present or absent. Post-hoc Tukey's comparisons showed that the Year Groups did not differ significantly from each other in either the Thought Bubble Present or Thought Bubble Absent conditions. Thought Bubble Present or Absent did not interact significantly with Bubble Type. There

was a significant Thought Bubble x Bubble Order interaction [$F(1,84) = 8.81$, $p<0.005$], which is illustrated in Figure 3.7.

Children were more likely to say 'no' when the first story they saw contained a thought bubble (mean = 0.66, SD = 0.66) compared to children who saw a story without a thought bubble first (mean = 0.27, SD = 0.66) and this effect was confined to the Thought Bubble Present condition. Mean scores in the Thought Bubble Absent condition were very similar for Bubble Absent first (mean =0.43, SD = 0.71) and Bubble Present first (mean = 0.48, SD = 0.71) trials. The effect for the Thought Bubble Present condition also maintained independently [$t(1,99) = 8.55$, $p<0.01$]. Bubble Order did not interact with Year Group, or with any other factor. There were no other significant interactions between any of the variables.

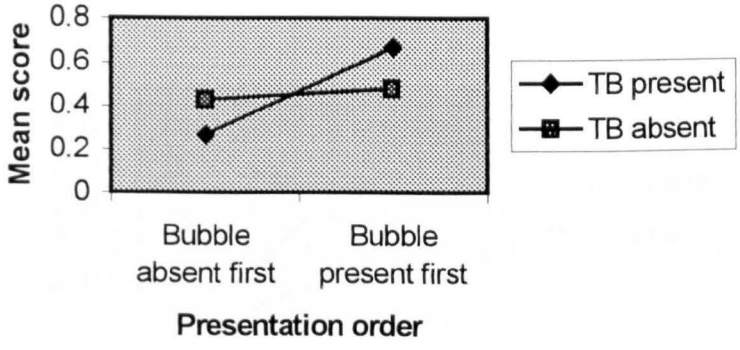


Figure 3.7: Mean scores for the Evaluation question (saying 'no') according to presentation order of trials, and presence or absence of Thought Bubble.

3.3.3.2. Knowledge question. The pattern of results for the Knowledge question was very similar to that found for the Evaluation question. There were no

significant main effects for Thought Bubble (present or absent), Year Group or Bubble Type. However, there was a significant Year Group x Thought Bubble interaction [$F(3,84) = 4.38$, $p<0.01$; see Figure 3.8].

Pairwise comparisons between Thought Bubble Absent and Thought Bubble Present scores were performed to determine whether mean scores differed for each Year Group independently. The 5-6-year-olds were significantly less likely to say that the listener would be able to pick the referent in the absence of a thought bubble (mean = 0.50, SD = 0.78) compared to when a thought bubble was present (mean = 0.25, SD = 0.53) [$t(23) = 2.30$, $p<0.05$]. The 6-7-year-olds showed a similar trend in responses (Absent mean = 0.37, SD = 0.57; Present mean = 0.17, SD = 0.38), but their scores did not differ significantly. By contrast, the 7-8-year-olds were significantly more likely to say that the listener would not be able to pick the referent in the presence of a thought bubble (mean = 0.46, SD = 0.65) compared to when a thought bubble was absent (mean = 0.19, SD =0.57)

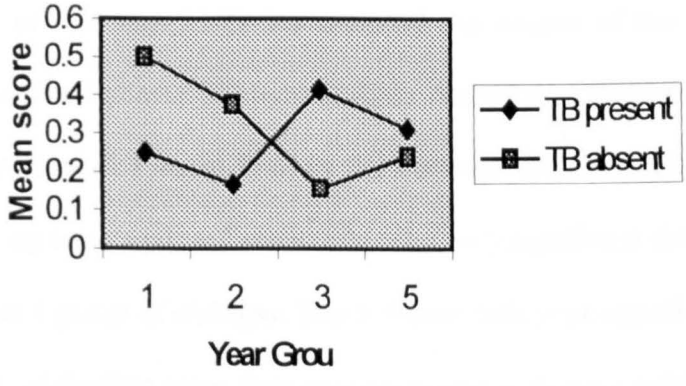


Figure 3.8: Mean scores for the Knowledge question (saying 'no') according to Year Group and presence or absence of Thought Bubble.

[t(25) = 2.27, p<0.05]. There was a trend for the 9-10-year-olds to answer in the same way (Absent mean = 0.23, SD = 0.51; Present mean = 0.31, SD = 0.47) but, again, this difference between scores was not significant. There was a significant main effect (between subjects) for Bubble Order, which was qualified by a Year Group x Bubble Order interaction [F(3,84) = 3.59, p<0.05; see Figure 3.9].

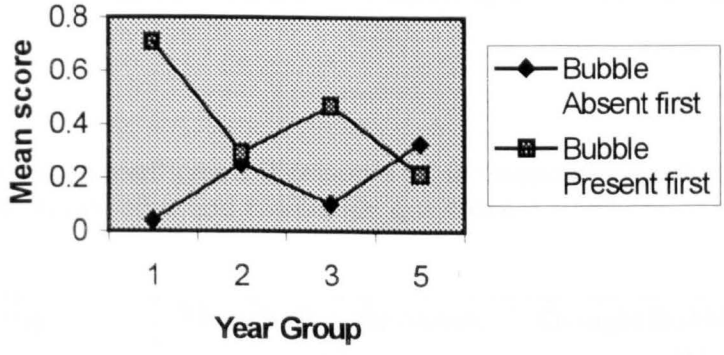


Figure 3.9: Mean scores for the Knowledge question (saying 'no') according to Year Group and presentation order of trials.

A series of one-way ANOVAs revealed the source of the interaction. Scores were combined across Thought Bubble Absent and Present trials, with Bubble Order (Absent or Present first) as the between subjects variable. Scores for each year group were analysed separately. The only significant difference was found in the Year 1 group of children. The 5-6-year-olds were significantly more likely to say 'no' if the first story they saw contained a thought bubble (mean = 1.42, SD = 1.44) compared to children who saw a story without a thought bubble first (mean = 0.08, SD = 0.29) [F(1,23) = 9.84, p<0.01].

Analysing the data for each Bubble Order separately across all four Year Groups, showed that there were no significant differences between any of the groups for either the Thought Bubble Absent first or Thought Bubble Present first trials. There were no other significant main effects or interactions.

3.3.3.3. Correlations between the Evaluation and Knowledge questions. Correlational analyses were performed to see whether children answered the two test questions similarly. A summary of results is contained in Table 3.4.

Table 3.4: Correlation coefficients (Pearson's r)* for associations between responses on the Knowledge and Evaluation questions.

Year Group	Thought Bubble Absent (r)	Thought Bubble Present (r)
1	0.43[a]	0.89
2	0.18[b]	0.41[a]
3	0.74	0.87
5	0.66	0.69
All	0.72	0.48

*All coefficients significant at $p<0.001$ except:
[a] Significant at $p<0.05$
[b] Not significant

Across all Year Groups there was a strong positive correlation between responses on the two questions, for both the Thought Bubble Absent and Present trials. This suggests that if participants answered the Evaluation question negatively they also tended to answer the Knowledge question negatively. This trend maintained for each Year Group independently in the Thought Bubble

Present trials and for all but the Year 2 children in the Thought Bubble Absent trials.

3.3.3.4. Justifications of responses. Data were coded in exactly the same way as in Experiment 3.2. There was a 92% agreement rate between the blind coder and the experimenter. Table 3.5 contains a summary of the types of justifications provided for each year group. There was no significant effect of bubble type on responses to the test questions, so justification data are combined across words and pictures.

Children showed an increasing tendency with age to say that the speaker did say exactly which item they wanted (e.g.) 'because there is only one blue one'. That is, children tended to justify their positive evaluations by making reference to the one item in the referent array that matched the speaker's utterance. This trend was found in the Thought Bubble present and absent trials, with around 33–37% of the 5-6-year-olds mentioning the array, moving through to 85-90% of the oldest group responding in this way. The only noticeable deviation from this trend was amongst the Year 3 children in the Thought Bubble Present trials. The 7-8-year-olds were significantly more likely to evaluate negatively in the presence of a thought bubble compared to when it was absent. This is reflected in the types of justifications provided: 62% (10/16) of Year 3 children who judged negatively said that the speaker had not provided enough information about the referent, compared to only 25-33% in the other three groups.

Table 3.5: Total number of justifications (and percentages)[a] for the Unambiguous condition according to Thought Bubble present or absent.

	Number of responses (%ge)	1 Multiple objects	2 Single object	3 Not enough info	4 Thinking it/not saying it	5 Listener's mental state	6 No answer/ don't know	7 Hypotheti-cals	8 Other	Totals
	Year 1									
	Yes[b]	0	13 (33)	0	0	7 (18)	7 (18)	1	11 (28)	39
	No[b]	2	0	1	0	2	1	2	2	10
TB Absent	*Year 2*									
	Yes	2	21 (66)	1	0	1	3	0	4	32
	No	2	1	7	0	0	2	2	1	15
	Year 3									
	Yes	0	37 (84)	0	0	4	0	3	0	44
	No	0	0	1	0	0	0	5	2	8
	Year 5									
	Yes	0	37 (90)	0	0	1	0	0	3	41
	No	1	0	5	0	0	2	3	1	12
	Year 1									
	Yes	0	15 (37)	0	0	5	7 (17)	2	11 (27)	40
	No	1	0	2 (25)	2	0	2	1	0	8
TB Present	*Year 2*									
	Yes	0	24 (67)	0	0	1	5	0	6	36
	No	4	1	4 (33)	0	1	1	0	1	12
	Year 3									
	Yes	0	25 (69)	1	0	3	1	1	5	36
	No	3	0	10 (62)	0	0	0	3	0	16
	Year 5									
	Yes	0	34 (85)	0	0	0	1	2	3	40
	No	2	0	4 (33)	0	1	1	3	1	12

[a] Combined across words and pictures. [b] Yes or No indicates response to the Evaluation Question.

3.3.4. Discussion

Unlike Experiments 3.1 and 3.2, there did not seem to be a general tendency to evaluate negatively when asked whether the speaker had said exactly which item he/she wanted. Nevertheless, the 7-8-year-olds did demonstrate a significant tendency to respond in this way – irrespective of whether the bubble appeared in text or picture format. Moreover, their tendency to focus on the utterance-thought relationship was supported by the types of justifications provided for their negative evaluations in the Thought Bubble Present condition. Almost two-thirds of these justifications made reference to the fact that the speaker did not give enough information, or did not say what was included in the thought bubble. The failure to find a general effect of judging negatively in the presence of a thought bubble could be due to the alteration in experimental design; from Thought Bubble present or absent as a between-subjects factor in Experiment 3.2, to a within-subjects factor in Experiment 3.3. It is possible that being able to contrast the two conditions directly in the second experiment enabled children to see that the information in the thought bubble was redundant, and therefore did not need to be considered in their utterance evaluations. Some support for this is provided by the Thought Bubble x Bubble Order interaction found for the Evaluation question. This showed that children, across all year groups, were significantly more likely to say 'no', in the Thought Bubble Present condition, if the first story they saw contained a thought bubble compared to children whose first story did not contain a thought bubble. In short, children

seemed to be responding in a similar way to those in the between-subjects design of Experiment 3.2.

Even with the opportunity to contrast conditions in a within-subjects design, the 7-8-year-olds continued to judge negatively on the Evaluation question significantly more often in the presence of a thought bubble. Additionally, the 7-8-year-olds were the only group who were significantly more likely to say that the speaker would not be able to pick the correct referent when they were privy to the speaker's internal representation. This result is not threatened by the significant Year Group x Bubble Order interaction, since the tendency to judge negatively if a thought bubble present story was presented first, rather than second, was confined to the 5-6-year-olds in Year 1. At least for the 7-8-year-old children, then, there seemed to be a difficulty in making accurate judgments about how well the utterance mapped on to the array, in terms of allowing a unique referent to be identified. It seems that the 7-8-year-olds focused heavily on the utterance-thought relationship at the expense of the utterance-array relationship. That is, they were unable to dissociate what they knew about the discrepancy between the utterance and the speaker's internal representation, from how well the utterance adequately specified a particular referent in the array. In short, knowledge of the speaker's representation seemed to interfere with children's ability to evaluate and interpret Unambiguous utterances.

By contrast, the 5-6-year-olds seemed to be helped by the presence of the thought bubble to make correct judgments about how the listener would be able to interpret the utterance. They were significantly more likely to say that the listener

would be able to pick the right referent when a thought bubble was present rather than absent. This seems to be consistent with the idea that younger children are more likely to judge utterances positively when they know the speaker's meaning (Beal & Flavell, 1984; Beal & Belgrad, 1990). Compared to these earlier studies, however, the novel finding is that this tendency was found in Unambiguous messages. This suggests that there may be at least some difficulty at this age with understanding how an utterance can be sufficiently Unambiguous in terms of allowing a unique identification of a referent in an array. It might be that younger children prefer to have more information about the referent before they can decide whether the utterance is 'good enough'. This is in line with previous research demonstrating that 4-5-year-olds tend to confuse quantity with quality and judge longer (but incomplete) messages to be more useful than shorter (but Unambiguous) ones (Sonnenschein, 1984). It seems that this tendency extends slightly beyond the age of 5 years though – at least for the children in the present sample.

Finally, even though there was a general tendency amongst many children in Experiment 3.3 to judge positively, there was still a noticeable developmental trend in the types of justifications given for the Evaluation question. The youngest children rarely mentioned the 'goodness of fit' between the utterance and the array when justifying their positive evaluations, whereas the oldest children almost always did so. Again, as in Experiment 3.2, this is suggestive of a gradually developing understanding of the pragmatic importance of referential

communication, rather than a conceptual leap in this knowledge at around the age of 6 or 7 (cf. Olson, 1981; cited in Bonitatibus, 1988).

3.3. General Discussion

The aim of this research was to investigate how children handle the triangular relation between an utterance, a speaker's meaning and reality when the speaker's internal representation (meaning) was explicitly presented in a thought bubble. In Experiments 3.1 and 3.2, children aged 5 to 10 years were shown stories in which a protagonist's utterance was either Ambiguous or Unambiguous. In support of earlier work by Beal and colleagues (Beal & Flavell, 1984; Beal & Belgrad, 1990), being informed of the speaker's meaning resulted in children overlooking the relationship between the utterance and the array. This was evidenced by children being much more likely to make reference to the utterance-meaning relation in their justifications for the Ambiguous condition. Moreover, this tendency was also found in the Unambiguous condition, leading many children to judge that a sufficiently informative utterance was inadequate. It would appear that whether speaker meaning is construed as intended outcome or internal representation, children are influenced by their knowledge about meaning and tend to disregard the pragmatic fit of the utterance to the array. Where the present results differ from Beal's findings however, is that children *older* than about 6 years continue to experience difficulty making evaluations about literal meaning, when the speaker's meaning is depicted as an internal representation via a thought bubble. This might be because, as Robinson and Whittaker (1987)

suggest, understanding speaker meaning as an internal representation may be a later developing ability than understanding meaning as intended outcome. Therefore, even 9-10-year-olds may struggle at times to understand how the speaker's internal representation relates to the utterance, the array and the success of the communication.

Arguably, it might be perfectly reasonable to expect participants to interpret the evaluation question as a request to compare the information in the thought bubble with the information in the utterance, resulting in the observed neglect of the utterance-array link. However, adults almost never judged in this way in the Unambiguous condition, preferring a pragmatic interpretation of the question instead (saying that the speaker had said exactly which item he/she wanted). This suggests that adults mostly interpreted the question as an evaluation of how well the utterance singled out an item in the array, rather than how fully the utterance characterized the speaker's meaning. Moreover, the 7-8-year-olds in Experiment 3.3 also tended to respond negatively to the Knowledge question. This suggests that children's difficulty taking a pragmatic perspective was not simply a result of interpreting the evaluation question in a peculiar way. Instead, there seemed to be a more general difficulty understanding how the speaker's internal representation influences the communication process. One limitation to this explanation is that the evaluation question always preceded the knowledge question. Thus, the 7-8-year-olds may have shown a significant tendency to respond negatively to the knowledge question because they had already evaluated the utterance negatively. The strong, positive correlations obtained between

responses to the two questions are consistent with children adopting such a strategy. Question order was fixed to avoid any influence that choosing a referent might have on children's evaluations (cf. Ackerman, 1993; Note 1), so it would not have been helpful to incorporate counterbalancing in the design. Instead, a further experiment is required that asks children the knowledge question only. If they continue to judge that the listener would not be able to pick the referent, even though the utterance is sufficiently informative, this would provide stronger evidence that there is a particular difficulty in responding pragmatically when the speaker's meaning is directly presented in a thought bubble.

Nevertheless, it might still be that the present results are congruent with a conceptually based explanation. In particular, these findings do not seem to fit comfortably with Olson's (1970) or Bonitatibus' (1988a) account of how children come to solve the problem of referential ambiguity. Both authors argue that the crucial component in understanding how an utterance may be inadequate, is the recognition of the importance of literal meaning, over and above what the speaker may mean. That is, since the literal meaning of a referential utterance is dependent on the context of alternatives available in the array (Olson, 1970), children need to understand that it is the utterance-reality link which determines how good (or not) an utterance might be. For example, Bonitatibus (1988b) reports that children capable of detecting referential ambiguity tended to request the speaker's words as a 'clue' to help them diagnose the cause of communication failure, whereas children who failed to detect ambiguity were more likely to request knowledge of the speaker's meaning. It should follow, then, that children around the age of 6 or

7 years would prefer to base their utterance evaluations on the link between utterance and reality rather than on the relationship between utterance and meaning, and this should be true of Unambiguous as well as Ambiguous utterances. This was not the case for many children in the present experiments, who tended to focus on utterance-meaning instead, both in the Ambiguous and Unambiguous conditions.

Although the general tendency for children aged 6-10 years to answer negatively in the presence of a thought bubble was not replicated in Experiment 3.3, the 7-8-year-olds did respond in this way. This discrepancy in results could be due to the within-subjects design of Experiment 3.3 (compared to the between-subjects design in Experiments 3.1 and 3.2), or to sample differences. Nevertheless, the finding that the 7-8-year-olds responded in the same way as the children in Experiments 3.1 and 3.2 *despite* such differences, suggests that there may be something particularly interesting occurring in the child's thinking around this age.

One possibility is that children apply an overly rigid mentalistic criterion for interpretation such that they overlook the utterance-reality relationship. That is, once children come to realise that both the literal and speaker meanings are important components of the communicative act they then seem to have difficulty understanding the value of the utterance in the context of the array. It seems that the array is not the main reference point for judging the adequacy of utterances, at least around the age of six or seven years. Related evidence comes from an allied 'theory of mind' domain in the form of reasoning about false beliefs. Mitchell,

Robinson, Isaacs and Nye (1996) found that adults, but not 5- or 9-year-old children, were significantly influenced by their own knowledge of whether an utterance was true or false when judging whether a listener protagonist would believe a message. Participants were presented with story scenarios in which they (the participants, but not the listener protagonist) were given privileged information about the truth or falsity of a speaker's utterance about the contents of a jug. Adults tended to judge that a message would be believed by a listener protagonist if they themselves knew that the message was true (and the listener was known to have no independent information regarding the truth of the message), and disbelieve a message that they knew to be false. By contrast, children were not influenced by their own knowledge in judging whether or not a listener would believe an utterance. Mitchell et al. argue that children were able to make 'uncontaminated' judgments about belief because they rigidly applied a newly learned rule that 'seeing leads to knowing' (also Sodian & Wimmer, 1987). The adults, instead, tended to adopt a more sophisticated role-taking perspective and tried to interpret the message from the perspective of the listener. This led them to believe that it would have been difficult to reach any other conclusion than the one they knew to be true.

The main message from Mitchell et al. (1996) was that adults seemed to be influenced by a 'reality bias' in their judgments whereas 5- and 9-year-old children were not. However, this is not the same as saying that children are immune from the influence of reality at this age, but rather that any decline in the relative salience of reality is overshadowed by the application of a newly learned

rule. Generally, the authors argue for a gradually developing understanding of the mind that occurs due to an '...important decline [throughout childhood and into adulthood] in the tendency to focus inappropriately on reality in preference to belief' (p.18). The difficulty for the child might occur in deciding when to apply reality as the criterion for making judgments (as in the case of referential communication) and when not to (in the case of judgments about beliefs).

Ironically, the adults more pragmatic, reality-based strategy may steer them away from negative ('incorrect') evaluations in the present procedure. It might be that the thought bubble procedure encourages negative judgments because a criterion for evaluation (the say-mean discrepancy) is suggested to the children (cf. Mitchell & Robinson, 1994). This may override any natural tendency to think about the pragmatic importance of the utterance. Adults may not be so susceptible to this cue either because they have a more robust understanding of the significance of the pragmatic relationship in referential communication, or because, more generally, they tend to be influenced by what they know about 'reality' (the state of the array).

3.4. Conclusion

Overall, children's understanding of literal meaning (the utterance-array relationship) may not be robust enough to withstand knowledge about the speaker's meaning. In particular children might find it difficult to consider the relationship between the utterance and the array when they know that what is said is discrepant with what is meant. It seems that children may find it especially hard

to understand how an utterance can be adequate even though thought and speech are not identical. This suggests that children's understanding of how the elements of the referential triangle contribute to the formation of utterances undergoes further development beyond the age of 6-7 years.

CHAPTER 4

When do children evaluate utterances according to the speaker's belief? Developments from middle childhood to adulthood.

4.1. Introduction

The experiments presented in Chapter 3 suggest that children as old as 10 years often overlook the relationship between an utterance and an array when a speaker's internal representation is shown directly via a thought bubble. Instead, children tended to focus on the relationship between what the speaker said and what the speaker meant. However, participants might respond quite differently in situations in which the speaker's internal representation is not shown directly in a thought bubble. Therefore, the following experiments explore whether children focus on the speaker's utterance-belief relationship when the speaker's internal representation is not presented pictorially. It is important to see to what extent children, and adults, include the speaker's meaning in their evaluations when an internal representation can only be inferred from the information provided in a story.

4.2. Background literature.

The tendency for children to focus on the speaker's underlying meaning or intent before they come to appreciate the literal meaning of an utterance has been summarised in Chapter 3. According to the wealth of research into referential ambiguity, children can only begin to appreciate literal meaning at around the age of 6 years. At this age, children can reflect on the utterance as an object of scrutiny in it's own right and begin to realise that the utterance is crucial for determining communicative success. However, it is also vital for the child to understand that an utterance is not simply a faithful copy of reality, but is a representation of a speaker's knowledge, thoughts or beliefs. Therefore, not only must the relation between the utterance and the array be taken into account but so too must the speaker's conceptual framework.

As Robinson and Whittaker (1987; and Robinson & Mitchell, 1992) point out, it is difficult to determine in standard ambiguity tasks whether the child participant is more concerned with utterance-array relations or utterance-meaning relations. This is because the speaker's internal representation is always isomorphic with the desired object in the array. The speaker knows what the array looks like, and they have only one object in mind, but may underspecify the referent. As a result, it is impossible to tell whether children are judging about ambiguity on the grounds that the utterance does not uniquely identify an object in an array, or whether there is a discrepancy between what is said and what is meant. One of the advantages of the thought bubble paradigm used in Chapter 3, is that it was possible to examine the extent to which children did include the

speaker's meaning in their utterance evaluations – even though the speaker's internal representation and the array were identical, as in traditional ambiguity research. The main finding from Chapter 3 was that children older than 6- or 7-years tended to concentrate on the utterance and the speaker's representation, rather than on the relationship between the utterance and the array.

Another way to tackle the issue of whether children attend to the speaker's internal representation when evaluating and interpreting utterances, is to present situations in which the speaker's meaning and the state of affairs in the real world are no longer identical (Robinson & Whittaker, 1987). It then becomes possible to investigate which of these elements children are more concerned with when responding to problematical messages. Mitchell and Russell (1989, 1991; Mitchell, Munno & Russell, 1991) utilised and improved a task developed by Ackerman (1979). The basic task involves a speaker protagonist who provides an utterance that mismatches a referent in an array. The mismatch occurs because the speaker's internal representation of the object is faulty in some way. In other words, the speaker's belief or meaning is discrepant with the appearance of the object in the real world, whilst the speaker's utterance is an adequate characterisation of what they believe.

For example, children in Mitchell and Russell's (1989) procedure were told a story about John who went to find a book that Mary had asked for. Mary said that the book had a dog on the cover, but the book John found actually had a picture of a cat on it. Some children were told that Mary had a poor memory whilst others were told that Mary's memory was perfect. In the first two

experiments from that study, children were asked whether the book was the one the speaker had intended. 5- and 6-year-olds very often said that it was not, but nevertheless judged that the listener would return to the speaker with the book. Therefore, the 5- and 6-year-olds seemed to be responding on the basis that if the speaker's description did not exactly match the found object, then the speaker did not actually *mean* that object. They would return with the book because they deemed it to be a good enough alternative. In other words, children did not seem to understand that even though the relationship between the utterance and reality could be discrepant, the utterance was, nevertheless, adequate from the speaker's point of view. Instead, children tended to focus on the discrepancy between the utterance and the object in the real world. Moreover, the 5-6-year-olds were not influenced by the information pertaining to the speaker's memory, '...they appeared not to recognize the connection between the informational state of the speaker and the accuracy of his/her ensuing description' (p.488). Generally, this group of 5-6-year-olds seemed to be more focused on utterance-world relations rather than utterance-meaning relations.

However, evidence from Robinson and Mitchell (1992, 1994) suggests that even 4-year-olds are adept at including the speaker's representation in their judgments whilst also recognising the separate literal meaning of a message. In their 'message-desire discrepant task' children were presented with a scenario in which Mum put her two different bags of material into two differently coloured drawers. Whilst Mum was absent from the room, Jane swapped the bags around so that they were in the opposite drawer to their original location. Mum, sewing in

the other room, asked Jane to fetch the bag in the blue drawer. Children were asked which bag Mum really wanted. As many as 50% of the 4-year-olds correctly said that Mum really wanted the bag that was now in the red drawer. They made a correct nonliteral interpretation of the utterance, which required processing the literal meaning of the utterance first and recognising that this was a product of Mum's false belief. That is, many of the 4-year-olds seemed to understand that utterances consist of a literal meaning and intended meaning; and they based their judgments on Mum's *representation* of the state of affairs rather than simply on the relationship between the utterance and the array.

There seems to be a disparity between the Mitchell and Russell (1989; Experiments 1 and 2) results on the one hand and the Robinson and Mitchell (1992, 1994) findings on the other. Both approaches presented children with scenarios in which the speaker's belief was discrepant in some way with reality, and both approaches required children to make an *interpretation* of the resulting discrepant utterance. However, the former studies indicated that 5-6-year-olds tend not to incorporate the speaker's intended meaning or belief state into their judgments, whereas the latter studies provided evidence that four-year-olds often utilised the speaker's representation as an aid to interpretation. It is arguable, however, whether the kind of interpretation required in both approaches is equivalent.

In the Robinson and Mitchell studies, children could simply point to one of two (or three) locations in response to the question 'Which bag did Mum really want?' What they needed to do was work out which bag matched Mum's (false)

belief. In other words, it is possible to make a correct response on this task by considering the relationship between belief and reality. The speaker's utterance and belief are the same and so a simple match between belief and reality is all that is required to give a correct answer.

By contrast, children needed to make a tacit *evaluation* of the utterance in the Mitchell and Russell studies when they were asked 'Is the book John found really the book Mary had been thinking about?' This question does not simply require a child to figure out which book matches Mary's belief since there is no alternative book, and there is no book in the real world that matches Mary's belief exactly. Instead, the child needs to make some kind of evaluation of *how close a match* the found object is to the one the speaker had in mind. This is more suggestive of a situation in which the referential triangle needs to be considered. The words of the utterance need to be compared to the speaker's belief and to the array because there is no direct match between belief and reality, or between the utterance and reality. All three components of the triangle need to be considered in order to work out the best answer to the question. It is worth noting that when children's understanding of the size of a discrepancy between what was said and the object in the real world was investigated directly, 5-6-year–olds, but not those aged 8-13 years, judged that communication would be successful irrespective of whether an utterance was slightly or grossly discrepant (Mitchell, Munno & Russell, 1991; Mitchell & Russell, 1991). Therefore, it seems that when judging about discrepant utterances, *evaluation* may be harder than *interpretation.*

Mitchell and Robinson (1994) investigated this possibility directly, using their message-desire discrepant (MDD) paradigm summarised above. In Experiment 1, children aged 4-7 years were presented with the MDD condition as well as a standard ambiguous condition. Following the stories, children were asked an evaluation question: 'Did (speaker) do a good job of saying which (item) she really wants?', and an interpretation question: 'Can you point to the (item) that (speaker) really wants?' Children of all ages were significantly more likely to interpret the utterance correctly in the MDD task (point to the item in the unspecified location), than they were to say that the speaker had done a bad job. In addition, performance on the MDD task and the ambiguous condition was significantly correlated in terms of types of evaluations. The younger children tended to evaluate positively (good job) on both occasions, whereas the older children evaluated negatively both times. Therefore, even though the older children could interpret the utterance on the basis of the speaker's representation, they tended to evaluate the utterance on the basis of a discrepancy with reality.

As Mitchell & Robinson (1994) suggest, though, evaluating a discrepant utterance according to the speaker's internal representation might be a sophisticated criterion, which may only be available to children older than seven years. Children could judge that the speaker had done a good job on the grounds that the utterance is unambiguous vis a vis the speaker's representation. By contrast, younger children judge the speaker did a bad job because the utterance is discrepant vis a vis current reality. In other words, there might be a U-shaped development in which children initially evaluate positively because they have

limited understanding of the role of the literal meaning of the message in referential communication. Next, they evaluate negatively because they notice that the utterance is discrepant with reality, and finally they adopt a sophisticated criterion of judging according to the speaker's internal representation and, therefore, evaluate the utterance positively once again. Older children were not included in Mitchell and Robinson (1994), and this issue was not investigated directly, so it is not possible to tell what criteria children use for utterance evaluation as they get older.

Nevertheless, Experiments 2 and 3 in Mitchell and Robinson (1994) indicated that children could evaluate discrepant utterances according to a more mentalistic criterion if they were cued into thinking that there was a discrepancy between what was said (and how that related to reality) and what was meant. Instead of the 'good job' question asked previously, children were asked 'Did (speaker) say the right thing for the (item) she really wants?'. The authors argue that this question alerts children to the idea that what is said is a clue or characterisation of what is meant and so they are able to focus on the say-mean discrepancy as the basis of their evaluation. Children with a mean age of 5 years showed no difference between interpretation and evaluation when the 'right thing' question was included. In addition, when the 'right thing' and 'good job' questions were directly compared in a within-subjects design, the 'right thing' question promoted significantly more correct (negative) evaluations than the 'good job' question. The main point is that, like with the Mitchell and Russell (1989) study described above, children were significantly better at making

judgments based on a consideration of the say-mean discrepancy when they were provided with clues about the appropriate criterion for evaluation.

It could be argued that the children in the thought bubble experiments in Chapter 3, were alerted to the possibility that there was a discrepancy between what was said and what was meant and so used this as the basis for their judgments, even when it was inappropriate to do so. In other words, children in those experiments were very good at recognising a mismatch between message meaning and intended meaning because it was clearly presented in front of them. Generally then, there are many examples of children incorporating the speaker's representation into their judgments when they are helped to do so by the provision of cues as to the 'correct' criterion for evaluation. However, the extant literature does not inform us about whether children spontaneously *do* evaluate utterances according to the speaker's representation when strong cues as to an appropriate criterion are absent.

In the Mitchell and Robinson (1994) study mentioned earlier, children around the age of seven years judged that the speaker had done a 'good job' when utterances were discrepant or ambiguous. As the authors note '...the "good job" question... leaves the onus with the child for identifying the appropriate criterion for evaluation' (p.1221). However, at least in the MDD task, it is not clear whether children were judging 'good job' because they had little understanding of the utterance as a separate object for evaluation, or because they noticed that the utterance was good according to the speaker's internal representation. The aim of the following experiments was to investigate this issue directly. The 'good job'

question was used since it allows participants to choose their own criterion for evaluation. This is important because everyday situations do not always provide clear clues about whether we should be wary of the possibility of a say-mean discrepancy occurring (and, therefore, we should be more mentalistic in our approach).

A vital component of understanding of the mind, and, ergo, being a successful communicator is the knowledge that utterances are things that are based on the speaker's mental model of reality, rather than on reality itself. As Mitchell (1999) points out 'the ability to be informed by an utterance often depends on an ability to take into consideration the speaker's conceptual framework' (p.191). Evaluating utterances according to a more mentalistic criterion might, therefore, be an important skill in terms of understanding the referential triangle and how its various components contribute to communicative acts. Specifically, asking participants to evaluate utterances provides an opportunity to explore how they handle literal meaning in relation to the speaker's representation and the array. It allows an explicit look at which aspect(s) of the referential triangle participants assign the greatest value. This is in contrast to an interpretation question which could be answered based on more simple, linear relationships as described on pages 109-110 above. It is unclear at this stage when we might expect this ability to appear developmentally and so older children, as well as adult participants, were included to see whether they would spontaneously evaluate according to the speaker's internal representation.

The following methodology was based on a task from Mitchell, Robinson and Thompson (1999). The 'superficially ambiguous' task involves a speaker protagonist who sees an item placed in location x. They leave the scene and, in their absence, more items are added to the array. The speaker asks for the item in location x not knowing that their utterance is now ambiguous. In Mitchell et al., children aged 4 to 6 years were asked to *interpret* the utterance by pointing to the item that the speaker wanted. In order to do this correctly, children needed to make reference to the speaker's belief to choose the right item. Even the youngest children were able to make correct interpretations of the utterance, suggesting that they had few difficulties acknowledging the importance of the speaker's belief in this context.

Children were not asked to *evaluate* the utterance, though, which has already been argued to allow great scope for participants to consider all three elements of the referential triangle. In the 'superficially ambiguous' condition, children can either evaluate according to the utterance-array relationship (objective criterion) or according to the utterance-meaning relationship (mentalistic criterion). If they evaluate according to the former, they would say that the speaker had done a 'bad job' because the utterance is discrepant with the current state of reality. However, if they take into account the speaker's belief state, they will say that he/she did a 'good job' because as far as they (the speaker) were concerned, there was only one item in location x. The implication of the results reported in Chapter 3 is that children may concentrate on representation in middle childhood. Additionally, the possibility of a U-shaped development in

terms of utterance evaluation – as suggested in Mitchell and Robinson (1994) – implies the same thing. That is, children in middle childhood may adopt a more mentalistic criterion once they have come to appreciate the role of literal meaning and how this is instantiated by the array (Bonitatibus, 1988a). Therefore, it is expected that only children aged about seven years and older will be able to utilise the more sophisticated mentalistic criterion in their utterance evaluations. Interestingly, Mitchell and Russell (1989) found that it was not until around 14 years that children could consistently base their evaluations on information pertaining to the speaker's internal representation (i.e. whether they had a good or a poor memory). Consequently, it might only be older children, and adults, who tend to consider the speaker's belief state when evaluating utterances under these circumstances. This may suggest that there are difficulties that extend beyond the age of about seven years in understanding how the speaker's internal representation influences communication.

Experiment 4.1.

4.1.1. Introduction

The 'superficially ambiguous' (henceforth, the Not See) condition was compared to a See condition and a Control. In the See condition, the speaker protagonist saw that more items were added to the array and yet still underspecified the referent. This is equivalent to a 'standard' ambiguity task in which the speaker's utterance is an adequate characterisation of the speaker's

internal representation (they have only one object in mind), but the desired item is not described adequately in order to establish unique reference. Therefore, the locus of the problem lies in the relationship between the utterance and the intended referent. In this condition, participants should judge that the speaker did a bad job because the utterance does not uniquely identify an utterance in the array. By contrast, in the Not See condition, the speaker's internal representation is discrepant with the current state of reality, but the utterance is again an adequate characterisation of the speaker's internal representation. Therefore, participants can adopt either a mentalistic or an objective criterion for evaluation. This condition allows a direct investigation of participants' utilisation of the components of the referential triangle: do they assess utterance quality according to representation or reality? The Control condition was included to check that participants were willing to say that the speaker had done a good job when no items were added to the array. In this condition, the speaker's internal representation, ensuing utterance and the state of external reality are all in perfect harmony and so there is no obvious reason to judge that the speaker's utterance was inadequate.

The first experiment reported below is a pilot study involving adult participants. This was completed for two main reasons. Firstly, in order to provide a point of comparison for child participants, it was necessary to see how adults responded in each of the experimental conditions. Secondly, it was important to establish the extent to which adults would incorporate the speaker's belief state in their utterance evaluations when they were given the option of responding

according to either (a) the relation between the utterance and reality or (b) the relation between the utterance and the speaker's belief.

An Interpretation question was included alongside the Evaluation question. This was slightly different to the interpretations required from children in the Mitchell and Robinson (1994) and the Mitchell et al., (1999) studies mentioned earlier. In those studies, children were asked to make an interpretation from their own perspective (and they seemed to have very few difficulties making correct interpretations), whereas the children included in the present procedure were asked to make an interpretation from a listener protagonist's perspective. This provides an extra source of information concerning the criteria children use to process utterances. To expand, if children are focused too heavily on the link between utterance and representation (as in Chapter 3), they might judge good job in the Not See condition, and wrongly assume this implies the utterance is interpretable from the listener's point of view. There was no such problem anticipated with the adult participants, but the following experiment provided a useful opportunity to pilot the question.

4.1.2. Method

4.1.2.1. Participants. 134 second year Psychology Undergraduates took part in the pilot study. There were 97 females and 34 males (3 students failed to provide this information), with a mean age of 20:9 (range 18:1 to 36:5).

4.1.2.2. Materials. All story scenarios were typed in 12pt Times Roman font and illustrated on A4 paper. Each story and accompanying questions appeared on a separate page, making a total of 3 pages for each participant.

4.1.2.3. Design. Participants received a total of three story scenarios, one in each of the following experimental conditions: See, Not See and Control. The order of experimental conditions was rotated between participants and participants were randomly assigned to one of the possible rotations. Condition was mapped on to the type of story scenario, which was fixed across all participants. Participants were asked two main test questions and the order in which these appeared was counterbalanced across subjects and conditions.

4.1.2.4. Procedure. All participants were seated in a lecture theatre, but completed the task individually. They were provided with a handout containing three story scenarios and were told to read the stories in the order in which they appeared. It was stressed that there were no right or wrong answers and that participants should respond according to the answer that they thought best. The three stories involved a boy seeing a bike in a shop window (Bike story); a girl displaying her picture for a school Open Evening (Picture story) and a boy wanting a biscuit from a jar in the kitchen (Biscuit story). The three experimental conditions can be illustrated using the Bike story. At the top of the page was a picture showing the initial state of reality e.g. one bike in a shop window. This was followed by a short story paragraph. In the Control condition participants were told:

'Dave is shopping in town with his friends when they walk past the toy shop. In the window is a new bike, which Dave stops to have a look at.* So, there is only one bike on display.* Later at home, Dave's Mum asks him if he has seen anything he would like to have for his birthday. Dave says "Yes, there was a bike in the toy shop window today. I'd like to have that."

Between the points marked * the Not See condition differed from the Control condition, by including the additional information:

'Whilst Dave walks home, the shopkeeper brings out more bikes to display in the window. There are lots of bikes on display now.'

By contrast, in the See condition, participants were told:

'Whilst Dave watches, the shopkeeper brings out more bikes to display in the window. There are lots of bikes on display now.'

The procedure then differed depending on the order in which test questions were asked. Half of the participants were told: 'The next day, Mum goes to buy the bike...'. On the page, there was an illustration of the state of reality at the time the utterance was made. For the See and Not See conditions, this was a display of three different bikes: the initial referent plus two others. For the Control condition, the picture contained the original bike since no bikes were added to the display. The picture of the bike/s was presented alongside a large, black question mark. This was to provide participants with the opportunity to say 'don't know' in answer to the Interpretation question.

Participants were then asked:

(1) Interpretation question: Which bike will Mum think Dave wants? (Please circle one of the pictured bikes or the '?' if you think Mum won't be able to tell)

(2) Evaluation question: Did Dave do a good job of saying exactly which bike he wants? Yes/No

(3) Justification: Why/why not?

The remainder of the participants were asked the Evaluation question followed by the Justification question. After the presentation of the main story, they were told 'The next day, Mum goes to buy the bike...'; and were presented with the second picture of bikes added (See and Not See) or no bikes added (Control), and a big question mark. Finally, they were asked the Interpretation question. All questions were presented on the same page as the story and participants were required to write down their answers in the spaces provided.

4.1.3. Results

Preliminary analyses showed that there were no significant effects associated with question order or rotation order.

4.1.3.1. Evaluation question: Participants could either say yes or no to whether the speaker had done a good job of saying exactly which item he/she wanted. Table 4.1 contains a summary of the total number of yes and no responses in the three experimental conditions.

Table 4.1: Total number of yes and no responses (and percentages) to the Evaluation question in the Control, See and Not See conditions.

	Good job	Bad job
Control	116 (87)	18 (13)
See	8 (6)	126 (94)
Not See	68 (51)	65 (49)

There were clear differences in patterns of responding between conditions. In the Control condition, participants judged that the speaker had done a good job of saying exactly which item he/she wanted on 87% of occasions compared to only 6% of 'good job' responses in the See condition. Participants rarely said that the speaker had done a 'bad job' in the Control condition, with only 13% of negative evaluations. They were much more likely to judge negatively in the See condition with 94% of responses saying that the speaker had done a 'bad job'. By contrast, responses were more evenly split in the Not See condition, with 51% of participants indicating that the speaker had done a 'good job' and 49% saying that the speaker had done a 'bad job'.

4.1.3.2. Comparisons Between Conditions: A number of multiple comparisons were performed on the data from each year group. Each chi-squared analysis included Yates' correction for continuity, so applying additional corrections (such as the Bonferroni) would be overly conservative.

It was more common for participants to judge 'good job' in Not See and 'bad job' in See (n = 63), compared to the reverse (n = 3) [McNemar χ^2

(corrected, df = 1, n = 66) = 55.4, p<0.005]. By contrast, more participants judged 'bad job' in the Not See condition and 'good job' in the Control (n = 58), compared to judging positively in Not See and negatively in the Control (n = 10) [McNemar χ^2 (corrected, df = 1, n =68) = 32.5, p<0.005]. There was also a very strong trend for participants to judge negatively in the See condition and positively in the Control (n = 112), compared to the reverse (n = 4) [McNemar χ^2 (corrected, df = 1, n = 116) = 98.7, p<0.005].

4.1.3.3. Interpretation question: Responses to this question were coded into three main categories: (1) Referent: indicating that the listener would think the speaker wanted the initial referent shown to participants; (2) Question Mark: indicating that the listener would not be able to tell which referent the speaker wanted and (3) Other: indicating one of the alternative referents in the display. Table 4.2 contains a summary of responses for each condition, according to whether participants answered yes or no to the Evaluation question.

Participants clearly discriminated between conditions in terms of whether the listener would be able to pick the referent that the speaker wanted. Overall, in the Control condition, the majority of responses (95%) indicated that the listener would pick the specific referent. By contrast, the majority of responses in the See (79%) and Not See (75%) conditions indicated that the listener would not be able to tell which item the speaker wanted.

Table 4.2: Summary of responses to the Interpretation question for each condition according to story scenario and response to the Evaluation question.

Story type	Biscuit		Picture		Bike		Total	
Good Job?	Yes	No	Yes	No	Yes	No	Yes	No
Referent								
Control	42	2	35	6	36	3	113	11
See	2	13	1	0	1	1	4	14
Not See	16	8	1	1	0	0	17	9
Can't tell								
Control	0	3	1	0	0	3	1	6
See	1	16	1	45	2	39	4	100
Not see	3	9	25	15	20	27	48	51
Other								
Control	0	0	0	0	0	0	0	0
See	0	9	0	0	0	0	0	9
Not See	1	5	1	0	0	0	2	5

Comparing the number of referent versus nonreferent responses in the See and Not See conditions (see Table 4.3), there was no significant preference to judge that the listener would choose the referent more often in the Not See compared to the See condition [McNemar χ^2 (corrected, df = 1, n = 38) = 2.13, n.s.]. This pattern would have been expected if participants judged message adequacy on the basis of the match between the speaker's utterance and internal representation. However, individual story scenarios seemed to produce markedly different patterns of responding. In particular, the biscuit story seemed to encourage participants to answer in a different way to the other two stories. For example, participants chose the referent a total of 39 times across the See and Not See conditions in the Biscuit story, compared to only 5 times for the Picture and Bike stories combined.

Table 4.3: Total number of referent versus nonreferent responses in the See and Not See conditions.

Condition		See	
	Response	Referent	Nonreferent
Not see	Referent	3	24
	Nonreferent	14	93

4.1.3.4. Justifications – Good Job Question: The justifications to the Good Job question were coded according to eight different categories by the experimenter and a blind rater. Following Camaioni, Ercolani and Lloyd (1998), a sample of 10% of justifications was scored independently by the second coder, with an interrater reliability of 96%[1]. Disagreements were discussed and, in most cases, the experimenter's choice of category was retained. The coding categories were as follows:

(1) Pertaining to multiple objects in the array e.g. 'there are lots of bikes there now.'

(2) Pertaining to one object e.g. 'there is only one biscuit in the jar,' ' there was only one bike there when he saw it.'

1: The method of using a 10% sample to verify the consistency of coding was utilised due to the almost identical coding categories in Chapters 3 and 4, and the high level of agreement between raters, reported in chapter 3. Since the same independent rater coded the justifications in both chapters, rating 100% of the justifications in Chapter 4 was felt to be unnecessary.

(3a) Speaker did not provide enough information e.g. 'she didn't describe it enough.'

(3b) Speaker described the referent adequately e.g. 'it was a clear description.'

(4) Explicit mention of the speaker's mental state e.g. 'he didn't know there were more in the jar.'

(5) Reference to the listener's mental state e.g. 'Dad won't be able to tell which picture is hers.'

(6) No answer/don't know

(7) Invoking hypotheticals e.g. 'there could be lots of bikes in the shop window now.'

(8) Other

Table 4.4 provides a summary of justifications for each experimental condition, according to whether participants answered yes or no to the Evaluation question.

In the Control condition, most of the participants (116/134) answered positively and the vast majority of their positive justifications (83%) made reference to the fact that there was only one object in the jar. In the See condition, the majority of participants (126/134) answered negatively, but their justifications were grouped into two main categories. The modal response for negative evaluations (47%) was to make reference to the multiple objects in the array, and the second most popular response (37%) was to say that the speaker had not provided enough information about the referent. In the Not See condition, just over half of the participants responded positively (68/133) and the majority of their justifications (76%) made reference to the speaker's mental state (i.e. that the

speaker did not know that more items had been added to the display). For those answering negatively (65/133), the modal response was to mention the multiple items in the array (48%), whilst the second most popular type of justification was to say that the speaker had not provided enough information (34%).

Table 4.4: Total number of justifications (and percentages) for the Control, See and Not See conditions.

Number of responses (% ge)*	1 Multiple objects	2 Single object	3a Not enough info	3b Enough Info	4 Speaker's mental state	5 Listener's mental state	6 No answer/ Don't know	7 Hypoth-eticals	8 Other
Control Yes	0	96 (83)	1	7	5	1	5	1	0
No	0	2	9	0	0	1	2	2	2
See Yes	0	3	1	0	2	0	0	0	2
No	59 (47)	0	47 (37)	0	13	2	3	1	1
Not See Yes	1	11 (16)	1	0	52 (76)	2	1	0	0
No	31 (48)	1	22 (34)	0	4	5	0	1	1

* For the sake of brevity, only the modal percentages of high response categories are provided.

Given that there were very similar patterns of responding in the See and Not See conditions in terms of the types of justifications provided for negative judgments, it was necessary to see whether different participants use the same criterion for their judgments, or whether the same participants use different criteria depending on experimental condition. Of the 62 participants who answered negatively in both the See and Not See conditions, 33 (53%) gave the

same category of justification and 29 (47%) gave justifications that fell into two different categories. Justifications in categories 1 and 3 made up the vast majority of the latter type of responses.

4.1.4. Discussion

Overall, adults showed a strong tendency to say that the speaker had done a good job of saying exactly which item he/she wanted in the Control condition, and a bad job of saying exactly which item he/she wanted in the See condition. Therefore, they were willing to make a positive evaluation when no items were added to the array, and a negative evaluation when items were added to the array and the speaker knew this. On the majority of occasions, participants justified their positive evaluations in the Control condition by mentioning the array. That is, they said that the speaker had done a 'good job' 'because there was only one bike in the window'. In the See condition, justifications based on the relationship between the utterance and the array can fall into two categories; either category 1 (mentioning the array) or category 3 (mentioning the inadequacy of the utterance, presumably in relation to the array). If these two categories are considered jointly[2], then 84% of justifications in the See condition were focused on the

2: It was crucial to keep these categories separate in Chapter 3 since, in the presence of a thought bubble, the lack of information refers to an underspecification of the speaker's representation rather than an underspecification of the array. By contrast, given that thought bubbles were not included in the present experiments, the lack of information from the speaker refers to an underspecification of the array. This response indicates a recognition of the ambiguity of the array, as does the explicit mention of the array in category 1. Therefore, it was useful to combine the two categories in the present experiments as a measure of how often participants recognised that communication could be problematic due to the relationship between the utterance and the array.

utterance-array relationship rather than on the fact that the speaker knowingly provided an ambiguous utterance. In these two conditions, then, adults seemed to focus on the importance of the relationship between the utterance and the array in determining whether the speaker had done a 'good job' or not.

In the Not See condition, where it was possible to evaluate according to the speaker's internal representation, around half of the group were happy to do so. They were significantly more likely to judge positively in the Not See condition compared to the See condition. That is, rather than basing their judgments simply on the appearance of the array, they chose to use the speaker's belief state as the criterion for their evaluations and, therefore, judged that the speaker had done a 'good job'. This is supported by the types of justifications given for positive evaluations of the utterance in the Not See condition, where 76% made explicit reference to the fact that the speaker had done a 'good job' because he/she was not aware that the array had been altered. It appears that around 50% of the adult group gave a higher priority to the role of the speaker's internal representation, compared to the utterance-array link, in tasks where the particular criterion for evaluation is deliberately ambiguous and unspecified. Nevertheless, roughly half of the group continued with a pragmatic strategy in the Not See condition. That is, they relied on the utterance-array relationship to judge the adequacy of the utterance. Given that adults are supposed to possess a mature understanding of the mind, which includes the knowledge that utterances are representations of speaker's beliefs, it is surprising that such a large percentage of

the group did not utilise a mentalistic criterion when making an evaluation of an utterance.

A further aim of the pilot study was to obtain the consensus response concerning the listener's interpretation of the utterance. As expected, adults judged that the listener would pick the referent in the Control condition (because it was the only one there), but that the listener would not be able to tell which referent the speaker wanted in the See and Not See conditions. As predicted, then, there was no evidence that the adults' judgments of pragmatic adequacy (on behalf of the listener) were contaminated by their knowledge of the adequacy of the utterance-representation link. However, the percentages of 'can't tell' responses in the See and Not See conditions were lower than would be expected from this group (79% and 75% respectively). The Biscuit story seemed to be responsible for this effect due to the fact that quite a few respondents chose the referent in this story, compared to very few responses of this kind in the other two stories. The most likely explanation for this is that the initial biscuits were described as being the protagonist's – Adam's - favourite. Presumably, the listener (in this case Adam's Dad) would know that this would be Adam's preference and so would know which biscuit Adam wanted, irrespective of whether more biscuits had been added to the array. Therefore, this story was changed in the subsequent experiments to omit any reference to a stated preference.

Additional changes to the stories were also needed in order to match the details of each scenario more closely. In the pilot study, there were only two items

added to the initial referent in the Bike story, three items added in the Picture story and multiple items added in the Biscuit story. This may not have influenced adults' responses, but such discrepancies may affect children since the baseline for picking the referent is altered in each case. Moreover, it is possible that children could judge that because only a couple of items were added to the array, this means that the utterance could be a 'better job' than if many items were added. The possibility that children can be sensitive to the magnitude of discrepancy between utterances and states in the world has some empirical support. Mitchell, Munno and Russell (1991; also Mitchell & Russell, 1991) report that 8 to13-year-olds (but not 5-year-olds) were more likely to judge a message to be adequate if it was only slightly discrepant with the real world state of affairs. By contrast, messages that were grossly discrepant were judged to be inadequate for uniquely identifying a particular target. Consequently, the following experiments contain stories in which 3 items were added to the array in the See and Not See conditions, making a total of four items in all three types of scenario.

Experiment 4.2.

4.2.1. Introduction

The pattern of results from Experiment 4.1 seem quite clear with respect to the criteria on which adults base evaluations of utterances. It seems appropriate to suggest that if participants answer positively in the Control condition, it is because

they notice that the utterance is pragmatically adequate. Similarly, negative responses in the See condition suggest a pragmatic focus also since participants say 'bad job' because the utterance does not uniquely specify a referent. The same is true of negative responses in the Not See condition. By contrast, when participants judge positively in the Not See condition, presumably that is because they consider that the speaker's belief state – or representation of meaning – is relevant to their evaluation.

This pattern provides a useful basis on which to interpret the children's responses in the following experiment. The main aim of Experiment 4.2 was to obtain a pattern of responses based on the evaluation and interpretation questions only to investigate to what extent children differ from each other and from adults. Children were not asked to justify their responses because it was thought that asking for additional information using the following procedure might prove to be too demanding, and therefore stressful, especially for the youngest group of children. This concern was based on Bonitatibus and Flavell's (1985) finding that 6- and 7-year-olds provided uninformative explanations about ambiguous utterances.

Sixth-form (Year 12) students were also included in the following experiment as an adult baseline group against which the child data could be compared. The inclusion of this group was necessary on the grounds that the stimuli had changed between Experiment 4.1 and Experiment 4.2. It was, therefore, necessary to check whether the pattern of responding obtained in the pilot study could be replicated with a different group of adult participants.

4.2.2. Method

4.2.2.1. Participants. 96 Sixth-form students and 149 children were included. There were 38 male and 47 female Sixth-form students (11 students did not provide this information), aged between 16:4 and 18:4 (mean = 17:4). All attended a LEA funded Comprehensive School in Birmingham, West Midlands. There were 53 Year 1 children (25 males, 28 females; mean = 5:8; range 5:2 to 6:6), 50 Year 3 children (28 male, 22 female; mean = 7:7; range 7:2 to 8:2) and 46 Year 5 children (23 male, 23 female; mean = 9:7; range 9:2 to 10:2), all of whom attended a LEA funded Primary School in Sandwell, West Midlands.

4.2.2.2. Materials.

Children: All story scenarios were presented on A4 laminated card and illustrated using cartoon pictures. Each story was presented on a maximum of four separate cards. The story narratives were printed on the stimulus cards in 16pt Times Roman font. Responses were recorded on pre-prepared data sheets.

Sixth-form Students: Exactly the same stimulus cards were photocopied onto acetates to be shown to the students on an overhead projector. Test questions and stimulus arrays were printed on A4 sheets of paper.

4.2.2.3. Design. This was identical to Experiment 4.1.

4.2.2.4. Procedure.

4.2.2.4. (a) Children: The experimenter introduced herself and explained to the child participant that they would be shown some stories in which someone 'tells us something that they like, or something that they want.' Children were told that 'some of the people are quite good at saying what they want and some of the

people are not so good. I need you to help me figure out who is doing a good job and who is not doing such a good job, OK?' There were three different story scenarios: Bike, Picture and Coat. The Bike and Picture stories were identical to those included in the pilot study. The Coat story replaced the Biscuit story and is outlined below. Children were shown the first page of the story and the experimenter read the story aloud. For example, in the Coat story the first picture shows the protagonist hanging his coat above some bookshelves in the staff room at school. There are no other coats on any of the available pegs. The narrative for all experimental conditions began: 'At school, Mr. Riley walks into the staff room and hangs his coat above the bookshelves. So, there is only one coat hanging there.'* In the Control condition, the next picture children saw was of Mr. Riley standing behind his desk, talking to a boy in his classroom. The narrative continued: *'Later, Mr. Riley wants to show his class something outside and asks Nathan if he would fetch his coat from the staff room. Nathan asks how he will know which coat is Mr. Riley's. "It's the one above the bookshelves" says Mr. Riley.

Between the points marked * the narrative included additional information in the other two conditions. In the See condition, children were shown a picture of Mr. Riley sitting on a chair in the staff room, next to the bookshelves. There were now four different coats on the pegs (including Mr. Riley's). Children were told 'Whilst Mr. Riley drinks a cup of coffee, more teachers come in and hang their coats up. There are lots of coats above the bookshelves now.' In the Not See condition, the picture showed the same four coats on the pegs and an empty chair

beside the bookshelves. This time, children were told 'After Mr. Riley has left to go to his classroom, more teachers arrive and hang their coats up. There are lots of coats above the bookshelves now.' For all conditions, children were then shown the final page of the story, which was divided vertically into two sections. On one side, there was a large black question mark and on the other side there was a picture of the listener protagonist, Nathan, looking at the coats above the bookshelves. The side of the page on which the question mark and the picture appeared was systematically varied between story scenarios. Participants were asked two test questions:

(1) Evaluation: Did (speaker) do a good job of saying exactly which (item) he/she wants?

(2) Interpretation: Which (item) will (listener) think (speaker) wants – or won't he/she be able to tell?

The order of these questions was counterbalanced with rotation of experimental conditions. Before the Evaluation question, there was a linking sentence in the story, for example 'Nathan goes to fetch the coat...'. Children were seen individually in a quiet room. After the procedure they were thanked for their help and given a sticker for their participation.

4.2.2.4. (b) Sixth-form Students: Students were told that they were helping with developmental research, in order to establish an adult baseline response. They were told that there were no right or wrong answers and that they should give the answer that they thought best. Stories were presented on overhead acetates and read aloud by the experimenter. Students responded on pre-prepared

answer sheets containing the test questions and pictures of the final referent array for each story. The Sixth-form students were also asked to justify their responses to the Evaluation question, by answering the question 'Why/Why not?' The procedure was completed in small groups of 7-15 students. Discussion between participants was discouraged so that there was minimal opportunity for copying.

4.2.3. Results

Data from 16 Sixth-form students was disregarded due to an administrative error in the presentation of experimental conditions.

4.2.3.1. Evaluation question: Participants could either say yes or no to indicate whether the speaker had done a good job of saying exactly which item he/she wanted. Table 4.5 contains a summary of the total number of yes and no responses in the three experimental conditions for the Sixth-form students and for the child participants.

The children in Year 1 tended to answer positively on most trials, irrespective of experimental condition. Around 80% of the 5-year-olds in each condition judged that the speaker had done a good job of saying exactly which item they wanted. By contrast, the 7- and 9-year-olds seemed to discriminate between conditions, in the sense that the majority answered positively in the Control (74% and 63% respectively) and negatively in the See condition (62% and 78% respectively). Amongst the three groups of children, there was a tendency to judge negatively in the Not See condition with increasing age. Whilst only 21% of the 5-year-olds evaluated negatively in this condition, 76% of the 7-

year-olds and 91% of the 9-year-olds said that the speaker did not do a good job of saying exactly which item they wanted. In the Not See condition, the Sixth-form students answered similarly to the 9-year-olds with 87% evaluating negatively. This was also true for the See condition, in which 99% of the students said that the speaker did not do a good job. Responses in the Control condition were much more mixed in the Sixth-form group with 53% judging positively and the remainder (47%) judging negatively.

Table 4.5: Total number of yes and no responses (and percentages) to the Evaluation question for all groups of participants.

	Year 1 (mean age = 5:8)		Year 3 (mean age = 7:7)		Year 5 (mean age = 9:7)		Sixth-form (mean age =17:4)	
Good Job?	Yes (% of total)	No	Yes	No	Yes	No	Yes	No
Control	44 (83)	9 (17)	37 (74)	13 (26)	29 (63)	17 (37)	42 (53)	38 (47)
See	43 (81)	10 (19)	19 (38)	31 (62)	10 (22)	35 (78)	1 (1)	79 (99)
Not See	41 (79)	11 (21)	12 (24)	38 (76)	4 (9)	42 (91)	10 (13)	70 (87)

4.2.3.2. Comparisons between conditions. (see Table 4.6 for a summary of data).

4.2.3.2. (a) See versus Not See: Only five of the 5-6-year-old children gave differing judgments in the two conditions. This number was too small for statistical analysis. Similarly, only a few (n = 10) 7-8-year-olds changed their responses between these conditions and there was no significant difference between answering positively on one task and negatively on another. There were equal numbers of 9-10-year-olds (6 in each case) who gave differing evaluations in the two tasks. Unlike the three groups of children, the Sixth-form students did show a significant tendency to answer in a particular way when their judgments on the two tasks differed. They were significantly more likely to answer 'good job' in the Not See condition and 'bad job' in the See condition (n = 10) compared to the reverse (n = 0) [McNemar χ^2 (corrected, df = 1, n = 10) = 8.1, p<0.005].

4.2.3.2. (b) Control versus Not See: Very few children answered differently on the two conditions (n = 8) so numbers were too small for analysis. By contrast, the 7-8 year-olds showed a strong tendency to respond differently to the two conditions; they were significantly more likely to say that the speaker had done a 'bad job' in the Not See condition and a 'good job' in the Control (n = 26) compared to the reverse (n = 1) [McNemar χ^2 (corrected, df = 1, n = 27) = 21.3, p<0.005]. The remaining groups showed a significant tendency to answer in the same way [Year 5: McNemar χ^2 (corrected, df = 1, n = 25) = 19.4, p<0.005; Sixth-form: McNemar χ^2 (corrected, df = 1, n = 43) = 23.8, p<0.005].

Table 4.6: Summary of the number of responses to the Evaluation question between conditions for all four participant groups.

		Patterns of responses between conditions					
Year Group		*See versus Not See		Control versus Not See		Control versus See	
	Response to the Evaluation question	Good Job	Bad Job	Good Job	Bad Job	Good Job	Bad Job
Year 1	Good Job	40	3	38	5	39	5
	Bad Job	2	8	3	6	4	5
Year 3	Good Job	10	8	11	26	18	19
	Bad Job	2	30	1	12	1	12
Year 5	Good Job	1	6	4	24	6	22
	Bad Job	6	32	1	16	4	13
Sixth-form	Good Job	0	0	4	38	0	42
	Bad Job	10	70	5	33	1	37

* The first Condition mentioned corresponds to the vertical and the second Condition mentioned refers to the horizontal.

4.2.3.2. (c) Control versus See: Only 9 of the 5-6-year-olds provided different evaluations: five judged 'bad job' in the See condition and 'good job' in the Control and four responded in the reverse pattern. The 7-8 year-olds, by contrast, showed a strong tendency to judge that the speaker had done a 'bad job' in the See condition and a 'good job' in the Control ($n = 19$) compared to the reverse ($n = 1$) [McNemar χ^2 (corrected, $df = 1$, $n = 20$) = 14.4, $p<0.005$]. Again, the oldest two groups showed a similar pattern of responding to the Year 3

children; they were significantly more likely to judge 'bad job' in the See condition and 'good job' in the Control, compared to the reverse [Year 5: McNemar χ^2 (corrected, $\underline{df} = 1$, $\underline{n} = 26$) = 11.1, $\underline{p}<0.005$; Sixth-form: McNemar χ^2 (corrected, $\underline{df} = 1$, $\underline{n} = 43$) = 37.2, $\underline{p}<0.005$].

4.2.3.3. Justifications: Sixth-form Students. Justifications to the Evaluation question were coded in an identical way to the adult pilot data (see pp.125-126 for a summary of coding categories). There was an interrater agreement of 97%. Table 4.7 provides a summary of justifications for each experimental condition, according to whether participants answered yes or no to the Evaluation question.

Table 4.7: Total number of justifications (and percentages) for the Sixth-form group, according to Condition and response to the Evaluation question.

Number of responses (% ge)	1 Multiple objects	2 Single object	3 Not enough info	3a Enough Info	4 Speaker's mental state	5 Listener's mental state	6 No answer /Don't know	7 Hypoth -eticals	8 Other
Control *Yes*	0	34 (81)	0	6	0	0	2	0	0
No	1	2	14 (37)	1	0	3	0	14 (37)	3
See *Yes*	0	0	0	0	0	0	0	0	1
No	50 (63)	0	26 (33)	0	1	0	0	2	0
Not See *Yes*	0	1	1	7	0	0	1	0	0
No	40 (57)	0	24 (34)	0	4	1	0	1	0

Just over half of the sixth form students answered positively in the Control condition, and the majority of the justifications (81%) for the utterance being evaluated as a 'good job' were concerned with there being only one referent on display. Justifications for the speaker doing a 'bad job' in the Control condition were bi-modally distributed between saying that the speaker did not describe the referent in enough detail (37%) and invoking hypothetical scenarios (37% e.g. 'there might have been more put in the window'). In the See condition, the majority of students responded negatively and the modal justification made reference to there being lots of similar items in the array (63%). The second most popular response was to state that the speaker had not provided enough information about the referent (33%). A very similar pattern of responding was found for the Not See condition. The modal justification for negative evaluations was to make reference to multiple items in the array (57%), followed by the second most popular response of the speaker not describing the referent in enough detail (34%).

4.2.3.4. Interpretation question: Responses to this question were coded into three main categories, as before: (1) Referent: indicating that the listener would think the speaker wanted the referent; (2) Question mark: indicating that the listener would not be able to tell which item the speaker wanted and (3) Other: indicating one of the alternative items in the display. Table 4.8 contains a summary of responses for each condition, according to whether participants answered yes or no to the Evaluation question.

Table 4.8: Summary of responses to the Interpretation question for each condition according to story scenario and response to the Evaluation question.

	Year 1		*Year 3*		*Year 5*		*Year 12*	
Good Job?	Yes	No	Yes	No	Yes	No	Yes	No
Control								
Referent	41	8	34	9	28	13	42	30
Can't tell	3	1	4	2	1	4	0	8
Other	-	-	-	-	-	-	-	-
See								
Referent	27	8	8	9	4	5	1	2
Can't tell	3	1	7	16	3	28	0	75
Other	13	0	4	6	3	2	0	2
Not See								
Referent	17	5	7	8	0	8	0	3
Can't tell	4	2	3	23	3	31	7	64
Other	20	4	2	7	2	2	2	4

The vast majority of participants, across all four groups and irrespective of response on the Evaluation question, indicated correctly that the listener would pick the referent in the Control condition. In Year's 1,3,5 and 12 respectively 92%, 88%, 89% and 90% of responses were in the 'referent' category. Chi-squared analyses revealed that the participant groups did not differ significantly from each other in this condition.

By contrast, the correct response in the See and Not See conditions was to say that the listener would not be able to tell which item the speaker wanted. Children's use of the 'can't tell' category showed an increase with age. In the See condition, the respective percentages of children answering 'can't tell' in Years 1,3,5 and 12 were as follows: 7.7, 46.0, 68.8 and 93.7. The corresponding percentages for the Not See condition were: 11.5, 52.0, 73.9 and 88.7. The 7-8 and

9-10-year-olds did not differ significantly from each other in either condition, in terms of the number of referent versus nonreferent responses (see Table 4.9 for a summary of between-group referent versus nonreferent analyses for the See and Not See conditions).The 7-8-year-olds also did not differ from the 5-6-year-olds in this respect in the Not See condition, but they did differ significantly in See. It is noticeable that the majority (35 from 52) of the 5-6-year-olds and a large minority (17 from 50) of the 7-8-year-olds said that the listener would pick the referent in the See and Not See conditions. This is the item that the child participants knew to be the one that the speaker meant. By contrast, this type of response was very rare amongst the 9-10 and 17-18-year-olds (9 from 46 and 3 from 96 respectively).

Table 4.9: Chi-squared values* for between-group comparisons of referent versus nonreferent responses to the Interpretation question: See and Not See conditions.

			See (upper diagonal)		
	Year Group	1	3	5	12
	1	-	11.1	19.9	59.0
Not See (lower diagonal)	3	1.2 [c]	-	1.4 [c]	18.3
	5	7.3 [a]	2.0 [c]	-	6.9 [a]
	12	28.0	15.6	4.5 [b]	-

*All values significant at $p<0.005$ ($df = 1$, corrected, in all cases) except for:
[a] significant at $p<0.01$.
[b] significant at $p<0.05$.
[c] not significant.

Finally, the data were analysed to see whether there was a tendency for participants to judge that the listener would choose the referent more often in the Not See compared to the See condition. This pattern of responding could arise if participants were influenced by their knowledge that the speaker's utterance was an adequate characterisation of their (the speaker's) meaning, and, therefore, judge the utterance to be adequate from a pragmatic perspective also. The breakdown of within-group referent versus nonreferent responses is included in Table 4.10. There was a significant tendency for the 5-6-year-olds to give the reverse pattern to that predicted above. That is, they were significantly more likely to say that the listener would pick the referent in the See condition and a nonreferent in the Not See condition, compared to the reverse [McNemar χ^2 (corrected, $\underline{df} = 1$, $\underline{n} = 26$) = 6.5, $\underline{p}<0.025$]. For the remaining three groups, there were very few participants who changed their response between conditions and no significant differences between those who did judge differently.

Table 4.10: Summary of within-group comparisons of referent versus nonreferent responses to the Interpretation question for the See and Not See conditions.

	Year Group	Year 1 See		Year 3 See		Year 5 See		Sixth-form See	
		Ref	Nonref	Ref	Nonref	Ref	Nonref	Ref	Nonref
Not See	Referent	16	6	11	4	4	4	2	1
	Nonreferent	20	11	6	29	5	33	0	77

4.2.3.5. Order Effects: Data from the Sixth-form students were analysed separately, since the unequal numbers assigned to each rotation may have provided a misleading view of order effects in the three younger participant groups. Each condition was analysed separately, but data were combined across Years 1, 3 and 5. All order effects were calculated using chi-squared analysis. For the children in Years 1, 3 and 5, there were no significant effects associated with either rotation or question order. For the Sixth-form students, there were significant effects of rotation for all three experimental conditions [Control: χ^2 (df = 2, n = 80) = 11.40, p<0.005; See: χ^2 (df = 2, n = 80) = 18.47, p<0.005; Not See: χ^2 (df = 2, n = 80) = 18.47, p<0.005]. There was also a significant question order effect for the Year 12 group, but only in the Control condition [χ^2 (corrected, df = 1, n = 80) = 4.2, p<0.05]. As Table 4.11 testifies, there was a tendency for Sixth-formers to say 'bad job' more often when the Evaluation question appeared before the Interpretation question, compared to the reverse.

Table 4.11: Frequencies of yes and no responses to the Evaluation question for the Sixth-form group in the Control condition, according to order of question presentation.

	Question order	Interpretation/ Evaluation	Evaluation/ Interpretation	Total
Good Job?	Yes	25	17	42
	No	13	25	38
	Total	38	42	80

4.2.4. Discussion

There was little evidence amongst the children and the Sixth-form group that participants were including the speaker's mental representation, or belief state, in their utterance evaluations. Unlike the adults in Experiment 4.1, the Sixth-form students rarely judged that the speaker had done a 'good job' in the Not See condition. This was also the case for the 7-8 and 9-10-year-old children who answered 'no' on the majority of occasions in the Not See condition. Although the criteria for evaluations were not directly assessed in the children, the similarity of responding between Years 3 and 5 and the Sixth-form group is suggestive of similar strategy use. The justifications of the Sixth-formers in the Not See condition point overwhelmingly to the fact that they were focusing on the relationship between the utterance and the array rather than the utterance-internal representation link. Considering Category 1 (array based) and Category 3 (utterance based) responses together, 91% of the Sixth-form justifications explicitly mentioned the utterance-array relationship. Even though there was some evidence in the Sixth-form group that they were starting to adopt the more sophisticated criterion of judging according to the speaker's belief state (10 out of 70 said 'good job' in the Not See condition), they, nevertheless, failed to mention the speaker's belief state explicitly when asked to justify their responses. It seems unlikely, then, that the positive responses in this condition seen in the younger groups of children are a result of incorporating the speaker's internal representation into utterance evaluations. This is especially true of the 5-6-year-olds who showed little discrimination between conditions, tending to respond

positively irrespective of whether the utterance was ambiguous. This pattern is consistent with previous literature, which suggests that younger children have little understanding of the role of the actual words of the utterance in referential communication tasks and so tend to say that messages are adequate (Mitchell & Robinson, 1994; Flavell, Speer, Green & August, 1981; Robinson & Whittaker, 1987).

All groups, apart from the 5-6-year-olds, discriminated between the See and Control conditions, with significant numbers in each age group judging that the speaker had done a 'bad job' in the See condition and a 'good job' in the Control. Consistent with the adults in the pilot study, the Sixth-form students judged negatively in the See condition on the grounds that the array was ambiguous or that the utterance was inadequate (96% of justifications fell into these categories combined). Given that there seemed to be a developmental increase in the number of negative responses in the See condition, it is conceivable that children were also judging negatively because they had noticed the discrepancy between the utterance and the array. In addition, the students, like the adults, judged positively in the Control condition 'because there was only one item there' (81% of justifications fell in this category). Again, it seems unlikely that children were using wildly different strategies to the adults and Sixth-form students and were proficient at noticing the adequacy of the utterance because there were no other items to which the utterance could refer.

There was, however, a surprising tendency to evaluate the utterance negatively in the Control condition with increasing age. Moving chronologically

through the groups, from youngest to oldest, 17, 26, 37, and 47% of evaluations indicated that the speaker had not done a good job of saying exactly what he/she meant. This is especially surprising given the overwhelming preference amongst the adult group in Experiment 4.1 to say that the speaker had done a 'good job' in the Control condition. The finding that many participants judged negatively in the Control condition suggests that utterance evaluation does not simply take the form of mapping the utterance on to reality, or judging according to a perceived say-mean discrepancy. Such is argued to be the key process in referential communication, allowing children to distinguish between messages that are ambiguous and those that are not (Olson , 1970). Generally, participants in the present experiment seemed unwilling to evaluate the utterance based on the information in front of them. Even though the utterance, the array and the speaker's internal representation were in harmony, and the 'current' state of reality was clearly visible to participants at the time they made their evaluations, the speaker was still judged to have done a 'bad job'.

The present data seem to be consistent with the notion that children, even up to the age of 18 years, do not think a simple, adequate message is 'enough' in communicative terms. One possible reason why participants may not have sufficient confidence in their utterances is suggested by the types of justifications provided by the Sixth-form students for negative evaluations in the Control condition. Just over a third of justifications included the invocation of hypothetical scenarios that could render the utterance ambiguous. For example, students stated that there 'might be more bikes added to the window' or 'more

coats put on the pegs.' It is possible that, with increasing age, children become more familiar with instances in which communication failure has occurred. As a result, people may become more cautious about accepting utterances as accurate representations of reality and, therefore, engage in verbal redundancy or reasoning about hypothetical scenarios to limit their chances of being 'wrong'. After all, experience of being on the receiving end of inadequate utterances may teach us to be more careful, as speaker's, about how we characterise what we mean. There is some evidence of this in Lloyd (1991) and Hedelin and Hjelmquist (1998), both of whom found that speakers were significantly more likely to give adequate messages if they had previously played the role of a listener.

A problem with this interpretation is the finding that the adults in the pilot study rarely responded negatively in the Control condition. If there was a general increase with age in the tendency to treat utterances more cautiously, then this effect should be most pronounced in the oldest group of participants. However, it might be that the tendency to judge negatively in the Control condition in the Sixth-form group is a result of experimental artifact, rather than an extension of a developmental trend. The Sixth-form group may have responded in a way more consistent with the adults in the pilot study except that this was obscured by noise in the data (as a result of significant rotation effects). A further experiment is required to see whether a different group of adult participants would respond in a similar way. There was also a significant Question Order effect in the Control condition for the Sixth Form group. Students tended to say that the speaker had done a 'good job' significantly more often when they were asked to interpret the

utterance from the listener's perspective before they were asked to evaluate whether or not the speaker had done a 'good job'. This suggests that utterance evaluations were influenced by outcome, a finding normally only associated with children younger than about six years (e.g. Robinson & Whittaker, 1986). Generally, then, there is need for caution in making conclusions about the Sixth-form data and, hence, doubt about the extent of a trend to evaluate adequate utterances negatively.

The data from the Interpretation question seem much clearer. The Sixth-form students almost never said that the listener would be able to choose the item that the speaker wanted in the See and Not See conditions. Only 6 responses from a total of 160 (4%) indicated that the referent would be chosen, compared to around 25% of the adults in Experiment 4.1. In other words, the vast majority of the Sixth-form students indicated that the listener would not be able to tell which referent the speaker wanted in the conditions in which the utterance was ambiguous. Therefore, removing any mention of a particular preference for a certain item, seemed to increase the number of students interpreting utterances according to referential ambiguity, rather than according to contextual detail in the narratives.

Amongst the three groups of children there seemed to be a gradually developing tendency to say that the listener would not be able to tell which referent the speaker wanted in both the See and Not See conditions. This trend was accompanied by a corresponding decrease with age in the number of children saying that the listener would choose the referent. This seems to be similar to the

trend outlined in Braine and Rumain's (1983) chapter on logical reasoning (see also Robinson & Mitchell, 1990). The authors suggest that a 'can't tell' or 'undecidable' response is a "...cognitively complex response category" (p.274) which requires understanding the distinction between <I don't know whether p or not> and <One can't know whether p or not>. Braine and Rumain (1983) suggest that the logic of an undecidable response category is not available to children until around the age of six years, but that it seems to be more or less in place by about 9-10 years. Nevertheless, they suggest that there still remains a response bias against 'can't tell' responses which extends into adulthood, perhaps because "Subjects' reasoning strategies are probably geared toward making rather than shelving decisions, ...they prefer poorly based decisions to no decision" (p.276). To extrapolate to the present experiment, participants could prefer to pick one of the items in the array rather than answer 'can't tell' because they are still susceptible to an outcome bias. That is, it might be better to pick an item, even if it is not the referent, because it might serve as a suitable alternative (Mitchell & Russell, 1989).

Despite some children's reluctance to use the 'can't tell' response category, many children were, nevertheless aware that the referent would be unlikely to be chosen by the listener. Although most of the children in the youngest group said that the listener would choose the referent in the See and Not See conditions, this type of response was much less common amongst the older children. Therefore, even though some children were unwilling to say 'can't tell' they tended to point to an item other than the referent as the one the speaker

would choose. This suggests that there was at least some awareness that the utterance was problematical. Additionally, children did not seem to be influenced by their own knowledge of the adequacy between the speaker's representation and their utterance in the Not See condition. Such an influence could have resulted in children judging that the listener would be able to choose the referent more often in the Not See compared to the See condition. However, very few children gave differing responses on the two conditions and there were no significant tendencies amongst these children to respond in the way described above. Therefore, the children in the present experiment seemed to have quite a good grasp of the pragmatic value of the utterance, and seemed to respond only to the information contained in the message instead of being influenced by knowledge of the speaker's representation. Interestingly, though, the 5-6-year-olds demonstrated a significant preference for saying that the speaker would pick the referent in the See condition and a nonreferent in the Not See condition, compared to the reverse. This suggests that although there was some difficulty for the youngest children in knowing how an ambiguous utterance influences the listener's knowledge state (there was a strong preference to say 'referent' in the See condition), they, nevertheless, seemed to recognise that the Not See condition was different, or problematical, in some way.

Generally, the data from Experiments 4.1 and 4.2 suggest that children aged 5-10 years and adults aged 17 years and upwards tended not to use the speaker's internal representation as a criterion for utterance evaluation. However, it is not possible to tell whether children were answering in a similar way to adults

for the same reasons. It could be that children were responding according to the utterance-array or utterance-meaning links, or they might have interpreted the Good Job question in a different way. For example, they could have said that the speaker had done a 'bad job' because he/she was not polite or because they didn't pronounce their words properly. Children were not asked to justify their responses for fear that this would be too difficult for them, however many children provided spontaneous explanations in Experiment 4.2. It was, therefore, decided that the same procedure would be repeated in Experiment 4.3 with the modification that children would be specifically asked to provide justifications to the Good Job question.

Also, children were asked an additional Evaluation question in Experiment 4.3. Rather than asking the children to *interpret* the utterance from the listener's perspective, as in Experiment 4.2, they were asked to *evaluate* the utterance from the listener's perspective. It seems clear that children have relatively few difficulties determining whether the listener will be able to choose the correct referent when they are asked to consider this directly. However, it is not clear how children might respond if they are asked to think explicitly about the quality of the utterance from the listener's point of view. It might be that children are more likely to consider the role of the speaker's representation if they are forced to think about the communication on behalf of one of the protagonists rather than simply as an uninvolved 'onlooker'. Keysar (1994) argues that a more naturalistic way of eliciting utterance evaluations from children is for them to consider what might happen next. As he points out '...people do not normally judge perceived

intentions in an explicit manner... but they do habitually predict others' behavior' (Keysar, 1994; p.184).

In Experiment 4.3, children were asked to predict whether the listener protagonist would ask the speaker for more information about the referent or whether they would just go to the mentioned location. This question requires an implicit inference about the listener's perception of the adequacy of the communication. If the child judges that the listener would ask for more information from the speaker, then it can be assumed that the child thinks the utterance is inadequate in some way (Lloyd, 1992). By contrast, if the child judges that the listener would just go to the mentioned location, it can be inferred that the message is perceived to be adequate. Evidence of children taking into account the speaker's internal representation should take the form of saying that the speaker did a 'good job' in the Not See condition and saying that the listener would just go to the location of the item. However, if the Ask More question allows children to consider the role of the speaker's representation more easily than the Good Job question, there might be a tendency to say that the listener would just go to the mentioned location, yet judge that the speaker had done a 'bad job'.

Experiment 4.3

4.3.1. Method

4.3.1.1. Participants. A total of 86 children were included. There were 29 children from Year 1 (8 males, 21 females; mean age 6:0; range 5:7 to 6:6), 29

children from Year 3 (15 males, 14 females; mean age 7:11; range 7:4 to 8:5) and 28 children from Year 5 (14 males, 14 females; mean age 9:11; range 9:6 to 10:6). All children attended a Voluntary Aided Catholic School in the West Midlands.

4.3.1.2. Materials. All story scenarios were presented on A4 laminated card and illustrated using cartoon pictures. Each story was presented on a maximum of three separate cards. The story narratives were printed on the stimulus cards in 16pt Times Roman font. Responses were recorded on pre-prepared data sheets.

4.3.1.3. Design. Participants received a total of three story scenarios; one in each of the following experimental conditions: See, Not See and Control. The order of experimental conditions was rotated between participants and participants were randomly assigned to one of the possible rotations. Condition was mapped on to the type of story scenario, which was fixed across all participants. Participants were asked two main test questions and the order in which these appeared was counterbalanced across subjects and conditions.

4.3.1.4 Procedure. Children were introduced to the task in the same way as the previous study and told that there were no right or wrong answers. The three story scenarios were Bike, Picture and Coat. The Bike and Picture stories were identical to those in the previous study, but the Coat story contained some modifications. In pilot work with adults, it was noted that there could be differential responding in the Coat story because of a status imbalance between the adult teacher (speaker) and the child (listener). As a result, the story was

changed to include two child protagonists. The modified Coat story is outlined below.

Children were shown the first page of the story and the experimenter read the story aloud. For example, in the Coat story the first picture shows the protagonist hanging his coat above some bookshelves at a Youth Club. There are no other coats on any of the available pegs. The narrative for all experimental conditions began:

'Nathan arrives at the Youth Club and hangs his coat above the bookshelves. His is the only coat hanging there.' * In the Control condition, the next picture children saw was of Nathan standing outside, talking to his friend Sam. The narrative continued: *'Whilst Sam and Nathan are outside playing, Sam decides to fetch a ball for their game. Nathan asks her if she could get his coat because he is cold. Sam asks how she will know which coat is Nathan's. He says "It's the one above the bookshelves." '

Between the points marked * the narrative included additional information in the other two conditions. In the See condition, children were shown a picture of Nathan looking at the hooks above the bookshelves, whilst more coats were added to the array by another person. There were now four different coats on the pegs (including Nathan's). Children were told 'Whilst Nathan waits for his friend, the youth leader moves some coats from a chair and hangs them up next to Nathan's. So there are lots of coats above the bookshelves now.' In the Not See condition, the picture showed the same person putting the four coats on the pegs, but Nathan was not shown to be watching the transfer. This time, children were told 'After

Nathan has gone outside, the club leader moves some coats from a chair and hangs them next to Nathan's. So there are lots of coats above the bookshelves now.' The picture showing the initial array and the picture showing the final array were both visible when children were asked to evaluate the utterances. Children were asked the following two test questions in counterbalanced order:

(1) Ask More: Do you think Sam just goes to fetch the coat, OR does she ask Nathan any more about it? (the order of the action option of the listener was counterbalanced.) If the child indicated that the listener would ask for more information, they were asked to elaborate what kind of information would be asked for.

(2) Good Job: Did Nathan do a good job of saying exactly which coat was his? Why/why not?

Children were seen individually in a quiet room. After the procedure they were thanked for their help and given a sticker for their cooperation.

4.3.2. Results.

4.3.2.1. Evaluation questions. Children could either judge that the listener protagonist would ask for more information about the referent, or that they would just go to find the requested item. In addition, they could judge whether the speaker had done a good or a bad job of saying exactly which item they wanted. Table 4.12 contains a summary of children's answers to the two test questions.

On a group basis, the 5-6 year-olds tended to judge positively on the Good Job question irrespective of experimental condition. Only one child in the Control

and three in the See and Not See conditions said that the speaker did not do a good job of saying exactly which item he/she wanted. By contrast, the 9-10-year-olds judged in the opposite way to the youngest children, with the majority judging negatively on all three conditions. The 7-8-year-olds were more divided in their responses. Just under half of the group (48%) responded positively in the Control condition, and 34% judged positively in the See condition. Children were slightly less willing to judge positively in the Not See condition with 31% judging that the speaker had done a good job of saying exactly which item he/she wanted.

Table 4.12: Summary of the total numbers (and percentages) of children in each Year Group answering yes or no to the Good Job question and ask more or just go to the Ask More question.

Number (%ge)	Year 1 (5-6 years)		Year 3 (7-8 years)		Year 5 (9-10 years)	
Good Job?	Yes	No	Yes	No	Yes	No
Control	28 (97)	1 (3)	14 (48)	15 (52)	4 (14)	24 (86)
See	26 (90)	3 (10)	10 (34)	19 (66)	3 (11)	25 (89)
Not See	26 (90)	3 (10)	9 (31)	20 (69)	2 (7)	26 (93)
Ask More?	Ask more	Just go	Ask more	Just go	Ask more	Just go
Control	16 (55)	13 (45)	20 (69)	9 (31)	21 (75)	7 (25)
See	21 (72)	8 (28)	22 (76)	7 (24)	26 (93)	2 (7)
Not See	16 (55)	13 (45)	23 (79)	6 (21)	23 (82)	5 (18)

For the Ask More question, the majority of 7-8 and 9-10-year-olds responded similarly across all three conditions saying that the listener would ask for more information from the speaker, rather than just go straight to the mentioned location. This tendency was also found amongst the 5-6-year-olds in the See condition. The youngest group showed a less clear preference in the Control and Not See conditions with 16 children judging that the listener would ask for more information in both cases, and 13 children judging that the listener would just go straight to the mentioned location.

4.3.2.2. Comparisons between conditions (see Table 4.13 for a summary of responses).

4.3.2.2. (a) Good Job Question; See versus Not See: Children in each year group gave very similar patterns of responses to their counterparts in the previous study. For all three groups, there were no significant preferences for answering 'good job' in one condition and 'bad job' in another.

4.3.2.2. (b) Good Job Question; Control versus Not See: There were very few children who gave differing responses in Years 1 and 5 (4 and 2 respectively). Slightly more children gave differing responses in Year 3, but there was no significant preference to answer in a particular way. Nine 7-8-year-olds judged 'bad job' for Not See and 'good job' for Control and four gave the reverse pattern, but this difference was not significant.

4.3.2.2. (c) Good Job Question; Control versus See: Amongst the 7-8-year-olds, more children judged 'bad job' for See and 'good job' for Control (n = 6) compared to the opposite pattern (n = 2), but numbers were too small for

analysis. This was also true for the children in Years 1 and 5, where very few children gave differing responses to the two conditions (4 and 3 children respectively).

Generally, the majority of the youngest children judged positively across all conditions, whilst the majority of the oldest children judged negatively across all conditions. The 7-8-year-olds were more mixed in their responses. There was a majority preference to judge negatively for all conditions, but this tendency was not as marked as for the 9-10-year-olds.

Table 4.13: Summary of the number of responses to the Good Job question between conditions for Years 1, 3 and 5.

		Patterns of responses between conditions					
Year Group		*See versus Not See		Control versus Not See		Control versus See	
	Response to the Good Job question	Good Job	Bad Job	Good Job	Bad Job	Good Job	Bad Job
Year 1	Good Job	25	1	25	3	25	3
	Bad Job	1	2	1	0	1	0
Year 3	Good Job	6	4	5	9	8	6
	Bad Job	3	16	4	11	2	13
Year 5	Good Job	2	1	2	2	2	2
	Bad Job	0	25	0	24	1	23

* The first Condition mentioned corresponds to the vertical and the second Condition mentioned refers to the horizontal.

4.3.2.2. (d) Ask More Question; See versus Not See: (Responses are summarised in Table 4.14). For all three groups, very few children changed their responses between conditions and numbers were too small to compute statistical analyses.

4.3.2.2. (e) Ask More Question; Control versus Not See: Again, there were only a few children in each Year Group who gave differing responses in the two conditions, and numbers were too small for analysis.

Table 4.14: Summary of the number of responses to the Ask More question between conditions for Years 1, 3 and 5.

		Patterns of responses between conditions					
Year Group		*See versus Not See		Control versus Not See		Control versus See	
	Response to the Ask More question	Ask more	Just go	Ask more	Just go	Ask more	Just go
Year 1	Ask more	15	6	12	4	15	1
	Just go	1	7	4	9	6	7
Year 3	Ask more	18	4	17	3	17	3
	Just go	5	2	6	3	5	4
Year 5	Ask more	23	3	20	1	20	1
	Just go	1	1	4	3	6	1

* The first Condition mentioned corresponds to the vertical and the second Condition mentioned refers to the horizontal.

4.3.2.2. (f) Ask More Question; Control versus See: The same pattern was repeated here, with only very small numbers of children altering evaluations between conditions. Generally, then, like the pattern for the Good Job question, the two older groups of children showed a preference for saying that the listener would ask for more information from the speaker across all conditions. However, unlike the Good Job question, the 5-6-year-olds seemed more inclined to evaluate the Ask More question negatively, with the modal response indicating that the listener would ask for more information.

4.3.2.3. Comparing responses on Good Job and Ask More: Each evaluation question can be considered a separate measure of message adequacy (see Table 4.15 for a summary of frequency of responses). Firstly, McNemar χ^2 analyses examined whether participants tended to answer the two questions differently, within each condition.

For all three experimental conditions, the 5-6-year-olds judged that the speaker had done a good job of saying exactly which item he/she wanted, but, nevertheless, that the listener would ask for more information from the speaker (compared to saying 'bad job' and 'just go') [Control: McNemar χ^2 (corrected, $\underline{df}$ = 1, $\underline{n}$ = 17) = 11.5, $\underline{p}$<0.005; Not See: McNemar χ^2 (corrected, $\underline{df}$ = 1, $\underline{n}$ = 19) = 7.6, $\underline{p}$<0.01; See: McNemar χ^2 (corrected, $\underline{df}$ = 1, $\underline{n}$ = 20) = 14.4, $\underline{p}$<0.005]. By contrast, very few 7-8 or 9-10-year-olds gave inconsistent replies to the questions. Amongst the children giving different responses to the two questions, numbers were either too small for analysis or there were no significant preferences to respond in a particular pattern for any of the experimental conditions.

Secondly, general χ^2 analyses, or Fisher's Exact test in the case of small expected frequencies, examined whether there was a significant contingency between type of response, within each condition. Even though a majority of 7-8 and 9-10-year olds answered both evaluation questions negatively (saying 'bad job' and 'ask more'), there were no significant trends within the data to suggest that participants treated the two questions similarly.

Table 4.15: Summary of responses to the Ask More and Good Job questions for each condition for children in Years 1, 3 and 5.

		Condition					
Year Group		Control		See		Not See	
		Ask more	Just go	Ask more	Just go	Ask more	Just go
Year 1	Good Job	16	12	19	7	16	10
	Bad Job	0	1	2	1	0	3
Year 3	Good Job	9	5	8	2	6	3
	Bad Job	11	4	14	5	17	3
Year 5	Good Job	3	1	3	0	2	0
	Bad Job	18	6	23	2	22	4

4.3.2.4. Order Effects: Data for the Good Job and Ask More questions were analysed separately. Frequencies for answering either 'good job' or 'bad job', and 'ask more' or 'just go' were summed across year groups, for each experimental condition, and submitted to chi-squared analysis. There were no

significant order effects associated with question order, that is, whether the Ask More question appeared first or second. There were also no significant effects associated with rotation for the Good Job question. There was, however, a significant effect of rotation for the Ask More question in the Control condition [χ^2 (df = 2, n = 86) = 13.55, p<0.005]. Examining the data in Table 4.16, below, it seems that this effect arises because of a peculiarity in responding in Rotation 2. When the Control condition appeared last in the sequence of stories, children were much less likely to judge that the listener would ask the speaker for more information about the referent, compared to the other two rotations in which the Control condition appeared earlier in the sequence.

Table 4.16: Numbers of responses to the Ask More question, across year groups, for each rotation.

	Rotation 1 Control/See/Not See	Rotation 2 See/Not See/Control	Rotation 3 Not See/Control/See
Ask More	24	11	22
Just Go	6	17	6

4.3.2.5. Justifications - Good Job question. Justifications were coded according to the criteria specified in the previous experiment and summarised in Table 4.17. There was a 94% level of agreement between raters. The 5-6-year-olds provided very few justifications since they mostly evaluated positively and found it very difficult to explain why. Therefore, there are no justification data from the youngest group of children. Also, there were very few positive

evaluations amongst the two older groups of children and so only justifications for negative evaluations will be included.

Table 4.17: Types of justifications provided in Years 3 and 5 for answering the Good Job question negatively.

Number of responses (% ge)	1 Multiple objects	2 Single object	3 Not enough info	3a Enough Info	4 Speaker's mental state	5 Listener's mental state	6 No answer/ Don't know	7 Hypoth -eticals	8 Other
Control *Year 3*	0	2	6 (35)	0	0	0	0	8 (47)	1
Year 5	0	0	12 (50)	0	0	4	0	8 (33)	0
See *Year 3*	12 (67)	0	0	0	0	0	6 (33)	0	0
Year 5	13 (52)	0	8 (32)	0	0	1	0	3	0
Not See *Year 3*	8 (42)	0	10 (53)	0	0	0	0	0	1 (5)
Year 5	6 (25)	0	10 (42)	0	0	4	0	3	1

Both groups of children gave very similar justifications in the Control condition for why the speaker had not done a good job of saying exactly which item he/she wanted. The 7-8 and 9-10-year-olds stated that the speaker either did not describe the referent in enough detail or they invoked hypothetical changes that could occur to render the utterance ambiguous. By contrast, the modal response in the See condition, for both groups, was to mention the multiple similar objects in the array. In addition, a third of the 9-10-year-olds preferred to mention explicitly the quality of the utterance, saying that the speaker did not provide

enough information about the referent. A very similar pattern of explanations was found in the Not See condition, with responses being divided mostly between the ambiguous array and the lack of information in the utterance. Only the Year 5 children made any mention of mental states, and even then, this was confined to the listener's mental state (i.e. the utterance was no good because the listener would not have known which item the speaker was referring to).

4.3.2.6. Elaborations - Ask More question. Following Ironsmith & Whitehurst (1978), children's requests for more information were coded into either general or specific categories. Category 1 – General – included statements that would prove largely uninformative for the listener in terms of disambiguating information e.g. 'When did you want to buy it?'; 'How much is it'; 'Where in the school is the picture?' Category 2 – Specific – included statements that mentioned a particular attribute of the referent, or indicated that the appearance of the object needed to be described in more detail, in order to help the listener disambiguate the utterance e.g. 'What colour is it?'; 'What does it look like?'; 'Is there only one picture next to the clock?' There was also a third category which included the children who did not give an answer and also 8 responses that did not seem to fit in either of the first two categories e.g. 'she tells her Dad to ask her more questions'; 'more people might have hung their coats up'; 'because all are above the bookshelf.' A 10% sample of elaborations was coded (cf. Camaioni et al., 1998) by an independent rater. A 94% level of agreement was obtained between the experimenter and the second coder. The kinds of information that children judged would be asked for by the listener are summarised in Table 4.18.

Table 4.18: Children's suggestions of information asked for by the listener, according to Year Group and experimental condition.

	Number (%ge)	1: General	2: Specific	3: Other	Totals
Control	*Year 1*	4 (25)	9 (56)	3 (19)	16
	Year3	6 (30)	13 (65)	1 (5)	20
	Year5	7 (33)	12 (57)	2 (10)	21
See	*Year1*	8 (38)	12 (57)	1 (5)	21
	Year 3	4 (18)	16 (73)	2 (9)	22
	Year 5	4 (15)	19 (73)	3 (11)	26
Not see	*Year 1*	7 (44)	8 (50)	1 (6)	16
	Year 3	6 (26)	16 (70)	1 (4)	23
	Year 5	6 (26)	15 (65)	2 (9)	23

Children in all three year groups gave very similar responses in the Control condition. Between a quarter and a third of the children gave general statements and between a half and two-thirds provided more specific information about the referent. Given that the utterance was unambiguous in this condition, however, it was not so crucial for children to recognise the need to provide more specific information. In the See and Not See conditions, where the utterance was ambiguous, the 7-8 and 9-10-year-olds tended to give more specific information than the 5-6-year-olds. There was also a corresponding decrease in more general questions amongst the two older groups.

To examine whether the older children were significantly more likely to provide specific suggestions for more information compared to the youngest group, the General and Other categories were combined and compared against the Specific category. Separate χ^2 analyses were performed between pairs of year groups within each condition. Yates' Continuity Correction was applied in each case. There were no significant differences between any of the year groups in any of the conditions.

4.3.3. Discussion.

The results for the Good Job question replicate those in Experiment 4.2. In the See and Not See conditions, there was an overall preference to judge positively in the youngest group of children, moving through to mainly negative judgments in the older children. The inclusion of the justification question confirmed that the 7-8 and 9-10-year-old children were answering the Good Job question in the same way as the adults in Experiments 4.1 and 4.2. Children judged that the speaker had done a 'bad job' in the See and Not See conditions either because of the ambiguity of the array or because of the failure of the utterance to adequately specify an item in the array. Generally, there was very little evidence to suggest that children evaluated utterances on the basis of the speaker's representation in the Not See condition, since very few of the older children said that the speaker did a good job of saying exactly which item he/she wanted.

Also in line with the results from Experiment 4.2, was the tendency to judge the Control condition more negatively with increasing age. However, this tendency was much more pronounced in the present sample of children with 86% of the 9-10-year-olds answering in this way (compared to 37% of this age group in the previous experiment). The 7-8-year-olds were fairly equally divided between saying 'good job' or 'bad job', whereas the 5-6-year-olds again judged positively. It is impossible to tell whether the youngest group were saying 'good job' for the right reasons (i.e. because the utterance can only refer to one item, it must be adequate), since they had great difficulty in providing explanations, or whether it was due to their lack of understanding about the utterance being an independent object of scrutiny. The latter is suggested by the finding that children this age did not discriminate between conditions in their evaluations, either in response to the Good Job or the Ask More question.

The older children did not discriminate significantly between conditions either, but their justifications suggest that they were evaluating negatively for sensible reasons. In all three conditions, the majority of the 7-10-year-olds answered that the speaker had not done a good job of saying exactly which item he/she wanted. In the See and Not See conditions these children were sensitive to the ambiguity between the utterance and the array, and often stated this as the reason why the speaker had done a 'bad job.' The tendency to judge that the listener would ask for more information in the Control condition seemed to derive from an alternative basis, at least for some of the children. Like the children in Experiment 4.2, the 7-8 and 9-10-year-olds tended to invoke hypothetical changes

that could render the utterance ambiguous. Therefore, many children were not saying that the listener would ask for more information because they considered the utterance to be inadequate *at that time* but because, *in the future*, it might be. It is interesting that children were less likely to reason in this way if they saw the Control story last. This suggests that once children had been given the opportunity to contrast the Control condition against the See and Not See conditions, they were more likely to think the utterance was adequate. Overall, then, it seems that children were unwilling to treat informative messages as sufficient either because the utterance did not describe specific attributes of the referent, or because they reasoned about the possibility that the array could change.

It is possible that the older children did not discriminate significantly between conditions because they were not sensitive to the differences between them. That is, the results could indicate that the conditions did not successfully represent different kinds of scenario, at least as far as the children were concerned. However, as described in the previous paragraph, the older children's justifications for their responses suggest that they understood the ambiguity of the utterances presented in the See and Not See conditions, and often stated different reasons for evaluating negatively in the Control condition. This suggests that children were aware of the differences between the scenarios in terms of *actual* versus *possible* ambiguity engendered by the utterance.

There was also a failure to demonstrate that children were treating the two evaluation questions consistently, despite the fact that there was a majority preference amongst the 7-10-year-old children, in most conditions, to evaluate

negatively both times. This could indicate that the evaluation questions were not interpreted as being about message adequacy, resulting in children responding negatively for other reasons. However, again, children's justifications of their responses to the Good Job question suggest that they were attuned to message adequacy. Additionally, the type of information that children said the listener would ask for from the speaker supports the idea that children were sensitive to the reason why the utterance was problematical. Around half of the 5-6-year-olds and between two-thirds and three-quarters of the older children provided specific suggestions for the kind of information that would be asked for by the listener in the See and Not See conditions; allowing them to disambiguate the utterance. This tendency was also found in the Control condition, which could indicate that children were simply responding according to a general 'conversational' rule. That is, children said that the listener would ask for more information because they thought that was how listeners should normally behave.

However, children were aware that there was something specific about the referent that could be described in order to help the listener pick it out; they gave *sensible* and *relevant* suggestions about the kind of questions a listener might ask. If they were simply responding according to the idea that the listener should ask for more information, but not know why the information is required, their suggestions would tend to be superfluous, nonsensical or irrelevant. These latter types of responses were more common amongst the youngest group of children compared to the other two groups, although not significantly so. Nevertheless, it is still reassuring that at least half of the 5-6-year-olds who said that the listener

would ask for more information were able to give specific suggestions about additional information that was required, and they did not differ significantly from the older children in this respect. Therefore, the lack of significant results concerning consistency of interpretation are likely to be due to the very small numbers of children who associated 'good job' and 'just go' when evaluating utterances, rather than due to children providing random answers.

Even though it was not possible to demonstrate a significant effect amongst the 7-10-year-old children, the results generally suggest that children older than 6 years tended to treat the two evaluation questions similarly. That is, their explanations in the See and Not See conditions were often concerned with the ambiguity between the utterance and the array. Contrary to the original prediction, then, there was no evidence that older children found the Ask More question 'easier' in terms of judging according to the speaker's representation in the Not See condition; they very rarely indicated the speaker would just go to the mentioned location. Instead, they seemed to be focused on the pragmatic utility of the utterance, whether they had to judge from their own, or the listener's, perspective.

The final experiment of this chapter explored whether adults answered in a similar way to the children. Adult participants were presented with story scenarios identical to those presented to children in Experiment 4.3. The inclusion of this group was necessary given the equivocal results from the adult group in Experiment 4.2. In addition, it was necessary to determine how adults answered the Ask More question. Given the lack of inclusion of the speaker's representation

by the children in the previous experiment, it was necessary to see whether this was a response that would be favoured even by the adult group. If adults do incorporate the speaker's representation into their judgments in the Not See condition, this will be indicated by judgments of message adequacy i.e. that the speaker did a 'good job' and the listener will just go to the mentioned location.

Experiment 4.4.

4.4.1. Method.

4.4.1.1. Participants. 193 students were included with a mean age of 19:8 (range 18:0 to 33:11). There were 125 females and 51 males (17 respondents did not provide this information). All were undergraduates at the University of Birmingham studying either Psychology or History.

4.4.1.2. Materials. All story scenarios were presented on a single A4 page (one story per page). Story narratives were printed in 12pt font beneath black and white cartoon illustrations. Test questions were printed beneath the stories, with sufficient space for answers to be written.

4.4.1.3. Design and Procedure. The adults participated in a between-subjects design, rather than the within-subjects design presented to the children. This facilitated administration of the task. Consequently, participants received one story scenario only, in one of the following experimental conditions: Control, See or Not See. Each condition appeared in each of the three story scenarios, which were identical to those presented to the children (Bike, Picture and Coat). Stimulus sheets were randomly distributed to students in a lecture theatre. Pre-task

instructions to participants emphasised that there were no right or wrong answers. Students were also told that they should read through the story before answering the questions in the order in which they appeared on the sheet. Participants were presented with exactly the same pictures and narratives as those presented to the children, and were asked identical test questions, presented in counterbalanced order:

(1) Evaluation - Good Job: Did (speaker) do a good job of saying exactly which (item) he wanted/ was his? Why/why not?

(2) Evaluation - Ask More: Do you think (listener) just goes to fetch the (item), or does he/she ask (speaker) any more about it? (the order of the action option of the listener was counterbalanced).

4.4.2. Results.

4.4.2.1. Evaluation Questions. There were no significant differences in the way the Psychology and History undergraduates responded and so results are combined across the two groups. Each participant was coded according to whether they answered yes or no to the Good Job questions, and whether they judged that the listener would ask the speaker for more information about the referent or not. A summary of the total number of responses for each condition is included in Table 4.19. Preliminary analyses showed that there were no significant effects associated with question order.

Generally, there was a tendency amongst the adult participants to say that the speaker had done a 'bad job' in the Control and See conditions, and a 'good

job' in the Not See condition. Also, there was a trend across all three experimental conditions to say that the listener would ask for more information about the referent.

Table 4.19: Total number of responses (and percentages) to the Ask More and Good Job questions for each condition.

Question	Response	Number of responses (%)		
		Control	See	Not See
Good Job?	Yes	23 (34)	9 (14)	36 (65)
	No	45 (66)	57 (86)	19 (35)
Ask More?	Yes	56 (82)	40 (61)	38 (69)
	No	12 (18)	26 (39)	17 (31)

4.4.2.2. Between-group comparisons. For the following analyses, each question was considered separately.

4.4.2.2. (a) Good Job Question. There was a significant difference between the Control and See conditions in terms of the number of yes and no responses [χ^2 (corrected, $df = 1$, $n = 134$) = 6.4, $p<0.025$]. The majority response for both conditions was to say that the speaker had done a 'bad job', but this response was more common in the See condition. Also, participants were significantly more likely to say that the speaker had done a 'good job' in the Not See compared to the Control condition [χ^2 (corrected, $df = 1$, $n = 123$) = 8.4, $p<0.005$]. Finally, the number of yes and no responses also differed significantly

between the See and Not See conditions [χ^2 (corrected, $\underline{df} = 1$, $\underline{n} = 121$) = 32.3, $\underline{p}<0.005$]. Participants were much more likely to say that the speaker had done a 'good job' in Not See than in See.

4.4.2.2. (b) Ask More Question. There were no significant differences in responding between the Not See and Control conditions, or between the Not See and See conditions. However, the Control and See conditions did differ significantly in terms of the number of 'ask more' and 'just go' responses [χ^2 (corrected, $\underline{df} = 1$, $\underline{n} = 134$) = 6.7, $\underline{p}<0.01$]. Paradoxically, adults were more likely to judge that the listener would ask for more information in the Control condition, and more likely to say that the listener would 'just go' in the See condition.

4.4.2.3. Comparing responses on Good Job and Ask More: The following analyses looked at whether participants responded inconsistently on the two questions (e.g. saying 'good job' and 'ask more'), and, if so, whether there was a tendency for them to respond negatively on one particular question (data are summarised in Table 4.20).

Participants in the Control and Not See conditions were significantly more likely to say that the speaker had done a 'good job' and the listener would ask for more information, compared to the reverse [Control: McNemar χ^2 (corrected, $\underline{df} = 1$, $\underline{n} = 23$) = 4.3, $\underline{p}<0.05$; Not See: McNemar χ^2 (corrected, $\underline{df} = 1$, $\underline{n} = 31$) = 10.4, $\underline{p}<0.005$]. By contrast, the opposite pattern was found in the See condition, where there was a significant trend for participants to say that the speaker had done a 'bad job' and yet the listener would just go to the location of the referent, compared to the reverse [McNemar χ^2 (corrected, $\underline{df} = 1$, $\underline{n} = 23$) = 11.1, $\underline{p}<0.005$].

Table 4.20: Summary of responses to the Ask More and Good Job questions by undergraduate students for each condition.

	Condition					
	Control		See		Not See	
Ask More? / Good Job?	Ask more	Just go	Ask more	Just go	Ask more	Just go
Good Job	17	6	3	6	25	11
Bad Job	39	6	37	20	13	6

As in Experiment 4.3, general χ^2 analyses or Fisher's Exact test in the case of small expected frequencies, examined whether there was a significant contingency between type of response, within each condition. The majority response in the Control and See conditions was to say 'bad job' and 'ask more', but this preference was not significant in either condition. The majority response in the Not See condition was to say 'good job' and 'ask more', but again this trend did not yield a significant result.

4.4.2.4. Justifications - Good Job Question. Justifications were coded according to the criteria specified in the previous experiment and are summarised in Table 4.21 below. There was a 96% interrater reliability.

The majority of participants who indicated that the speaker had done a 'good job' in the Control condition, justified their response with respect to the array. That is, they mentioned that there was only one item on display (64%). The justifications for those answering negatively in the Control condition were mainly split between two categories: 42% said that the speaker had done a 'bad job'

because they had not provided enough information about the referent, and a further 44% mentioned hypothetical scenarios (e.g. the message was no good because there 'could be more bikes added to the window').

Table 4.21: Types of justifications provided by undergraduate students according to condition and response to the Good Job question.

Number of responses (% ge)	1 Multiple objects	2 Single object	3 Not enough info	3a Enough Info	4 Speaker's mental state	5 Listener's mental state	6 No answer/ Don't know	7 Hypoth -eticals	8 Other
Control *Yes*	0	14 (64)	0	4 (18.2)	1	0	0	1	2
No	1	0	19 (42)	0	0	2	0	20 (44)	3
See *Yes*	0	2	0	0	4	0	0	0	3
No	26 (46)	0	19 (33)	0	6	2	0	1	3
Not See *Yes*	0	9 (25)	0	1	24 (67)	0	0	0	2
No	4 (21)	0	6 (31)	0	4 (21)	1	0	1	3

In the See condition, the vast majority of participants gave negative evaluations and the modal justification was to mention the multiple similar items in the array (46%). A substantial proportion (33%) preferred to focus on the utterance, saying that the speaker did not provide enough information about the referent. There were also a few respondents who explicitly mentioned the speaker's mental state (10%). For example, they said that the speaker did not do a 'good job' because 'he knew there were other coats there.'

By contrast, most of the people evaluating positively in the Not See condition made reference to the speaker's mental state (67%). That is, they tended to say that the speaker had done a 'good job' because, for example, 'her description was sufficient according to her knowledge.' A further 25% of 'good job' justifications fell into the Single Object category. For example, statements such as 'there was only one when he walked past', and 'there was only one bike at the time he looked' were included in this category rather than in category 4 because the speaker's mental state was not mentioned explicitly. Justifications for negative evaluations in the Not See condition were more varied, with the majority split between three different categories: Multiple objects in the array (21%), Speaker's mental state (21%) and Speaker not providing enough information (31%).

4.4.3. Discussion.

In line with the adults in Experiment 4.1, there was a significant tendency to judge the utterance positively in the Not See condition compared to See. Therefore, even though the relationship between the utterance and the array was ambiguous in both cases, the adult participants often preferred to use the speaker's belief state in the Not See condition as the main criterion for evaluation, compared to the utterance-array relationship in the See condition. Again, this is supported by the types of justifications provided, since the majority of explanations for positive evaluations made explicit reference to the speaker's mental state. However, like the children in Experiment 4.2 and the children and the Sixth-form students in

Experiment 4.3, there was also a significant tendency for adult participants to judge that the speaker had done a 'bad job' in the Control condition (66%). Just under half of the justifications for these negative evaluations made reference to future hypothetical scenarios, whereas a similar percentage indicated that the speaker had not provided enough information. Therefore, like the older groups of children in Experiment 4.3 and the Sixth-form students in Experiment 4.2, there was a noticeable trend for participants to treat the sufficiently informative utterance as not 'good enough', and to respond according to an hypothetical state of affairs, not included in the story.

There was a strong tendency for participants to treat the informative utterance as inadequate when they responded to the Ask More question. Indeed, judgments that the listener would ask for more information were significantly more common in the Control condition than in the See condition. There was a preference for negative judgments in response to the Good Job question also, perhaps suggesting that the majority of adult participants treated the utterance similarly in the Control and See conditions. That is, the modal response in each condition was to say that the speaker had done a 'bad job' and the listener would ask for more information. However, this trend was not significant in either the Control or the See condition. Overall, treating the two evaluation questions similarly was no more common than treating the questions inconsistently. This should not be surprising, given the significant trends to treat the questions differently in all three conditions. Possible reasons for why the questions were treated differently in each condition are explored in section 4.3.

In the Not See condition, adults were significantly more likely to say that the speaker had done a 'good job', yet judged that the listener would ask for more information, compared to the opposite response of 'bad job' and 'just go'. The hypothesis concerning adults was that if they incorporated the speaker's internal representation into their 'good job' evaluations, then they would also tend to say that the listener would just go to the mentioned location, since both responses imply perceived message adequacy.

However, there were only 11 participants who indicated that the speaker had done a 'good job' and the listener would just go to the mentioned location. The general patterns from the See and Not See conditions suggest that adults treated these conditions similarly in response to the Ask More question, but very differently in response to the Good Job question. There was a strong preference for adults to judge that the listener would ask for more information in both conditions, but evaluate positively ('good job') in Not See and negatively ('bad job') in See. In other words, when making judgments from the listener's perspective (Ask More question), adults evaluated problematical utterances pragmatically. This pragmatic focus was also evident in judgments from their own perspective in the See condition (Good Job question), but not in the Not See condition where participants often made reference to the speaker's belief state, rather than focusing on the relationship between the utterance and the array.

4.3. General Discussion.

The main aim of the four experiments, was to explore the extent to which children, and adults, incorporate the speaker's internal representation in judgments of message adequacy when a criterion for evaluation is not specified by the test question. Participants were asked whether the speaker had done a 'good job of saying exactly which item he/she wanted' in a scenario in which the utterance was adequate as far as the speaker was concerned, but inadequate insofar as far as the referential array was concerned. Consequently, participants could choose to respond either according to the link between the utterance and the array, or according to the link between the utterance and the speaker's meaning. The adult participants responded quite differently to the children in one main respect. Generally, they were much more willing to evaluate according to the speaker's internal representation (say 'good job' in the Not See condition), when there was a choice of doing so, compared to the 7-8 and 9-10-year-old children, who almost never judged in this way. In other words, as far as the referential triangle is concerned, children seemed to focus on the relationship between the utterance and the array at the expense of the speaker's belief state.

By contrast, the adults often preferred to make evaluations based on the utterance-meaning relationship. In terms of the suggestion made in the Introduction concerning the possibility of a U-shaped development of responses to the Good Job question (cf. Mitchell & Robinson, 1994), it was not clear at what age children might be more inclined to judge according to the speaker's representation rather than the utterance-array relationship. It turns out that moving

beyond the literal, pragmatic, interpretation of a problematical utterance (judging 'bad job') to include the speaker's representation (judging 'good job') is a sophisticated response only available to adult participants.

Nevertheless, between 35% and 87% of the adult participants still did not utilise the speaker's internal representation as the central criterion for evaluations, under some circumstances. The literature summarised in the Introduction suggests that even four-year-olds can attend to the speaker's nonliteral meaning if they are asked to make an interpretation of the utterance (Mitchell et al., 1999; Robinson & Mitchell, 1992; 1994), or if they are directed towards a specific criterion when making an evaluation of the utterance (Mitchell & Robinson, 1994). However, the present results suggest that even adults may not spontaneously attend to the speaker's internal representation for the purpose of utterance evaluation when they are not directed towards a specific criterion either by the test question or contextual detail. Instead, they often seem to assign greater weight to the literal reading of the utterance (i.e. how the very words of the utterance relate to the array; Olson, 1970). Mitchell et al., (1999) suggest that '...children who are aged about 7 years and older would not necessarily be captured by their initial literal reading because they are effective in considering a range of possible options, and selecting from them the one that seems most appropriate after considering additional factors (e.g. the state of the speaker's belief)' (p.61). For many adults in the present experiments, then, the speaker's belief state was not considered to be sufficiently important to merit a criterion for evaluation when an alternative literal criterion could be utilised.

Gilbert (1991) suggests that we initially accept a literal representation as true as part of the comprehension process. Arguing from a Spinozan perspective, he claims that '...the acceptance of an idea is part of the automatic comprehension of that idea and...the rejection of an idea occurs subsequent to, and more effortfully than, its acceptance' (p.107). That is, propositions expressed (e.g. in sentences) are automatically accepted and can only be 'unaccepted' in the presence of very strong cues or contrary evidence. In the context of the present experiments, most participants accepted the literal representation of the utterance and failed to move beyond it to consider the speaker's representation. It seems that when strong cues concerning the appropriate criterion for evaluation are absent, an acceptance of the literal reading of the utterance may be a default setting that occurs as part of a general cognitive process as we come to believe propositions that we see or hear. Moving beyond the literal reading requires more effort, and further processing, that can only be applied once the literal has already been comprehended (see also Mitchell, et al., 1999).

Another main finding of interest was the tendency amongst the majority of groups to judge that the sufficiently informative utterance in the Control condition was inadequate. This was evidenced by the large percentage of older children and adults saying that the speaker had done a 'bad job' and judging that the listener would ask for more information from the speaker. In other words, many participants did not consider the utterance to be 'good enough' for effective communication and frequently required more information about the referent, even though this was redundant in terms of allowing a listener to select the correct item.

There is related evidence that children, and even adults, are not always satisfied by simple, contrastive messages and tend to prefer verbal redundancy in some circumstances. Whitehurst (1976) and Sonnenschein (1985) found that older children and adult speakers continued to include redundancy in their messages, because this was cognitively less demanding and often, more helpful, than excluding redundant information. The preference for redundancy, especially with large stimulus arrays, was also found for 8-year-olds and University students (Sonnenschein & Whitehurst, 1982) and 9-year-olds (Sonnenschein, 1982) when they were assigned to a listener's role. That is, older participants (but not those aged around 6 years) were more likely to find the correct target in a stimulus array when they heard a redundant message rather than a contrastive one (mentioning the minimum necessary to be informative). Perhaps more importantly, the 9-year-olds in Sonnenschein (1982) were significantly more likely to judge that a message did not describe any of the pictures in an array when a message was contrastive rather than redundant. Simply, the children judged the message to be inadequate even though it was sufficiently informative. Lloyd (1991; 1992), using a more naturalistic paradigm coupled with discourse analysis, found that children, and adults '...continued to negotiate meaning after the standard measures of completion had been reached' (1991; p.188). That is participants continued to engage in task-related conversation after an adequate utterance had been formulated and/or a referent selected. Lloyd suggests that this is because speakers may often lack confidence in their utterances, but this cannot usually be demonstrated using traditional referential communication paradigms.

At the very least this finding suggests that children are not simply judging according to a mapping strategy in the Control condition. Such a strategy involves comparing the utterance to the array to determine if an item is uniquely specified or not. If this were the case, the majority of evaluations in this condition should have been positive. This strategy seemed to predominate in the See and Not See conditions, especially amongst the 7-8 and 9-10-year-olds, and was supported by the tendency to focus on the utterance-array link when justifying responses. However, the justifications provided for negative evaluations in the Control condition strongly hint that rather than relying on the information provided in front of them, participants often appealed to plausible, imaginary scenarios that could result in the utterance being rendered ambiguous. To borrow Kahneman and Tversky's (1982) terminology, participants seemed to use a simulation heuristic, in which they incorporated 'downhill changes' to make the scenario more realistic. A downhill change is defined as '...one that removes a surprising or unexpected aspect of the story, or otherwise increases its internal coherence' (p.205). Participants may have considered that the likelihood of only one bike remaining in the window (for example) was low, and instead, introduced changes to the scenario that would result in a more probable end-state (that more bikes would be added to the window). These changes were completed for the participants in the Not See and See conditions, perhaps making it more likely that they would respond only to the information contained in the story.

However, there was still a significant minority of adult participants who said that the speaker had done a 'bad job' in the See condition, but also said that

the listener would just go to the mentioned location of the referent. Again, this is suggestive of participants including extra-contextual information in their reasoning. For example, the speaker in the See condition knowingly gives an ambiguous message, since they clearly witness the addition of more items to the array and yet still underspecify the referent. Participants could reason that the speaker is untrustworthy, and, presumably, the listener would also know this and so there would be little point in asking for more (equally uninformative) information. Sa and Stanovich (1999) and Nelson, Plesa and Henseler (1998) present direct evidence for the tendency of adult participants to 'complexify' theory of mind tasks by including other situational variables into story scenarios. In both studies, adults were asked to provide justifications for their responses to false belief tasks and many provided narratives that went beyond the given information. The finding that 7-8 and 9-10-year-olds also included hypothetical states in their justifications in the present experiments, suggests that this tendency also exists amongst younger participants. The main implication of this is that there needs to be some caution in assuming that children and adults are responding only on the basis of the logical or causal reasoning that is purposely manipulated by experimenters, and is argued to underlie the cognitive processing that is required by tasks like these.

Despite the significant minority of adults making inconsistent evaluations in the See condition, the majority of the adults and the 7-10-year-old children judged negatively in this condition, saying that the speaker had done a 'bad job' and the listener would ask for more information. This pattern of responding was

common amongst the older children in the Not See condition, indicating that there was very little incorporation of the speaker's internal representation in children's judgments.

This was not the case for the adult participants in Experiment 4.4 who showed a significant (and majority) preference to say that the speaker had done a 'good job' but that the listener would ask for more information from the speaker in the Not See condition. This suggests that whilst the speaker's representation was included in Good Job evaluations, the utterance was not deemed adequate from the listener's perspective. If the speaker was judged to have done a good job of saying exactly which item she wanted, then, assuming that the listener also treats the utterance as a product of the speaker's mental model of reality, the listener should consider the utterance to be sufficiently informative. Instead, participants tended to use their own knowledge of reality as the primary criterion for their judgments; behaving as if the ambiguity of reference was transparent to the listener. In other words, judgments were based on the referential success of the utterance, rather than on the specific belief framework within which the utterance was produced. This tendency is consistent with research which suggests there is a 'realism bias' which persists into adulthood in related theory of mind domains, for example, false beliefs (Mitchell, Robinson, Isaacs & Nye, 1996) and appearance-reality (Taylor & Mitchell, 1997). In both studies, adults were found to be significantly influenced by their own knowledge of reality, resulting in a difficulty making unbiased judgments either about what a story protagonist would believe, or about the shape of an object.

However, this interpretation remains highly speculative given the tendency for adult participants to judge that the listener would ask for more information in the Control condition also. The justifications for responses to the Good Job question suggest that evaluations in the Control condition might have a different basis: namely, that participants often tend to invoke hypothetical changes rather than consider the status of the array. Therefore, the reality bias could be evident in adults' evaluations of ambiguous utterances, but the Control condition does not provide a sufficiently unambiguous contrast to allow the bias to be seen in the Not See condition. This possibility could be checked in a follow-up study, in which the likelihood that participants could respond according to hypothetical states of affairs in the Control condition is limited. It might be that adults find it difficult to evaluate utterances with respect to the belief states of the speakers and listeners involved. Instead, participants may make judgments based on their own knowledge of the referential success of the utterance, rather than the specific belief framework within which the utterance is produced. This suggests that there could be a more widespread difficulty with knowing how to tailor communication to the informational needs of a listener.

4.4. Summary and Conclusion

In summary, children in their middle school years tend to be heavily focused on the relationship between the utterance and the array when evaluating problematical utterances. This was the case whether children were asked to make

an evaluation from their own, or a listener protagonist's perspective. This tendency was evident even in the Not See condition where it was equally appropriate to respond either according to the array, or to the speaker's internal representation. Even children aged 9-10 years did not consider the role of the speaker's belief in the formation of an utterance, when there was no criterion for evaluation specified by the test question.

By contrast, although around half of the adult participants did evaluate according to the speaker's representation when judging whether the speaker had done a 'good job' in the Not See condition, they tended to overapply a reality criterion when evaluating on behalf of the listener. In other words, even though many adult participants did consider the match between the speaker's belief and their utterance to be important, their knowledge about the discrepancy between utterance and reality tended to dominate their judgments. It seems that when thought content is not depicted directly, as in Chapter 3, children and adults tend to focus mostly on the utterance-array relationship rather than on the utterance-meaning relationship. Consequently, there might be a tendency to under-include thought content by virtue of the fact that it is usually intangible and unobservable. Reality might be particularly salient for children, drawing their attention away from thought content (Mitchell, 1996). Similarly, Russell, Mauthner, Sharpe and Tidswell (1991) suggest that "...for the young child physical knowledge is more salient than mental knowledge so that in circumstances where the two are in competition, the former wins out" (p.343). At least under the circumstances

presented in these experiments, it seems that this process might also apply to older children and adults.

CHAPTER 5

Does the realist bias in adult reasoning extend to evaluations of ambiguous utterances?

5.1. Introduction

In Experiment 4.4, a majority of adult participants evaluated a superficially ambiguous utterance positively, but, nevertheless, judged that the listener would ask the speaker for more information about the referent. It was suggested that this tendency may arise because the participants' own knowledge about the change in reality renders them unable to adopt the role of the listener objectively. That is, since the adults were quite happy to judge that the speaker had done a good job of saying exactly which item he/she wanted, they should also judge the message to be adequate from the perspective of the listener. However, they rarely judged in this way and, instead, seemed to be more influenced by the fact that the state of reality had altered – even though the speaker and listener protagonists did not know this.

The problem with this interpretation was that participants also judged that the listener would ask for more information from the speaker in the Control condition. If participants' own knowledge of the state of the array formed the main basis of evaluations, then they should have said that the listener would 'just

go' to the mentioned location since there is only one item on display. However, participants tended to 'complexify' the stories in Experiment 4.4 by introducing extra-contextual information and hypothetical states of affairs. Consequently, judgments that the listener would ask for more information may have been based on information not actually contained in the story.

In principle, then, the paradigm utilised in Chapter 4 allows scope for adult realism. It might be that adults are influenced by their own knowledge of reality in this context, but the Control condition did not provide a suitably unambiguous contrast to allow the bias to appear as a significant difference. The aim of the following two experiments was to investigate the possibility of a realism bias directly, with an improved Control condition. If adults show a significant tendency to be influenced by their own knowledge of reality when responding to utterances from other people, this could provide evidence that they experience difficulty treating utterances as the products of other minds.

5.2. Background literature

The tendency to base judgments on reality rather than belief originates in early childhood. Children aged 3-4 years demonstrate great difficulty understanding that a belief can be false, that is, that a belief can be discrepant in some way from the current state of reality (Wimmer & Perner, 1983). The interesting point is that this difficulty manifests itself in a very particular way. Children do not just make random errors or guesses when they get false belief tasks 'wrong'; instead, they make systematic realist errors (reporting the current

state of reality) when they should be reporting the content of the protagonist's false belief. For example, if Paul's chocolate has been moved from the drawer to the fridge in his absence, and the child participant is asked where Paul thinks the chocolate is, a typical 3-year-old will say that Paul thinks the chocolate is actually in the fridge. Of course, this is the location where the child knows the chocolate is currently residing. However, a typical 4-year-old will be able to answer according to Paul's belief and say that he will think the chocolate is in the drawer (where he left it). In short, the 3-year-old seems unable to separate their own knowledge of reality from someone else's.

This problem in separating belief from reality extends to judgments about the appearance of objects. For example, in the classic appearance-reality task (Gopnik & Astington, 1988), a sponge is painted to look like a rock and 3-4-year-olds are asked what the object looks like, what the object feels like and what they thought the object was when they first saw it. Normally developing 4-year-olds are able to make the distinction between what the object looks like and what they know about the object's true identity. By contrast, 3-year-olds will say that they thought the object was a sponge when they first saw it, even though upon first presentation they said it was a rock. Simply, they tend to confuse their own knowledge of reality (rock is really a sponge) with their prior belief about reality (the object is a rock).

Perner (1991) argues that the early difficulty with treating beliefs and reality as separable representations is a result of a deficit in conceptual understanding prior to the age of four years. Perner's argument is that the child

undergoes a radical conceptual shift around the time of her fourth birthday, furnishing her with the new concept that beliefs are actually only representations of reality and may, therefore, also misrepresent reality. This view entails the acceptance that concept acquisition is an all-or-none phenomenon. That is, the child will fail a false-belief task before acquiring the concept and pass it with ease once the concept has appeared.

However, work by Mitchell and colleagues has challenged the notion of a radical representational shift at age four years. In Mitchell, Robinson, Isaacs and Nye (1996), adult participants demonstrated a significant difficulty separating their own knowledge of reality from someone else's. The authors found that adults were significantly influenced by their own knowledge of whether an utterance was true or false when judging whether a listener protagonist would believe a message. Participants were presented with story scenarios in which they (the participants, but not the listener protagonist) were given privileged information about the truth or falsity of a speaker's utterance about the contents of a jug. Adults tended to judge that a message would be believed by a listener protagonist if they themselves knew that the message was true and disbelieved if they (the participant) knew the message to be false. The interpretation favoured by Mitchell et al. was that the adults tended to adopt a more sophisticated role-taking perspective and tried to interpret the message from the perspective of the listener. This led them to believe that it would have been impossible to reach any other conclusion than the one they knew to be true.

Additionally, Taylor & Mitchell (1997) presented adults with a sophisticated appearance-reality task. They found that participants who knew that an elliptical shape they were asked to view was really a circular disc, tended to significantly exaggerate the circularity of the disc when asked to reproduce the shape on a computer screen. This was contrasted with a group of participants who were not initially exposed to the actual shape of the disc. Their ability to match the viewed shape with the one on the computer screen was significantly more accurate than the informed group. These two studies together suggest that difficulties in separating belief from reality can continue into adulthood, albeit in a more subtle form.

There is also evidence more closely related to utterance evaluation which suggests that a similar process occurs when adults are asked to make judgments about perceived intentions of speakers when reading story vignettes. Keysar (1994) asked participants to take the perspective of a listener who heard a statement about the quality of a particular restaurant (for example) from a speaker. The statement could either be sarcastic or not sarcastic, but information relevant to a particular interpretation was only available to the participant. Results showed that when participants perceived the speaker to be making a sarcastic remark, they were significantly more likely to judge that a naïve listener would also perceive sarcasm. This was the case even though the utterance was presented in written form and so there were no paralinguistic cues present, which could have indicated one interpretation over another.

It is possible that a similar process could operate in the paradigm outlined in Experiment 4.4, although there was no direct evidence to suggest this was the case. In the previous experiment, adults were asked to adopt the perspective of a naïve listener, who heard an utterance that only the participant knew was ambiguous or unambiguous. Participants were also asked to provide an evaluation of the utterance, based on a question that did not specify from whose perspective the judgment needed to be made. Whilst many of the participants judged that the speaker had done a good job of saying exactly which item he/she wanted, they nevertheless judged that a listener would ask for more information about the referent from the speaker. It might be the case that adults try to adopt the role of the listener, but are unable to give an unbiased judgment because they (the adult participant) know that more items have been added to the display. Consequently, they realise that the listener will need to ask for more information if they are to successfully choose the correct referent from the display.

As noted in the Introductory paragraph to the present chapter (section 5.1), a limitation to this interpretation is that there was also a strong tendency for participants to judge that the listener would ask for more information in the Control condition in Experiment 4.4. This was the case even though participants could see that no more items had been added to the array. A possible clue to why participants responded in this way comes from the types of justifications they provided for their 'good job' evaluations. 44% of those who said that the speaker did a bad job in the Control condition, invoked hypothetical changes to the display. This was because the scenarios were such that it was entirely likely that a

change to the array could occur between the utterance being made and the listener seeing the array. In the following two experiments it was, therefore, important to try to establish a Control condition in which participants were less likely to think that changes could take place between the speaker seeing the display and telling the listener about it. Pilot work on earlier versions of the stories confirmed that participants were likely to create intermediate hypotheses about this interim period. Therefore, it was necessary to create scenarios in which the 'blanks' were filled in as much as possible. This could enable participants to focus on the information presented to them, rather than invoking hypothetical changes. In an attempt to make the Control condition as unambiguous as possible, two main features of the story were emphasised: (a) the speaker's belief that there was only one jumper and (b) the fact that no more jumpers were added to the display.

The aim of the following two experiments, then, was to directly investigate the potential phenomenon of a realist bias in utterance evaluation in the adult population. Following the methodology of Chapter 4, adult participants were presented with story scenarios in which they (but not a speaker or listener protagonist) know that something has happened to render the utterance ambiguous. This was contrasted with a scenario in which the utterance remains unambiguous. For example, if Rachel sees a jumper in the shop window (it is the only one there) and goes home to tell her friend about it, her utterance will be unambiguous if she states "there's a really nice jumper on display in the window in Stevenson's clothes shop. You'll spot it easily because it is the only one there." However, if during Rachel's walk home the shopkeeper added more jumpers to

the display (without Rachel's knowledge), then her utterance is rendered ambiguous by the change in reality. The important point is that in both conditions the belief states of the speaker and listener are identical. The only difference between the two groups is that in the second scenario the participants are privy to information that rendered Rachel's utterance objectively ambiguous; that is, it was no longer adequately specifying the intended referent. If adults are uninfluenced by their knowledge of the change in reality, they should judge that a listener would ask for more information no more often in the Ambiguous (Not See) than in the Unambiguous (formerly 'Control') condition. However, if adults find it difficult to divorce their own knowledge of reality from what an uninformed listener may know about it, they will be more likely to judge that the listener would ask for more information in the Ambiguous compared to the Unambiguous condition.

Experiment 5.1.

5.1.1. Method

5.1.1.1. Participants. 57 mature students were included, aged between 19:10 and 61:3 with a mean age of 38:6. There were 49 females and 8 males. All students attended a LEA funded College of Further Education in Birmingham, West Midlands.

5.1.1.2. Materials. All story scenarios were typed and printed in 12 pt font on an A4 sheet of paper. There were no cartoon illustrations. Two different stories appeared on each sheet of paper.

5.1.1.3. Design and Procedure. Participants were randomly allocated to either the Unambiguous or Ambiguous condition. A between-subjects design was chosen for ease of administration and, importantly, Tversky and Kahneman (1983) note that: " Judgment problems in real life do not usually present themselves in the format of a within-subjects design or a direct test of the laws of probability. Consequently, subjects' performance in a between-subjects test may offer a more realistic view of everyday reasoning" (p.311).

Each participant was opportunistically sampled from the college cafeteria and asked to take part in a study looking at adult reasoning. They were told to read the stories on the sheet provided and to answer the questions that followed. They were also encouraged not to discuss their answers with their neighbour. Each participant saw two scenarios, a Jumper story and a Bar story. Type of condition (Ambiguous or Unambiguous) was mapped on to story scenario, and the order in which Jumper and Bar appeared was counterbalanced between subjects. For example, in the Unambiguous condition, participants were presented with the following story:

"Julie went to the night-club bar to get a round of drinks. There was a really good-looking guy behind the bar. He was the only guy there. All of the other bartenders were women. Julie couldn't carry all of the drinks herself so she asked the barman

if he could look after them. Then Julie walked back to her friends.* No other guys were working behind the bar that night. Claire offered to fetch the remaining drinks from the bar and wanted to know who she should ask. Julie said **"There is a really cute guy serving the drinks. You'll spot him easily because he is the only guy there."** Claire was glad to discover that the drinks were waiting for her at the bar. She was eager to get the drinks as soon as possible so that she could see the guy Julie was talking about."

The Ambiguous version was identical except for the additional information at * that more male bar staff had arrived for work and so there were a few guys serving behind the bar now. Participants were then presented with the Evaluation question, which required a categorical response:

What do you think happened next? (circle either A or B).

(A) Do you think Claire asked Julie any more about the appearance of the guy behind the bar?

OR

(B) Do you think Claire went straight away to fetch the drinks?

The order in which the action options appeared was the same for each participant, but was systematically varied between participants. Participants were debriefed upon completion of the task and thanked for their help.

5.1.2. Results

5.1.2.1. Responses to the Ask More question. Participants could either indicate that the listener would ask the speaker for more information about the

referent, or they could say that the listener would just go straight to the mentioned location. Students were given a score of 1 each time they judged that the listener would ask for more information, resulting in a possible score from 0-2. A breakdown of the total number of responses in each of these categories is included in Table 5.1.

Compared to the adults in Experiment 4.4, far fewer participants indicated that the listener would ask for more information from the speaker in the Unambiguous condition. The two experiments cannot be compared directly, but from the total number of Unambiguous stories presented to participants in the present experiment (n = 58), the listener was judged to ask for more information on only 16 occasions (28%). The equivalent percentage of participants who indicated that the listener would ask for more information in the Control condition in Experiment 4.4 was 82%. The aim of the present experiment was to create a condition in which the default response was not 'ask more.' It seems that by highlighting the belief of the speaker and by emphasising the state of the array (e.g. there being only one jumper on display), participants were less likely to include extra-contextual information and more likely to consider the utterance to be adequate.

Table 5.1: Total number of participants judging that the listener would ask for more information on neither, one or both of the stories.

Ask for more information on...	Neither story	One story	Both stories
Unambiguous	15	12	2
Ambiguous	7	15	6

The scores in Table 5.1 were submitted to a 2(Version: Ambiguous or Unambiguous) x 2(Story Order: Barman first or Jumper first) x 2(Question Option Order: 'ask more' first or 'just go' first) ANOVA. All three factors were between-subject classifications.

There was a significant main effect for Version [$F(1,49) = 7.88$, $p<0.01$]. Mean scores (see Table 5.2) confirm that this was in the predicted direction with adults significantly more likely to say that the listener would ask for more information when the utterance was Ambiguous compared to when it was Unambiguous. There was also a significant main effect for Story Order [$F(1,49) = 5.84$, $p<0.05$]. Participants were significantly more likely to say that the listener would ask for more information when the Bar story preceded the Jumper story than when the Jumper story appeared first (see Table 5.2 for a summary).

Table 5.2: Mean 'ask more' scores, from a total of 2, (and standard deviations) for Version and Story Order

	Version		Story Order	
	Unambiguous	Ambiguous	Jumper/Bar	Bar/Jumper
Mean scores	0.55 (0.63)	0.96 (0.69)	0.57 (0.63)	0.93 (0.70)

Breaking the data down further (see Table 5.3), this main effect can be attributed to the tendency for participants to judge that the listener would ask for more information specifically in the Jumper story compared to much less frequent

judgments of 'ask more' in the Barman story. There was no effect of Question Option Order and no significant interactions between any of the independent variables.

Table 5.3: Total number of responses to the Ask More question according to Story Order.

Story Type	Response to Ask more question	Story Order	
		Jumper/Bar	Bar/Jumper
Jumper	Just Go	17	11
	Ask more	11	18
Bar	Just Go	23	20
	Ask more	5	9

5.1.3. Discussion

The aim of this experiment was to see whether adult participants were able to make unbiased judgments about the adequacy of an utterance, when they had to adopt the perspective of an uninformed listener. Only the adult participants, but not the speaker or listener protagonists, knew that a change in the array had rendered an utterance ambiguous. If adults are 'uncontaminated' by their knowledge of the addition of extra items, they should treat this situation identically to one in which the utterance remains unambiguous. However, the participants in the present study did not respond in this way (at least in the Jumper story). Instead, they were influenced by their privileged knowledge of the updated state of reality and tended to judge that the listener would ask for more information from the speaker significantly more often in the Ambiguous condition

compared to the Unambiguous condition. Given that there was no reason to expect that the speaker was deliberately trying to mislead in the Ambiguous condition, the utterance should have been treated identically to the one in the Unambiguous condition. However, by treating the objectively Ambiguous utterance as inadequate, adults tended to use reality as the primary criterion for their judgments; behaving as if the ambiguity of reference was transparent to the listener.

It could be argued that the observed order effect accounts for this finding. However, Order did not significantly interact with Version. This suggests that whilst experimental design may lead to noise in the data, this does not threaten the theoretical interpretation due to this noise being equally distributed across versions because of counterbalancing and random assignment of participants.

According to Olson (1970) the problem of referential ambiguity can be solved by comparing the words of the utterance to the array, and it is only when children understand the role of literal meaning in referential communication that they begin to judge messages to be inadequate. It seems that adults are very proficient at this comparison process also. That is, the participants knew that there was an element of referential ambiguity and so they tended to respond as if they were mapping the words of the utterance on to the referent display. This is exactly the correct procedure for determining whether an utterance is adequate or not, in most tests of referential ambiguity. However, in the present task, it is no longer sufficient simply to compare the words of the utterance with the referent display in order to decide how to evaluate the utterance. Instead, participants have to

make a judgment from the perspective of the listener who, in the context of the story, does not have access to the same information as the adult participant. Therefore, it is necessary to consider how the listener would behave based on the information available to him/her *in the utterance only*. The mapping of the utterance on to reality becomes a redundant process, because the information concerning the referent display is not available to the listener.

The finding that adults often preferred to base their judgments on a comparison between the utterance and the array, rather than on the belief state of the speaker, is surprising given that even children aged below 6 years seemingly have little difficulty utilising the speaker's belief when asked to make an interpretation or an evaluation of an utterance (Mitchell, Robinson & Thompson, 1999; Robinson & Mitchell, 1992; Mitchell & Robinson, 1994). Also, in the previous experiment, adults often incorporated the speaker's belief in their evaluations of whether the speaker had done a 'good job', suggesting that they are able to recognise the importance of the speaker's belief under certain circumstances. Perhaps the difference in findings arises from the fact that participants were required to 'step into the shoes' of the listener protagonist in order to judge the adequacy of the message in the present procedure, whilst in the procedures mentioned above participants could take a more objective perspective. As noted in the Introduction to this chapter (p.187), it could be that by being explicitly asked to assume the perspective of the listener, participants engage in advanced role-taking which makes it hard for them to consider alternative outcomes to the one they know to be true (Mitchell et al., 1996).

A reductive explanation for the present results might argue that a similar pattern of data could be obtained if participants were simply trying to demonstrate that they had comprehended the alteration in the state of reality. That is, they might just say that the listener would ask for more information because they wanted to show that they had noted more jumpers had been added to the window. One way to solve this problem would be to add more items to the display in the Unambiguous version, such that the addition of the items does not influence the (un)ambiguity of reference. Therefore, if participants still demonstrate a significant tendency to say that the listener would ask for more information in the Ambiguous condition, this cannot simply be because they are trying to demonstrate that they have acknowledged the alteration in reality. If this were the case, participants would be expected to perform similarly on the Ambiguous and Unambiguous versions.

An additional concern with the present procedure is that the change in reality that occurs in both stories might unwittingly provide information about the predictability of the event. Participants could reason that if the change in the array happens *this* time, it is probably a fairly common occurrence. Presumably if the listener knows this is common also, they would be more likely to ask for more information from the speaker. This is reminiscent of the types of cognitive heuristics that adults commonly use in reasoning tasks when they are uncertain about their response, or important information is unavailable (e.g. Kahneman & Tversky, 1973; Taylor, 1982). For example, Tversky & Kahneman (1982) suggest that people use an 'availability heuristic' to reduce the uncertainty about making a

judgment. That is, they tend to judge that an event is highly probable if examples of similar events come quickly and easily to mind. This heuristic is employed even if the actual occurrence of the event is relatively rare according to veridical or objective criteria (Tversky & Kahneman, 1983; Kahneman & Tversky, 1973). This heuristic (and the simulation heuristic; Kahneman & Tversky, 1982) produces characteristic biases in favour of events which are thought to be plausible and predictable, with a corresponding tendency to underestimate the likelihood of events that are produced in a multitude of unlikely ways. As a result, the participants in the present study could reason that more staff turning up at the bar, or more jumpers being added to the shop window are events that are highly plausible and, therefore, predictable. Presumably, the listener would also know this and would need to ask for clarifying information from the speaker in order to anticipate such a likely occurrence.

To control for this problem, predictable and unpredictable versions of the same scenario would need to be produced. In addition, participants could be asked to evaluate the perceived probability of the change in reality actually occurring. If participants still show a tendency to judge that more information would be asked for in an ambiguous, yet unpredictable, condition, this would provide much stronger evidence that adults are specifically contaminated by their own knowledge of reality when making judgments of this kind. Moreover, it would also suggest that 'realism' is a process that works in addition to the biases that Kahneman & Tversky talk about in their work. That is, if participants were to continue to ask for more information in an unpredictable, ambiguous scenario, it

would suggest that their knowledge of reality tends to override their tendency to underestimate the likelihood that unpredictable events will occur.

Experiment 5.2.

5.2.1. Method

5.2.1.1. Participants. 95 undergraduate students were included, aged 18:6 to 41:6 with a mean age of 21:9. There were 78 females and 17 males. All were studying psychology at the University of Nottingham.

5.2.1.2. Materials. Story scenarios and related questions were printed in 12pt on A4 paper. All the information required by each participant was included on one sheet only.

5.2.1.3. Design and Procedure. A completely between-subjects design was employed, with participants randomly assigned to one of the four experimental conditions. All four conditions were based on the same story scenario, which involved Mark (the speaker) and Jim (the listener) talking about books in the library.

In the Unambiguous version, books are added to a different desk whilst Mark is absent. Hence the addition of books to the desk does not alter the informative value of the utterance. In the Ambiguous version books are added to Mark's desk, without his knowledge. Therefore, his utterance is rendered ambiguous by the change in reality. The Predictable change version described how books were placed on the desk as Sue left the library. The desk was described

as being by the exit, with the implication that this was a frequent occurrence because people often put their books in a convenient location as they leave the building. The Unpredictable change version described how Sue was struggling past the desk with a heavy box and books were accidentally spilled onto the desk. Participants are told that the books are being moved to a new room in the library with the implication that this is a relatively infrequent occurrence. To illustrate, the basic story - Predictable/Unambiguous - is outlined below in addition to the specific variants for each condition:

"Mark is reading a book at desk 31 in the busy college library. It is the only book on his desk. He decides he needs some more information and goes to the bookshelves on the next floor. *Whilst Mark is walking to the shelves, Sue is on her way out of the library. She has finished reading her books and puts them down on desk 32, the desk next to Mark's.* At the bookshelves, he sees his friend Jim, who is reading about a similar topic. Jim asks Mark if there is a book he can recommend. Mark says "Yes. I have a great one downstairs on my desk. You'll spot it easily because it's the only one there." Mark tells Jim he is sitting at desk 31, near the exit."

The remaining three versions were identical except for the change in information included between the points marked *:

Unpredictable/Unambiguous: Whilst Mark is walking to the shelves, a Librarian struggles past desk 32 with a heavy over-filled box. She is transferring some books to a new room in the library. Some of the books fall onto desk 32, the desk next to Mark's.

Predictable/Ambiguous: Whilst Mark is walking to the shelves, Sue is on her way out of the library. She has finished reading her books and puts them down on desk 31, next to the book that Mark was reading.

Unpredictable/Ambiguous: Whilst Mark is walking to the shelves, Sue the Librarian struggles past desk 31 with a heavy over-filled box. She is transferring some books to a new room in the library. Some of the books fall onto Mark's desk, next to the book that Mark was reading.

The response section was then divided into three parts, which was presented in fixed order for all participants:

Participants were asked to indirectly evaluate the utterance. They were asked:

(1) What do you think happened next? (circle either A or B)

(A) Do you think Jim went straight away to fetch the book from the desk?

OR

(B) Do you think Jim asked Mark for more information about the book? (the order of these options was counterbalanced between subjects).

Next, they were asked to evaluate the percentage likelihood of the change in reality occurring:

(2) How often do extra books end up on people's desks like this?

[If 0% represents Very Rarely, 50% represents Sometimes and 100% represents Very Frequently, give a percentage value anywhere between 0-100 that estimates the frequency with which you would expect this to happen.]

Finally, participants were asked to categorise their percentage response according to a number of pre-specified frequencies. This was to control for the possibility

that participants may misunderstand what is required from them in Question 2 resulting in widely varying data. They were told to:

(3) Indicate on the list below the description that most closely represents the percentage value that you have stated above.

Every day
Every 2-3 days
Once a week
Once a fortnight
Once a month
Less than once a month
Less than once every six months
Less than once a year
Almost never

Participants completed the task, without any opportunity for collaboration, in a lecture theatre. They were told that there were no right or wrong answers and that they should complete the questions in the order they were presented on the sheet. After completing the task, participants were debriefed and thanked for their cooperation.

5.2.2. Results

5.2.2.1. Responses to the Ask More question. Table 5.4 contains a summary of the number of participants in each condition who judged that the listener would ask for more information from the speaker. Overall, there were very few participants from the total group who answered in this way (29%). There was no tendency to judge that the listener would ask for more information from the speaker when the utterance was Ambiguous. Indeed, more people judged that the listener would ask for more information when the utterance was

Unambiguous. This is in the opposite direction to the experimental hypothesis and so no analyses were carried out on this data. In addition, there was almost no difference in the numbers of participants judging 'ask more' between the Predictable and Unpredictable conditions. Combining across the Unambiguous and Ambiguous conditions, 15 people said 'ask more' in the Predictable version and 13 people gave this response in the Unpredictable condition.

Table 5.4: Total number of responses to the Ask More question according to predictability of change and ambiguity of utterance.

Story Version	Predictable		Unpredictable	
	Ask More	Just Go	Ask More	Just Go
Ambiguous	4	18	4	20
Unambiguous	11	14	9	15

5.2.2.2. Percentage estimations of probability of change. The participants estimated the probability of the change in reality, described in the story, actually occurring as a percentage. The mean percentages for each condition are summarised in Table 5.5. Overall, there was very little difference between judged frequencies according to whether the utterance was Ambiguous or Unambiguous. However, there was a marked difference in estimated percentages between the Predictable and Unpredictable versions of the story. Participants judged that the change described in the story would occur less frequently in the Unpredictable version compared to the Predictable version. This difference was

confirmed in a 2(Ambiguity of Utterance) x 2(Predictability of Change) x 2(Order of alternatives in Question 1) ANOVA. All three factors were between-subject variables. The only significant result was a main effect for Predictability [$F(1,87) = 18.73$, $p<0.001$] in the direction described above. There was no effect of Ambiguity and no significant interactions between any of the variables.

Table 5.5: Mean percentages (and standard deviations) for the judged likelihood of the change in reality, as described in the story, actually occurring.

Version	Predictable	Unpredictable	Overall mean
Ambiguous	39.4 (24.9)	22.2 (22.8)	30.4 (25.1)
Unambiguous	40.7 (24.1)	18.0 (20.7)	29.6 (25.0)
Overall mean	40.1 (24.2)	20.1 (21.6)	-

5.2.2.3. Category judgments of probability of change. Participants were also asked to categorise their estimated percentages according to a pre-specified list of varying frequency labels. The list was divided into High Frequency and Low Frequency categories. High Frequency ranged from 'Every day' to 'Once a month' inclusive, and Low Frequency ranged from 'Less than once a month' to 'Almost never.' High and Low Frequency responses were classified (a) according to Predictability, to see whether the manipulated level of predictability was reflected in participants' perception of the likelihood of change actually occurring, and (b) according to Ambiguity, to see if categorical responses were influenced by the ambiguity of reference (summarised in Table 5.6).

Firstly, there was a significant correlation between Predictability and perceived categorical Frequency [χ^2 (corrected, $\underline{df} = 1$, $\underline{n} = 95$) = 8.2, $\underline{p}<0.005$]. Participants were significantly more likely to judge that Predictable scenarios would occur more frequently than Unpredictable scenarios. Secondly, in line with the lack of an effect for Ambiguity on the mean percentage scores outlined above, there was no significant correlation between Ambiguity and perceived categorical Frequency [χ^2 (corrected, $\underline{df} = 1$, $\underline{n} = 95$) = 0.3, n.s.].

Table 5.6: Number of participants giving High or Low Frequency responses to Question 3, according to Predictability of version and Ambiguity of utterance.

Response / Version	High Frequency	Low Frequency
Predictable	36	11
Unpredictable	22	26
Unambiguous	28	21
Ambiguous	30	16

5.2.3. Discussion

There was no evidence to suggest that adults were influenced by their own knowledge of reality when making evaluations of ambiguous utterances. Very few participants suggested that the listener would ask for more information from the speaker in the Ambiguous condition. Indeed, there were more 'ask more' responses in the Unambiguous condition. This suggests that these adults did not

seem to be judging according to their knowledge of the referential success of the utterance.

The predictability of the alteration to reality occurring did not seem to influence evaluations either. Participants were not more likely to judge that the listener would ask for more information if the change was a predictable rather than an unpredictable one. The lack of an effect for predictability cannot be attributed to participants' lack of sensitivity to the manipulated likelihood of the event occurring. There was a significant correlation between the experimenter's and the participants' ideas of what constituted an unpredictable or a predictable change. Moreover, there was a significant difference between the unpredictable and predictable scenarios in terms of the percentage frequencies participants provided to indicate the likelihood that more books would be added to the desk in the manner described. Therefore, the stimuli provided seemed to be successful in manipulating perceived predictability, but this did not influence utterance evaluations from the perspective of the listener.

5.3. General Discussion

The aim of these two experiments was to see whether adults are 'contaminated' by their own knowledge of reality when making evaluations of utterances. Adults were provided with scenarios in which they knew the utterance was ambiguous, but neither the speaker or listener story protagonists were aware of this fact. If adults are not influenced by their knowledge of the ambiguity of reference, they should treat this utterance identically to one in which the speaker

and listener have identical belief frameworks to the ambiguous scenario, with the only difference being that the utterance remains pragmatically adequate.

There was some evidence that adults were influenced by their privileged knowledge in the first experiment. They were significantly more likely to judge that the listener would ask for more information from the speaker in the Ambiguous condition compared to the Unambiguous condition. However, this effect was confined to one particular story scenario involving buying a jumper from a shop. Participants may have judged that the listener would ask for more information because it is quite plausible that more jumpers could be added to the display in the period between the utterance being made and the listener seeing the shop window. Therefore, judgments may have been based on the perceived predictability of the change occurring rather than on the knowledge that the change had actually occurred.

This possibility was assessed directly in the second experiment, where the predictability of the change to the array was manipulated in addition to the ambiguity of the utterance. Utterance evaluations were not influenced by the level of predictability, suggesting that the significant effect found in the first experiment may have been a result of a realism bias. However, this is an extremely tentative argument given that the story in the second experiment did not produce a realism effect. It might still be that participants considered a change in reality to be more likely in the Jumper story than in the Bar or Library story, resulting in the significant tendency to judge that the listener would ask for more

information. This possibility would need to be tested directly using predictable and unpredictable versions of the Jumper story.

The failure to find an effect in the Library (Experiment 5.2) and Bar (Experiment 5.1) stories suggests that, if there is a realism bias in utterance evaluation, it is only a weak and context-specific effect. The elusive nature of the trend may be exacerbated by the mode of presentation of the stories. In the present procedure, participants were asked to read vignettes from a sheet of paper. This may encourage a more deliberative and analytical approach because the relevant information can be isolated within the text and important points can be re-read by the participant. Consequently, the belief states of the speaker and listener protagonists may be especially clear to participants, because they can reflect on the circumstances within which the beliefs are formed.

Additionally, the array (at the time the speaker made their utterance) was not depicted, limiting the likelihood that participants could use an utterance-array comparison as the basis for their evaluations. Again, this might explain why the Jumper story generated a significant effect. It may have been easier for participants to imagine an altered array in the Jumper story compared to the Bar or Library stories, based on their familiarity with such a scenario (Kahneman & Tversky, 1982).

Consequently, follow-up work would need to provide a clear depiction of the array, preferably within a more realistic setting. For example, videos provide an ideal medium for the presentation of scenarios, allowing a more naturalistic display of communication. Mitchell, et al. (1996) used video scenarios to elicit a

realist bias in false belief reasoning in adults. It could be that the 'contaminating' effect of privileged knowledge works best when participants do not have the opportunity to pore over details of the story. Instead, they are required to respond in a way that is more reminiscent of the rapid and dynamic nature of real human interaction.

Methodological concerns aside, it might be that the lack of a consistent significant effect reflects the fact that adults are quite resistant to the influence of their privileged knowledge when making utterance evaluations. This might not be so surprising given the importance of considering the speaker's own conceptual framework during communication. As Mitchell (1999) suggests, 'What people say enlightens us about the perspective they have on reality; it tells us what kind of interpretation they favour of the events all around' (p.189). Therefore, if communication is to be meaningful and successful for the people involved, it is crucial that the speaker's belief, or meaning, is considered in addition to the literal words of the utterance.

5.4. Conclusion

It is perhaps reassuring that adults in the present studies seemed to be well attuned to the role of the speaker's belief in the creation of utterances. Rather than favouring a literal interpretation of the utterance (and, as a result, focusing on the utterance-array link), adults often responded in a way that suggested they were able to separate their own knowledge from that of a naïve listener, attributing more importance to the speaker's conceptual framework. This suggests that this

group of adults, at least, have a good understanding about the role of the utterance as an externalisation of the speaker's internal representation.

CHAPTER 6

General Discussion

6.1. Introduction

This thesis was primarily concerned with the referential triangle that characterises verbal communication. The three elements of the triangle - the utterance, the speaker's internal representation and reality - contribute to the production and interpretation of utterances. Sometimes it may not be sufficient to consider only the speaker's utterance and how it relates to reality. If the utterance underspecifies reality in some way, then a consideration of the speaker's belief or internal representation may be central to understanding what the speaker really means.

A crucial developmental milestone involves the child coming to understand that what is said is a representation of what is meant. That is, the child needs to understand that utterances are externalisations of a speaker's internal representation. Additionally, the developing child needs to understand that an utterance has a special relationship with reality insofar as the words of the utterance must adequately describe an object in order to allow a unique identification of the object from a set of alternatives. Therefore, understanding an utterance as an externalisation of an internal representation requires an awareness

of how the utterance, the speaker's meaning and reality relate to each other within referential communication. One of the limitations of traditional referential communication research is that usually only one of these relationships is considered at a time. Consequently, it is not clear how children treat these relationships when they have to consider them simultaneously. In other words, little is known about how children and adults handle the triangular relationship when all three elements need to be considered.

The thesis consists of two main approaches to investigating how people respond when all three elements of referential communication are manipulated. The first approach involved presenting the utterance, the array and the speaker's meaning as separate, substantive elements with the help of cartoon thought and speech bubbles. The second approach involved presenting children and adults with discrepant utterances: what was said and what was meant were isomorphic, but discrepant with the current state of reality. As a result, it was possible to examine whether people considered the (discrepant) relationship between the array and the utterance, or the (consistent) relationship between the belief and the utterance to be more important in judging message adequacy.

Both approaches required participants to make evaluations about utterances. Evaluation forces participants to be explicit about why an utterance may be adequate or inadequate. That is, *evaluation* provides a greater opportunity to consider all three elements of the referential triangle. In contrast, participants can *interpret* utterances correctly without necessarily having to consider all three

elements of the triangle and the resultant relationships between the elements (see pp. 109-110 for an expansion of this point).

The thesis was not designed to test a specific theory of referential communication, although different theories are relevant to each chapter. The aim was to provide an exploration of how people evaluate utterances when the tripartite relationship between the utterance, the array and the speaker's meaning has to be considered directly. Nevertheless, the overall picture from the results can be interpreted within certain theoretical frameworks. The main results from each of the two approaches to investigating the referential triangle will be summarised below before considering the theoretical, as well as the broader, implications of the findings.

6.2. Presenting the referential triangle as substantive elements

The value of using cartoon thought bubbles to depict thought content was demonstrated in Chapter 2. Normally developing children and children with autism were significantly helped to answer correctly about a story protagonist's false belief when they saw the thought content depicted in a thought bubble, compared to a standard false belief task in which no thought bubble was present. Both groups of children were able to answer correctly about reality also, suggesting that they understood how the protagonist could hold a belief that was discrepant with the current state of reality. Therefore, they seemed to understand thought bubbles in an appropriately representational way.

Given this facilitation, thought bubbles provided an ideal opportunity to present the speaker's meaning in a concrete and tangible way in a referential communication task. In Chapter 3, thought bubbles allowed the speaker's meaning to be represented separately from the utterance and from reality. In this approach, the speaker's meaning could be understood as an internal representation and not just as the intended outcome of the communication. This is a crucial distinction within the referential triangle, since meaning and reality were no longer confounded in terms of how they were presented to the children (cf. Beal & Flavell, 1984). Therefore, even though speaker meaning and reality were isomorphic in this procedure – as in traditional ambiguity research (Robinson & Whittaker, 1987) – it was possible to investigate whether children were more concerned with utterance-array relations or utterance-meaning relations because children could justify their responses with reference to either of these links.

In the ambiguous condition, and in the absence of a thought bubble (Experiment 3.2), the majority of children aged 6-10 years indicated that the speaker had not said exactly which item he/she wanted. These children overwhelmingly indicated that this was because the array was ambiguous (e.g. 'there were four yellow bags'). By contrast, although the majority of 6-10-year-olds again evaluated ambiguous utterances negatively in the presence of a thought bubble, their stated reasons for doing so were markedly different. Rather than concentrating on the relationship between the utterance and the array, they focused on the say-mean discrepancy instead. That is, children said that the speaker had not provided enough information or had not said exactly the same

words as those contained in the thought bubble. This tendency was especially strong amongst the 6-7-year-olds and then diminished with increasing age. The older children showed an increasing preference with age for stating the ambiguity between the utterance and the array as the source of the problem with the message, in the presence of a thought bubble. In other words, they overlooked the discrepancy between what was said and what was meant and focused on the pragmatic relationship instead. This suggests that the 9-10-year-olds, and a group of adults, treated the literal meaning (i.e. how the utterance maps on to the array) as more important for determining whether an ambiguous message is adequate or not (cf. Olson, 1970).

Given this understanding when responding to ambiguous messages, it was perhaps even more surprising that children aged 6-10 years evaluated the utterance negatively in the unambiguous condition. Rather than demonstrating a diminishing tendency to evaluate according to the utterance-meaning relationship, as in the ambiguous condition, a consistent proportion of the 6-10-year-olds (around 50%) judged unambiguous utterances negatively. The main reason given for a negative evaluation was the speaker not describing the item in enough detail. Very few children answered in this way in the absence of a thought bubble, suggesting that evaluations and justifications were heavily influenced by seeing the speaker's meaning written in a thought bubble. In other words, children seemed to respond as if they thought that an utterance can only be adequate if what is said and what is meant are identical. They seemed to have difficulty understanding that an utterance can be referentially adequate (it uniquely

identifies the referent in an array) even though speech might underspecify thought.

In a more stringent within-subjects test of the influence of seeing the speaker's meaning encapsulated in a thought bubble (Experiment 3.3), only the 7-8-year-olds showed a similar tendency to judge unambiguous utterances negatively. The same group of 7-8-year-olds were also significantly more likely to judge, in the presence of a thought bubble, that a listener would not be able to pick the right object. This was despite the fact that the speaker's utterance allowed a unique identification of the item in the array. This group of children seemed to overlook pragmatic adequacy when they were required to make an evaluation of the utterance, and they also overlooked the pragmatics of the communication when they were specifically focused on the utterance-array relationship.

Overall, Chapter 3 presented children with situations in which utterance and meaning differed, but reality and meaning were isomorphic. In the ambiguous condition, the utterance underspecified the array such that the referent could not be uniquely identified. Hence, there was a discrepancy between the utterance and the array as well as between the utterance and thought content. At around the age of 6 or 7 years, children focused on the say-mean discrepancy when evaluating ambiguous utterances. For example, they evaluated the utterance negatively 'because she didn't say 'flowers' (i.e. the description contained in the thought bubble). After this age, there was a decreasing tendency to state the say-mean discrepancy as the reason why the message was inadequate and a corresponding

increase in statements that it was the ambiguous array that was the source of the problem.

In the unambiguous condition, 6-10-year-old children focused on the say-mean discrepancy to the extent that they overlooked the pragmatic adequacy of the utterance. That is, they did not seem to realise that the redundant information in the thought bubble could be ignored because the utterance alone allowed a unique identification of the referent. These children behaved as if the presence of a say-mean discrepancy indicated that the message was no good, irrespective of the adequacy between the utterance and the array. Robinson and Whittaker (1986) suggest that "In order to realize that messages can be ambiguous, it is necessary for children to see that messages are representations of, and hence distinct from, intended meanings" (p.42). This understanding should apply equally to situations in which the utterance is unambiguous. However, this seems to be especially difficult for children, perhaps because an isomorphism between thought and speech is always assumed in the case of unambiguous utterances.

Of course, the situation of depicting thought content directly is somewhat artificial. Perhaps it should not be surprising that children found it difficult to reconcile the idea of a say-mean discrepancy with a referentially adequate utterance. Interestingly, though, Robinson (1994) suggests that experimental designs which give "children...privileged access to the speaker's belief, may allow children to demonstrate understanding of links between utterance, reality, and belief in a way in which they could not in more complex and everyday situations" (p.377). In other words, showing the speaker's meaning in a thought

bubble may provide an ideal situation in which it is possible to investigate the strengths and weaknesses in children's understanding of utterances as externalisations of internal representations.

6.3. Discrepant utterances

The source of the discrepancy within the referential triangle differed in Chapter 4 compared to Chapter 3. In the Not See condition, the discrepancy arose between reality and meaning, whilst speech and meaning were isomorphic. Consequently, there was also a discrepancy between speech and reality. Chapter 4 utilised a different paradigm to the one included in Chapter 3 so that it was possible to see whether participants spontaneously evaluated utterances according to the relationship between thought and speech when speaker meaning had to be inferred rather than depicted directly. A test question was included that deliberately did not specify which criterion was the 'correct' one for utterance evaluation (Did speaker do a good job of saying exactly which item he/she wanted?). This was so that the relative importance of the relationships within the referential triangle could be ascertained because it was up to the individual to decide which link was of primary significance. It was possible that participants could have used their knowledge that the speaker held a false belief to evaluate the utterance according to the speaker's representation, rather than according to the utterance-array relationship.

Children aged 6-10 years overwhelmingly chose to focus on the relationship between utterance and reality when evaluating an utterance that was

isomorphic with the speaker's belief, but outdated in terms of how well it characterised the current state of reality (Not See condition; Experiments 4.2 and 4.3). That is, children judged that the speaker had not done a good job of saying exactly which item he/she wanted because the array was ambiguous ('there are four bikes there') or because the utterance underspecified the item in the ambiguous array ('he didn't say the colour'). This pattern of responding was also found with a question that asked children to evaluate the utterance from an uninformed listener's perspective (Do you think the listener just goes to fetch the item, or does she ask the speaker any more about it?; Experiment 4.3). Children judged according to the pragmatic adequacy of the utterance. They said that the listener would ask the speaker for more information about the utterance because the utterance was ambiguous with respect to the array. Only adult participants evaluated utterances in the Not See condition according to the good match between the speaker's belief and their ensuing utterance. The adult group said that the speaker had done a 'good job' despite the fact that what was said was at odds with the known state of reality (Experiments 4.1 and 4.4). However, this response pattern was characteristic of only around half of the adult sample at best; the remainder of the group preferred to focus on the discrepancy between speech and the array instead.

The tendency to respond according to the utterance-array relationship was also very strong in the See condition. In this condition, the listener protagonist saw the change to the array and produced an ambiguous utterance. The speaker's internal representation of the array was consistent with reality but their utterance

did not allow a unique identification of the referent. There was no obvious reason why participants *should* consider the speaker's internal representation in this case, since the locus of the problem lay between the utterance and the array (Robinson & Mitchell, 1992; Robinson & Whittaker, 1987). Indeed, most children and adults specifically mentioned the ambiguous array or the underspecification of the utterance in relation to the array when explaining why the speaker had done a 'bad job'. There seemed to be few problems with 'mapping' the utterance on to the array when the speaker's internal representation was consistent with reality.

Curiously, many participants (children and adults) did not use this 'mapping' strategy between utterance and array in a Control condition in which all three elements of the referential triangle were in agreement. That is, the speaker's utterance was consistent with their belief, which was, in turn, consistent with reality. Rather than judging that the speaker had done a good job of saying exactly which item he/she wanted, children aged 6-10 years, and many of the adult participants, suggested that the speaker had done a 'bad job' based on the likelihood that a change to the array could occur in the future that would render the speaker's utterance ambiguous. In other words, many participants chose to invoke a discrepancy between the utterance and the array rather than judge according to the information that was presented to them.

It was hypothesised that the tendency to base judgments on knowledge of a discrepancy between the utterance and the array might be so strong that participants would continue to focus on this even in a situation where it was inappropriate to do so. There was some suggestion in the results of Experiment

4.4 that at least half of the adult participants evaluated the utterance according to the speaker's belief state in the Not See condition. They judged that the speaker had done a 'good job' because he/she was ignorant about the change to the array. However, when the same adults were asked to evaluate the utterance from an uninformed listener's perspective, they failed to consider the speaker's internal representation and, instead, judged according to the mismatch between the utterance and the array. That is, they said that even though the speaker had done a good job of saying exactly which item he/she wanted, the listener would, nevertheless, ask the speaker for more information about the item. Given that the speaker was not being deliberately vague or misleading and that the participant was aware of the match between the speaker's belief and utterance, the responses to the two different questions concerning message adequacy should have been the same. The listener protagonist did not know about the change in reality and so reference to the array in terms of how the naïve listener would interpret the utterance, was no longer appropriate. In short, it seems participants' knowledge about the discrepancy between the speaker's utterance and the altered array influenced their judgments of message adequacy from the perspective of the listener.

A problem with this interpretation was that adults also said that the listener would ask for more information from the speaker in the Control condition. If participants were judging message adequacy according to the utterance-array link they should have judged the message to be adequate in this case because there was only one object that matched the speaker's description. However, it has already

been noted above that there was a strong tendency in the Control condition to invoke hypothetical changes to the array that might render the (unambiguous) utterance ambiguous. Consequently, the trend for participants to judge that the listener would ask for more information from the speaker in the Control condition, could be based on the future, hypothetical discrepancy between the utterance and the array. In other words, although both the Control and Not See conditions produced response patterns consistent with perceived discrepancies between the utterance and the array, each pattern could have a different basis: *real* in the Not See condition and *assumed* in the Control. As a result, there might have been a tendency for adults to overapply their knowledge of the utterance-world discrepancy, but the Control condition did not provide a suitable contrast due to the ease with which extra-contextual information was included in the stories.

Experiments 5.1 and 5.2 were devised to directly test the possibility that adults were inappropriately evaluating the utterance according to their own knowledge of the utterance-world discrepancy. Based on the results of Experiment 4.4, it was crucial to try to establish a Control condition in which the likelihood that participants would consider future states of affairs was minimised. In Experiment 5.1, the speaker's belief that there was only one item that matched his/her utterance was emphasised, in addition to the fact that no more items were added to the array in the unambiguous condition. Under these circumstances, participants were much more likely to judge that the listener would just go to the mentioned location rather than ask the speaker for more information. By contrast, even though the speaker's and listener's belief states were identical in the

unambiguous and ambiguous conditions, participants were significantly more likely to judge that the listener would ask the speaker for more information about the utterance in the ambiguous condition, when they (the participant) knew that the utterance was discrepant with reality. In other words, participants once again seemed to overlook the isomorphism between speaker's utterance and belief, in favour of their own knowledge pertaining to the discrepancy between utterance and the array (a 'realism' effect; cf. Mitchell et al, 1996).

Experiment 5.2 investigated the possibility that participants in Experiment 5.1 were simply responding according to the predictability of the change to the array occurring. In the ambiguous condition, a change to the array takes place without the speaker's knowledge. However, because the alteration has happened, the participant might think that this is a common event which, presumably, the listener would also realise. Therefore, the judgment that the listener would ask for more information from the speaker in the ambiguous condition may arise because participants reason it is likely that the change *would* occur, rather than because they *know* the change has actually occurred. In Experiment 5.2, predictable and unpredictable versions of a story scenario were crossed with unambiguous and ambiguous conditions, to see whether adults would still be susceptible to a 'realism' effect. This would be evidenced by judgments that the listener would ask for more information from the speaker in the ambiguous condition, even when the alteration to the array was unpredictable.

Adult participants were sensitive to the manipulated predictability of the change to the array actually occurring. They discriminated significantly between

unpredictable and predictable versions of the story when asked to estimate the frequency with which the change in the story might occur. Importantly, though, utterance evaluation was not influenced by the predictability of change. There was no difference between predictable and unpredictable versions in terms of judgments of whether the listener would ask the speaker for more information about the desired object. Also, contrary to the hypothesis, there was no significant tendency for adults to judge that the listener would be more likely to ask for additional information in the ambiguous compared to the unambiguous condition. This group of adults appeared not to be influenced by their knowledge that the array had changed when considering utterance adequacy from the perspective of a listener protagonist. However, there were a number of methodological flaws, which may have contributed to the failure to find a significant effect. Perhaps the most important of these was the fact that the array was never depicted, thereby limiting the opportunity for utterance-array comparisons to take place.

Overall, the results in this thesis show that children attend to the speaker's meaning under some circumstances, whilst under different circumstances they attend to the literal meaning of the utterance. In summarising a similar pattern of findings, Robinson and Whittaker (1986) suggest that children's difficulties with knowing which relationship to focus on might arise because understanding referential communication requires "...seeing the message both as a stimulus to be interpreted in its own right and also as representation of a speaker's intended meaning which may on some occasions be inadequate" (p.43). In other words, the problem with understanding the tripartite nature of referential communication is

due to the complicated relationship between the elements of the referential triangle! This is tautological and does not explain why the difficulties might arise. What follows, therefore, is a summary of two, broad, competing theories that might go some way to explain why the tripartite relationship that exists in referential communication could be problematical for utterance evaluation.

6.4. Theoretical explanations

The first account that follows is based on the 'reality masking' hypothesis proposed by Mitchell (e.g. 1994). He suggests that children and adults may find reality particularly salient when making judgments about knowledge and beliefs. This may result in the content of the mental representation being overlooked in favour of what is known about reality. In terms of the referential triangle, a focus on reality would be suggested by judgments based on the array rather than the speaker's belief. Additionally, the 'reality masking' hypothesis predicts that if a mental representation can be made as salient as reality, a judgment based on the mental representation is more likely because reality no longer dominates responses. The implication for referential communication, is that if the speaker's belief and reality are equally salient, evaluations based on the speaker's belief might predominate. These possibilities might help to explain some of the results in this thesis and are explored in Section 6.4.1.

A second explanation for the range of results obtained is based on the idea that understanding the tripartite nature of referential communication may be hard because discrepancies may exist between the elements in the triangle.

Understanding that what is said can differ from what is meant is an important developmental milestone. Hence, a say-mean discrepancy can provide a strong clue about message adequacy. There might be some circumstances, though, under which a discrepancy between thought and speech can be overlooked because the utterance is pragmatically adequate. The important distinction that needs to be made is not whether there is a discrepancy between certain elements in the triangle, but whether it is important to use the discrepancy as the main criterion for evaluation. This account is discussed in more detail in Section 6.4.3.

6.4.1. The Reality Masking Hypothesis.

Mitchell (1994; 1996; also Saltmarsh, Mitchell & Robinson ,1995) argues that reality holds an attentional salience for children because it is more important for children to negotiate the real world in their early years, rather than the mental realm. It is crucial for the child to understand how things work in the world that surrounds them, so that they can function safely and effectively. A bias towards reality may help to reduce the unpredictability of a dynamic and complex environment. Moreover, current reality, rather than belief, may be salient because, until the age of around 3 or 4 years, a child is usually looked after by her family who meet all of her physical needs. This is not because the child has had to engage in any subtle or manipulative behaviours to coerce the family to behave in this way, but rather this situation arises by default. Accordingly, beliefs have a low attentional priority for the pre-school child because her needs are met automatically by the family. It is only when the child enters a different

environment, such as nursery education or school, that this situation begins to change. In such an environment, it is extremely unlikely that all of the people surrounding the child will cater solely for her needs and desires. Instead, their time is divided between many individuals, compelling children to give a high priority to other people's beliefs in order to make friends and influence others' behaviour.

According to this account, the pre-school child has some insight into the representational nature of the mind but this is often masked by the salience of reality. This stands in contrast to Perner (1991) who argues that children have no representational understanding of the mind before the age of about 4 years. He suggests the child undergoes a cognitive revolution at the age of four years which furnishes them with the necessary conceptual understanding that beliefs are representations of reality. Hence demonstrations that children could understand the mind as representational before the age of four years would prove damaging to Perner's claim, whilst providing supporting evidence for the reality masking account.

Such evidence involves paradigms that elevate belief to an equivalent par with reality. For example, Mitchell and Lacohee (1991), and Freeman and Lacohee (1995), found that 3-year-old children showed impressive abilities to pass false belief tasks when their previous belief had been bolstered by a counterpart in reality (e.g. a picture of the child's belief was posted into a toy postbox before the test questions were asked). Saltmarsh and Mitchell (1999) found similar results when children were presented with video evidence of their

false beliefs. Generally, equating reality and belief in terms of level of physical salience and tangibility allowed children to demonstrate their understanding of the mind because their attention was no longer dominated by reality.

Robinson (1994) suggests that utterances may be difficult to understand as clues to internal representations because they offer no protection from the "...salience of immediately perceptible reality..." (p.373). In other words, because utterances are (usually) ephemeral and non-substantive it is difficult to treat them as emanating from, and relating to, the mind because reality (in this case, the array) captures the child's attention. In Chapter 3, it might be that many children did not respond according to how the message related to the array because the presentation of the utterance alongside the belief/meaning of the speaker encapsulated in a thought bubble protected children from the perceptual draw of reality. That is, children could consider how the utterance was related to speaker meaning because both were depicted at an equivalent salience to reality.

A similar effect was observed in Chapter 2, where the presence of a thought bubble in a false belief task seemed to protect children against realist errors and enabled them to judge appropriately about belief. Ironically, the operation of this effect in Chapter 3 may have meant that children judged 'incorrectly' in the unambiguous condition because the meaning of the utterance was made as salient as the array. Consequently, the pattern of results in Chapter 3 may be more indicative of children focusing on a mentalistic criterion because the array no longer dominated their attention, rather than a fundamental problem with understanding the pragmatic value of the utterance (but maybe not – see below).

The results of Chapter 4 are also consistent with the idea that social cognition may be biased towards reality. On the majority of occasions, children and adults evaluated ambiguous utterances according to their relation to the array (i.e. reality) rather than according to the speaker's meaning. It seems unlikely that participants did not understand that what was said was a representation of what was meant, especially since signs of this kind of understanding emerge at around the age of four years (e.g. Mitchell, Robinson and Thompson, 1999). Instead, when there was a choice between judgments based on reality versus judgments based on belief, participants tended to focus on the former rather than the latter. According to Russell et al. (1991) and Mitchell's reality masking account, this is because physical reality is more salient than the mental domain.

It is worth noting that the conditions presented to participants in Chapter 4 resembled the 'real world' more closely than the paradigm utilised in Chapter 3, in the sense that 'reality' was perceptually salient, whilst the speaker's belief had to be inferred from their behaviour and from their utterance. By contrast, the array was not depicted in Chapter 5, perhaps making it less likely that participants would focus on reality at the expense of belief. Indeed, in Experiment 5.2 where reality and belief were equally salient because both had to be generated from a story narrative, participants behaved in a way more consistent with them taking the speaker's belief into account. That is, in both the unambiguous and ambiguous conditions participants judged that the listener would just go to the mentioned location rather than ask the speaker for more information. This suggests that adults assumed that the speaker was being truthful and sufficiently informative in

both cases and so judged in accordance with the speaker's (inferred) belief rather than in accordance with their own privileged knowledge of reality.

6.4.2. Limitations to the Reality Masking account.

Many of the present results can be accommodated within this explanatory framework, but there are some aspects of the findings that do not fit so comfortably within the 'reality hypothesis.' In particular, The Control condition was often treated as inadequate in Chapter 4. Given that reality was perceptually salient for participants in this paradigm, and they were responding in line with the predictions of the reality hypothesis, participants should have focused on the (adequate) relationship between the utterance and the array and judged the utterance positively. Instead, many child and adult participants evaluated the utterance negatively in this condition either because the speaker did not describe the desired item in enough detail or because they created imaginary, future scenarios in which the utterance was rendered ambiguous by a change to the array. It is difficult to see how the reality hypothesis can account for this tendency.

6.4.3. The Discrepancy account.

An alternative explanation that could account for the tendency to treat the Control condition as inadequate in Chapter 4 is based on the idea that children may be focused on discrepancies rather than reality. Overall, the majority of the evidence presented in the thesis is consistent with the idea that a known

discrepancy between elements in the referential triangle may be particularly difficult to overlook when making judgments about message adequacy. Children seem to have difficulty overlooking the discrepancy between thought and speech when evaluating unambiguous utterances (Chapter 3); and children and adults evaluated messages according to an utterance-array discrepancy over and above a match between utterance and belief (Chapter 4; Experiment 5.1). In Chapter 4 and Experiment 5.1, participants may have focused on the discrepancy between the utterance and the array, not because of the salience of reality, but because that was the source of the discrepancy. Moreover, it is possible that a preference for discrepancy within the referential triangle is so strong that participants chose to *invoke* a discrepancy within the referential triangle rather than respond to the information provided for them. Hence, the utterance in the Control condition in Chapter 4 was often evaluated as inadequate.

Preferences for discrepancies are not unusual and exist in domains other than utterance evaluation. In social psychology, Hansen and Hansen (1988) showed that adult participants easily picked out a single angry face amongst a crowd of happy faces. Bargh (1989) suggests that this tendency arises because our attention is automatically drawn to things that are unusual or different. So, in the above example, it is not that we notice an angry face in the crowd and choose to pay attention to it. Rather, our attention is 'captured' by the fact that the face is different from the others, and only then do we notice that the expression on the face is one of anger. Friedman (1979) demonstrated a similar phenomenon in a paradigm investigating encoding processes in memory. Participants were required

to remember complex representations of rooms and were then shown a number of stimuli that either matched or differed from the original map. At the most basic level, the results indicated that inconsistent or unexpected information was remembered more frequently than predictable and compatible information.

The discrepancies that exist between different elements within the referential triangle might dominate response criteria simply because they are salient and 'capture' our attention (cf. Bargh, 1989). However, the story might not be as simple as that statement suggests. In Experiment 3.2, children showed a developmental trend in their evaluations of ambiguous utterances. In the presence of a thought bubble showing the speaker's meaning, 6-7-year-olds had a strong preference to judge that the speaker had not said exactly which item he/she wanted on the grounds that the utterance-belief relationship was discrepant. By contrast, older children were much less likely to state the utterance-belief link as the source of the problem, and showed a growing tendency towards saying that the speaker had not said exactly which item he/she wanted on the grounds that the utterance was discrepant with the array. The main point is that there were two sources of discrepancy in this condition. The younger children focused on the utterance-belief relationship, whilst the older children focused on the utterance-array relationship. It is unlikely that one discrepancy is intrinsically more salient than another. If this were the case, most children, irrespective of age, would notice the more salient discrepancy and judge accordingly. The change in response criteria with age is more suggestive of a change in children's understanding of

which (discrepant) relationship was important for evaluating the quality of a message.

In the unambiguous condition (Experiments 3.2 and 3.3), in the presence of a thought bubble, there was only one source of discrepancy - between the utterance and the speaker's meaning. Around half of the children aged 6-10 years judged in accordance with this discrepancy and seemingly did not apply their understanding of the importance of the relationship between the utterance and the array. This tendency could arise because discrepancy is more salient to children than, in this case, the pragmatic adequacy of the message. Therefore, even though children might know that the utterance-reality link is more important under these circumstances, they might not apply this knowledge because their attention is drawn to the say-mean discrepancy. However, again there is some suggestion that the problem is not just one of the salience of the discrepancy, but rather, that children might have more of a conceptual difficulty in understanding the role of discrepancies in referential communication.

Many of the children indicated that the speaker had said exactly which item he/she wanted in the presence of a thought bubble. However, a surprising number of these children (including 58.6% of the 9-10-year-olds in Experiment 3.2) were unable to explain why they had given a positive response. Therefore, unlike the ambiguous condition, in which the vast majority of children were able to give some indication of why they thought the message was bad, the children who answered positively in the unambiguous condition had problems saying why they thought the message was good. It was as if they had difficulty reconciling

their knowledge of the discrepancy between thought and speech with their knowledge of the adequacy between the utterance and the array.

Generally, children seem to accept quite readily that discrepancies can occur between the elements in the referential triangle. However, the reason why understanding utterances as externalisations of internal representations might be difficult is because this also requires an understanding of when it is *appropriate* to give the discrepancy primary importance, versus situations in which the discrepancy should be overlooked in favour of an adequate match between other elements in the triangle. When children were given the opportunity to compare thought bubble absent and present trials in Experiment 3.3, they were much better at overlooking the say-mean discrepancy and, instead, judged the unambiguous utterance positively because it was pragmatically adequate. Being able to compare the conditions may have provided clues about when it was acceptable to ignore the discrepancy between what was said and what was meant.

Some researchers suggest that coming to understand that what is said can differ from what is meant is fundamental to understanding referential ambiguity (e.g. Beal, 1988; Bonitatibus, 1988a). Once children understand that say-mean discrepancies are important for determining communicative success, they can use the existence of such a discrepancy as a clue to the fact that the message might be problematical. Children might initially focus on the discrepancy because they realise that the say-mean link can tell them a great deal about the adequacy of verbal messages. After this time, children realise that the existence of a say-mean discrepancy simply serves as a useful clue to the fact that a referentially poor

utterance has been produced, and that it is the link between the utterance and the array that is crucial as far as picking the correct object is concerned (cf. Olson, 1970). The results from the unambiguous conditions suggest that children's understanding of the role of say-mean discrepancies in referential communication might not be fully formed. It seems that they have difficulty understanding how an utterance can still be adequate, even though what is said and what is meant might differ. Instead, children seem to assume that if a say-mean discrepancy exists, the utterance must be inadequate.

An awareness of the relevance of discrepancies within the referential triangle must develop alongside an understanding of the potential fallibility of referential communication. Experience of communication failures may teach us to be more cautious in accepting exactly what we are told. As a result, older children and adults might have a greater understanding of the importance of discrepancies within referential communication in addition to a more sceptical approach towards the veracity of verbal utterances. A combination of these insights could help to explain why the adult participants and older children in Experiments 4.2, 4.3 and 4.4 were more likely than younger children to judge that the utterance was inadequate in the Control condition. They could reason that communication is rarely as simple as the Control scenario suggests. The possibility that all three elements in the triangle could be in perfect harmony is unlikely and so a discrepancy is invoked between the utterance and the array in order to demonstrate an understanding that communication is likely to go wrong in this

instance. As a consequence, it is the discrepancy that reveals the source of the problem with the utterance and so the message is evaluated negatively.

Note also that the discrepancy account only applies to those children who recognise that discrepancies between elements in the triangle are possible. This explains why 5-6-year-olds did not judge in accordance with the discrepancies. In response to ambiguous utterances, they judged that the speaker had said exactly which item he/she wanted (Chapter 3) and that the speaker had done a good job of saying exactly which item he/she wanted (Chapter 4). This again suggests that judgments based on discrepancies are not made simply because the discrepancies are salient. If this were the case, presumably the 5-6 year-olds would also focus on the salient aspects of the stimuli and respond accordingly. Instead, it seems that there is a developmental progression in terms of understanding *that* discrepancies can occur and *when* it is appropriate to evaluate utterances in accordance with them.

6.4.4. Limitations to the Discrepancy account.

Perhaps the finding that around half of the adult participants in Experiments 4.1 and 4.4 did not make utterance evaluations in accordance with the discrepancy between the utterance and the array in the Not See condition suggests that the discrepancy account is limited. However, the intention is not to suggest that a discrepancy criterion is unthinkingly applied to utterance evaluations, but rather that there is an increasing understanding with age about the relative importance of discrepancies within the referential triangle.

It was suggested in Chapter 4 that making a 'good job' evaluation in the Not See condition appears to be a sophisticated response criterion, available only to adult participants. This response may require quite a substantial amount of processing in order to move past the literal reading of the utterance to access the speaker's belief (cf. Gilbert, 1991). Given that the discrepancy between the utterance and the array is usually a good indicator that the message is problematic, participants may adhere to the cognitively less challenging heuristic of judging on the basis of the discrepancy rather than engaging in the extra work required to evaluate according to the speaker's belief. In short, a discrepancy may serve as a particularly useful clue concerning the adequacy of the message and so remains the dominant criterion on which evaluations are based. Overriding the powerful cue of the discrepancy may require extra effort, and may not even be especially useful. For example, even though the speaker may have done a good job according to his/her own knowledge, this might not be helpful to the listener because they are unable to uniquely identify the desired referent.

6.5. Is it possible to arbitrate between the Discrepancy and Reality accounts?

It would be naïve to think that evaluation of utterances in referential communication could be explained solely on the basis of one of the above accounts. There are many factors that influence communication and how it is interpreted and evaluated by the interlocutors. For example, the decision to adhere to a discrepancy or reality-based evaluation is likely to be influenced by contextual factors, including the shared common ground between speaker and

listener and memory factors influencing the access to such shared meaning or knowledge (Ackerman, 1993). Additionally, the present discussion focuses on utterance evaluation from an observer's or listener's perspective. At no point is there any claim being made about a speaker's *production* of utterances and how the role of discrepancy within the referential triangle might be important in the formulation and shaping of utterances. That would require a whole other thesis!

Nevertheless, the Reality and Discrepancy accounts both seem to have quite wide explanatory value, at least as far as the present results are concerned. To move the research forward, it would be interesting to explore (a) whether participants do seem to have a preference for discrepancy and (b) whether this preference for discrepancy might be favoured over knowledge of reality. Possibilities for exploring these issues are presented below.

6.5.1. A preference for Discrepancy?

It would be useful to consider directly whether it is knowledge of the say-mean *discrepancy* that results in children overlooking the pragmatic adequacy of the utterance. Experiment 3.3 presented children with thought bubbles in pictorial and textual format. The discrepancy between what was said and what was meant was not so clear in the pictorial condition and so children might have been more likely to focus on the relationship between the utterance and the array. However, there was no difference between the textual and pictorial conditions in terms of children's tendency to evaluate unambiguous utterances negatively. Overall, though, the total number of negative responses was relatively low and the raw

data showed that the majority of these responses were confined to the textual condition. Consequently, there might have been a significant trend to say that the speaker had not said exactly which item he/she wanted in the textual rather than the pictorial condition, but the numbers were too small to generate a significant difference.

This issue could be explored directly by presenting children with thought content that is either discrepant or isomorphic with an ambiguous or unambiguous utterance. Children might only overlook the relationship between the utterance and array when the say-mean relationship is discrepant. This would suggest that it is specifically the discrepancy between thought and speech, rather than simply knowing the speaker's meaning that presents a problem for children (cf. Beal & Flavell, 1984).

6.5.2. Reality versus Discrepancy?

It would also be interesting to see how participants handle situations in which the quality of the reality-based relationship was systematically varied with source of discrepancy. For example, the paradigm utilised in Chapter 4 would be useful for investigating this issue, but with some modifications. One of the problems with the Not See condition as it stands is that the source of the discrepancy is confounded with the utterance-array relationship. Consequently, an experimental design is required that separates the source of the discrepancy from the reality-based relationship. One possible way to achieve this would be to combine the approaches of Chapters 3 and 4 in order to present children with a 2 x

2 design in which Type of discrepancy (consistent or discrepant) is crossed with Source of discrepancy (utterance-array or utterance-belief). Thought bubbles would reveal the speaker's meaning so that the utterance-belief relationship could be manipulated.

If children have a tendency to focus on a discrepancy over and above the reality-based relationship this will be evidenced by a trend to answer according to the discrepancy across conditions. It would be crucial to try to make all three elements of the triangle equally salient. Whether this could be achieved by presenting words, pictures or a combination of both requires careful thought because type of medium could have important implications for ease of comparison between, and salience of, the elements within the triangle. Overall, this kind of approach might help to provide new insights into which aspects of referential communication children struggle with.

6.6. Broader implications of Referential Communication

It seems that there are still many unanswered questions within referential communication research. Although the research focus on referential communication declined somewhat during the late 1980's and 90's, the importance of referential communication should not be underestimated or forgotten. Lloyd (e.g. 1992, 1994; Lloyd, Boada & Forns, 1992) has continued to promote the importance of referential communication, suggesting that it should form the basis of a standardised test for school children. Given that the skills involved in referential communication – sending and receiving clear and adequate

verbal messages - are especially relevant in a school environment (see also Robinson & Whittaker, 1986) developing such a test seems like a good way to augment assessments of educational achievement.

The importance of retaining a focus on this paradigm becomes especially important in the light of evidence that suggests children's understanding of referential communication is by no means fully developed by the end of the primary school years (Lloyd, Camaioni & Ercolani, 1995), or even by the age of thirteen years (Anderson, Clark & Mullin, 1994). As Lloyd et al. point out, the majority of referential communication research typically restricts participant groups to children aged around 6-7 years. However, it seems that there are developments in understanding that occur beyond this time.

Consequently, there is a need for continued basic research in this area because it may be that too much has been taken for granted. In particular, it is usually assumed that children's understanding of the adequacy of unambiguous utterances is exemplary. Some of the results presented in this thesis suggest that this might not be the case. This type of approach can usefully inform, and benefit from, more ambitious, large-scale projects such as the standardised test of referential communication proposed by Lloyd et al. Continued advocacy of the importance of referential communication is crucial, in addition to the stimulation of new research in this area. These approaches will hopefully provide a more representative idea of what children understand about communicating referentially.

References

Ackerman, B.P. (1993). Children's understanding of the speaker's meaning in referential communication. Journal of Experimental Child Psychology, 55, 56-86.

Ackerman, B.P. (1981). Performative bias in children's interpretations of ambiguous referential communications. Child Development, 52, 1224-1230.

Ackerman, B.P. (1979). Children's understanding of definite descriptions: Pragmatic inferences to the speaker's intent. Journal of Experimental Child Psychology, 28, 1-15.

Anderson, A.H., Clark, A. & Mullin, J. (1994). Interactive communication between children: learning how to make language work in dialogue. Journal of Child Language, 21, 439-463.

Bargh, J.A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J.S. Uleman & J.A. Bargh (Eds.), Unintended thought. New York: Guildford Press.

Baron-Cohen, S., Leslie, A.M. & Frith, U. (1985). Does the Autistic Child have a 'theory of mind'? Cognition, 21, 37-46.

Baron-Cohen, S. (1989). Are Autistic children 'Behaviourists'? An examination of their Mental-Physical and Appearance-Reality distinctions. Journal of Autism and Developmental Disorders, 19, 579-600

Beal, C.R. (1988). Children's knowledge about representations of intended meaning. In J.W. Astington, P.L. Harris, & D.R. Olson (Eds.), Developing theories of mind. Cambridge: Cambridge University Press.

Beal, C.R. & Belgrad, S.L. (1990). The development of message evaluation skills in young children. Child Development, 61, 705-712.

Beal, C.R. & Flavell, J.H. (1984). Development of the ability to distinguish communicative intention and literal message meaning. Child Development, 55, 920-928.

Bonitatibus, G. (1988a). What is said and what is meant in referential communication. In J.W. Astington, P.L. Harris, & D.R. Olson (Eds.), Developing theories of mind. Cambridge: Cambridge University Press.

Bonitatibus, G. (1988b). Comprehension monitoring and the apprehension of literal meaning. Child Development, 59, 60-70.

Bonitatibus, G. & Flavell, J.H. (1985). Effect of presenting a message in written form on young children's ability to evaluate its communication adequacy. Developmental Psychology, 21, 455-461.

Braine, M.D.S. & Rumain, B. (1983). Logical Reasoning. In J.H. Flavell & E.M. Markman (Eds.), Handbook of Child Psychology: Cognitive Development. (3). New York: John Wiley.

Camaioni, L., Ercolani, A.P., & Lloyd, P. (1998). The development of referential communication: Learning to speak and learning to process verbal information are not the same thing. Cahiers de Psychologie Cognitive, 17, 3-30.

Charman, T. & Baron-Cohen, S. (1992). Understanding drawings and beliefs: A further test of the metarepresentation theory of autism. Journal of Child Psychology and Psychiatry, 33, 1105-1112.

Dunn, L., Dunn, L., Whetton, C. & Pintilie, C. (1982) British Picture Vocabulary Scale. Windsor: NFER-Nelson.

Eisenmajer, R. & Prior, M. (1991). Cognitive linguistic correlates of 'theory of mind' ability in autistic children. British Journal of Developmental Psychology, 9, 351-364.

Flavell, J.H. (1977). Cognitive Development. Englewood Cliffs, N.J.:Prentice-Hall.

Flavell, J.H., Green, F.L. & Flavell, E.R. (1995). Young children's knowledge about thinking. Monographs of the Society for Research in Child Development, 60 (1).

Flavell, J.H., Speer, J.R., Green, F.L. & August, D.L. (1981). The development of comprehension monitoring and knowledge about communication. Monographs of the Society for Research in Child Development, 46, (5).

Freeman, N.H. (1994). Associations and dissociations in theories of mind. In C. Lewis & P. Mitchell (1994). (Eds.), Children's early understanding of mind: Origins and development. Hove: Lawrence Erlbaum.

Freeman, N.H. & Lacohee, H. (1995). Making explicit 3-year-olds' implicit competence with their own false beliefs. Cognition, 56, 31-60.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. Journal of Experimental Psychology: General, 108, 316-355.

Frith, U. (1989) Autism: Explaining the Enigma. Oxford: Blackwell.

Frith, U., Morton, J. & Leslie, A.M. (1991). The cognitive basis of a biological disorder: Autism. Trends in Neuroscience, 14, 433-438.

Gilbert, D.T. (1991). How mental systems believe. American Psychologist, 46, 107-119.

Gopnik, A. & Astington, J.W. (1988). The development of children's understanding of representational change. In J.W. Astington, P.L. Harris, & D.R. Olson (Eds.), Developing theories of mind. Cambridge: Cambridge University Press.

Grant, C.M., Riggs, K.J. & Boucher, J. (in press). Counterfactual and mental state reasoning in children with autism. Developmental Psychology.

Hadwin, J., Baron-Cohen, S., Howlin, P. & Hill, K. (1996). Can we teach children with autism to understand emotions, belief, or pretence? Development and Psychopathology, 8, 345-365.

Hansen, C.H. & Hansen, R.D. (1988). Finding the face in the crowd: An anger superiority effect. Journal of Personality and Social Psychology, 54, 917-924.

Happe, F.G.E. (1994). An advanced test of Theory of Mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. Journal of Autism and Developmental Disorders, 24, 129-154.

Happe, F.G.E. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. Child Development, 66, 843-855.

Hedelin, L. & Hjelmquist, E. (1998). Preschool children's mastery of the form/content distinction in communicative tasks. Journal of Psycholinguistic Research, 27, 421-452.

Holroyd, S. & Baron-Cohen, S. (1993). Brief report: How far can people with autism go in developing a theory of mind? Journal of Autism and Developmental Disorders, 23, 379-385.

Ironsmith, M. & Whitehurst, G.J. (1978). The development of listener abilities in communication: How children deal with ambiguous information. Child Development, 49, 348-352.

Kahneman, D. & Tversky, A.(1982). The simulation heuristic. Chapter in Kahneman, D., Slovic, P. & Tversky, A. (Eds.), Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press.

Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80, 237-251.

Keysar, B. (1994). The illusory transparency of intention: Linguistic perspective taking in text. Cognitive Psychology, 26, 165-208.

Leekam, S.R. & Perner, J. (1991). Does the autistic child have a metarepresentational deficit? Cognition, 40, 203-218.

Leslie, A.M. (1994). Pretending and believing: issues in the theory of ToMM. Cognition, 50, 211-238.

Leslie, A.M. (1987). Pretense and representation: The origins of "theory of mind". Psychological Review, 94, 412-426.

Leslie, A.M. & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. Cognition, 43, 225-251.

Lewis, C. & Mitchell, P. (1994). (Eds.) Children's early understanding of mind: Origins and development. Hove: Lawrence Earlbaum.

Lloyd, P. (1994). Referential communication: Assessment and intervention. Topics in Language Disorders, 14, 55-69.

Lloyd, P. (1992). The role of clarification requests in children's communication of route directions by telephone. Discourse Processes, 15, 357-374.

Lloyd, P. (1991). Strategies used to communicate route directions by telephone: a comparison of the performance of 7-year-olds, 10-year-olds and adults. Journal of Child Language, 18, 171-189.

Lloyd, P., Boada, H. & Forns, M. (1992). New directions in referential communication research. British Journal of Developmental Psychology, 10, 385-403.

Lloyd, P., Camaioni, L. & Ercolani, P. (1995). Assessing referential communication skills in the primary school years: A comparative study. British Journal of Developmental Psychology, 13, 13-29.

Markman, E.M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. Child Development, 50, 643-655.

Markman, E.M. (1977). Realizing that you don't understand: a preliminary investigation. Child Development, 48, 986-992.

McGregor, E., Whiten, A. & Blackburn, P. (1998a). Teaching theory of mind by highlighting intention and illustrating thoughts: A comparison of their effectiveness with three-year-olds and autistic subjects. British Journal of Developmental Psychology, 16, 281-300.

McGregor, E., Whiten, A. & Blackburn, P. (1998b). Transfer of the "picture-in-the-head" analogy to natural contexts to aid false belief understanding in autism. Autism, 2, 367-387.

Mitchell, P. (1999). The role of language and communication in children's ability to reason about the mind. The Annual Report of Educational Psychology in Japan, 38, 188-194.

Mitchell, P. (1997). Introduction to theory of mind: Children, autism and apes. London: Arnold.

Mitchell, P. (1996). Acquiring a conception of mind: A review of psychological research and theory. Hove: Psychology Press.

Mitchell, P. (1994). Realism and early conception of mind: A synthesis of phylogenetic and ontogenetic issues. In C. Lewis & P. Mitchell (1994). (Eds.), Children's early understanding of mind: Origins and development. Hove: Lawrence Erlbaum.

Mitchell, P. (1992). The psychology of childhood. London: The Falmer Press.

Mitchell, P. & Isaacs, J.E. (1994). Understanding of verbal representation in children with autism: The case of referential opacity. British Journal of Developmental Psychology, 12, 439-454.

Mitchell, P. & Lacohee, H. (1991). Children's early understanding of false belief. Cognition, 39, 107-127.

Mitchell, P., Munno, A & Russell, J. (1991). Children's understanding of the communicative value of discrepant verbal messages. Cognitive Development, 6, 279-299.

Mitchell, P. & Robinson, E.J. (1994). Discrepant utterances resulting from a false belief: Children's evaluations. Child Development, 65, 1214-1227.

Mitchell, P., Robinson, E.J., Isaacs, J.E. & Nye, R.M. (1996). Contamination in reasoning about false beliefs: an instance of realist bias in adults but not children. Cognition, 59, 1-21.

Mitchell, P., Robinson, E.J. & Thompson, D.E. (1999). Children's understanding that utterances emanate from minds: using speaker belief to aid interpretation. Cognition, 72, 45-66.

Mitchell, P. & Russell, J. (1991). Children's judgments of whether slightly and grossly discrepant objects were intended by a speaker. British Journal of Developmental Psychology, 9, 271-279.

Mitchell, P. & Russell, J.(1989). Young children's understanding of the say-mean distinction in referential speech. Journal of Experimental Child Psychology, 47, 467-490.

Mitchell, P., Saltmarsh, R. & 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.

Nelson, K., Plesa, D., & Henseler, S. (1998). Children's theory of mind: an experiential interpretation. Human Development, 41, 7-29.

Olson, D.R. (1981). A conceptual revolution in the early school years: Learning to differentiate intended meaning from meaning in the text. Paper presented at the International Society for the Study of Behavioural Development. Cited in Bonitatibus, G. (1988b). Comprehension monitoring and the apprehension of literal meaning. Child Development, 59, 60-70.

Olson, D.R. (1970). Language and thought: Aspects of a cognitive theory of semantics. Psychological Review, 77, 257-273.

Olson, D.R. & Hildyard, A. (1983). Writing and literal meaning. In Martlew, M. (Ed.), The psychology of written language: Developmental and educational perspectives. New York: John Wiley.

Parsons, S. (1996). What children with autism understand about thoughts and thought bubbles. Unpublished B.Sc. (Hons.) research project. University of Birmingham.

Perner, J. (1991). Understanding the Representational Mind. London: MIT Press.

Perner, J., Leekam, S., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. British Journal of Developmental Psychology, 5, 125-137.

Reid, L. (1996). The effects of focusing children's attention on the literal meaning of the message. Merrill-Palmer Quarterly, 42, 295-318.

Robinson, E.J. (1994). What people say, what they think, and what is really the case: children's understanding of utterances as sources of knowledge. In C. Lewis & P. Mitchell (1994). (Eds.), Children's early understanding of mind: Origins and development. Hove: Lawrence Erlbaum.

Robinson, E.J. (1981). The child's understanding of inadequate messages and communication failure: A problem of ignorance or egocentrism? In W.P. Dickson (Ed.), Children's oral communication skills. New York: Academic Press.

Robinson, E.J. & Mitchell, P. (1994). Young children's false belief reasoning: Interpretation of messages is no easier than the classic task. Developmental Psychology, 30, 67-72.

Robinson, E.J. & Mitchell, P. (1992). Children's interpretations of messages from a speaker with a false belief. Child Development, 63, 639-652.

Robinson, E.J. & Mitchell, P. (1990). Children's failure to make judgments of undecidability when they are ignorant. International Journal of Behavioral Development, 13, 467-488.

Robinson, E.J. & Whittaker, S.J. (1987). Children's conceptions of relations between messages, meanings and reality. British Journal of Developmental Psychology, 5, 81-90.

Robinson, E.J. & Whittaker, S.J. (1986). Children's conceptions of meaning-message relationships. Cognition, 22, 41-60.

Robinson, E.J., Goelman, H. & Olson, D.R. (1983). Children's understanding of the relation between expressions (what was said) and intentions (what was meant). British Journal of Developmental Psychology, 1, 75-86.

Robinson, E.J. & Robinson, W.P. (1983). Communication and metacommunication: Quality of children's instructions in relation to judgments about the adequacy of instructions and the locus of responsibility for communication failure. Journal of Experimental Child Psychology, 36, 305-320.

Robinson, E.J. & Robinson, W.P. (1977). Children's explanations of communication failure and the inadequacy of the misunderstood message. Developmental Psychology, 13, 156-161.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The "windows task" as a measure of strategic deception in preschoolers and autistic subjects. British Journal of Developmental Psychology, 9, 331-350.

Sa, W.C. & Stanovich, K.E. (1999). Why do adults "fail" some theory of mind reasoning tasks? An individual differences analysis. Unpublished manuscript, University of Toronto.

Sacks, O. (1995). An Anthropologist on Mars. Picador.

Saltmarsh, R. & Mitchell, P. (1999). Indelible evidence of false belief: confronting young children with video recordings of themselves. Psychologia, 42, 145-159.

Saltmarsh, R., Mitchell, P., & Robinson, E.J. (1995). Realism and children's early grasp of mental representation: Belief-based judgements in the state change task. Cognition, 57, 297-325.

Sodian, B. (1988) Children's attributions of knowledge to the listener in a referential communication task. Child Development, 59, 378-385.

Sodian, B. & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. Child Development, 58, 424-433.

Sonnenschein, S. (1984). Referential communication: Why do preschoolers confuse quantity with quality? In Feagans, L., Garvey, C., & Golinkoff, R. (Eds.), The origins and growth of communication. New Jersey: Ablex Publishing.

Sonnenschein, S. (1985). The development of referential communication skills: Some situations in which speakers give redundant messages. Journal of Psycholinguistic Research, 14, 489-508.

Sonnenschein, S. (1982) The effects of redundant communications on listeners: When more is less. Child Development, 53, 717-729.

Sonnenschein, S. & Whitehurst, G.J. (1982). The effects of redundant communications on the behaviour of listeners: Does a picture need a thousand words? Journal of Psycholinguistic Research, 11, 115-125.

Sparrevohn, R. & Howie, P.M. (1995). Theory of mind in children with autistic disorder: Evidence of developmental progression and the role of verbal ability. Journal of Child Psychology and Psychiatry, 36, 249-263.

Swettenham, J.G., Baron-Cohen, S., Gomez, J-C. & Walsh, S. (1996). What's inside someone's head? Conceiving of the mind as a camera helps children with autism acquire an alternative to a Theory of Mind. Cognitive Neuropsychiatry, 1, 73-88.

Taylor, L.M. & Mitchell, P. (1997). Judgments of apparent shape contaminated by knowledge of reality: Viewing circles obliquely. British Journal of Psychology, 88, 653-670.

Taylor, S.E. (1982). The availability bias in social perception and interaction. Chapter in Kahneman, D., Slovic, P. & Tversky, A. (Eds.), Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press.

Tversky, A. & Kahneman, D. (1983). Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. Psychological Review, 90, 293-315.

Tversky, A. & Kahneman, D. (1982). Availability: A heuristic for judging frequency and probability. Chapter in Kahneman, D., Slovic, P. & Tversky, A. (Eds.), Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press.

Wellman, H.M. (1990) The Child's Theory of Mind. Cambridge, MA: MIT Press.

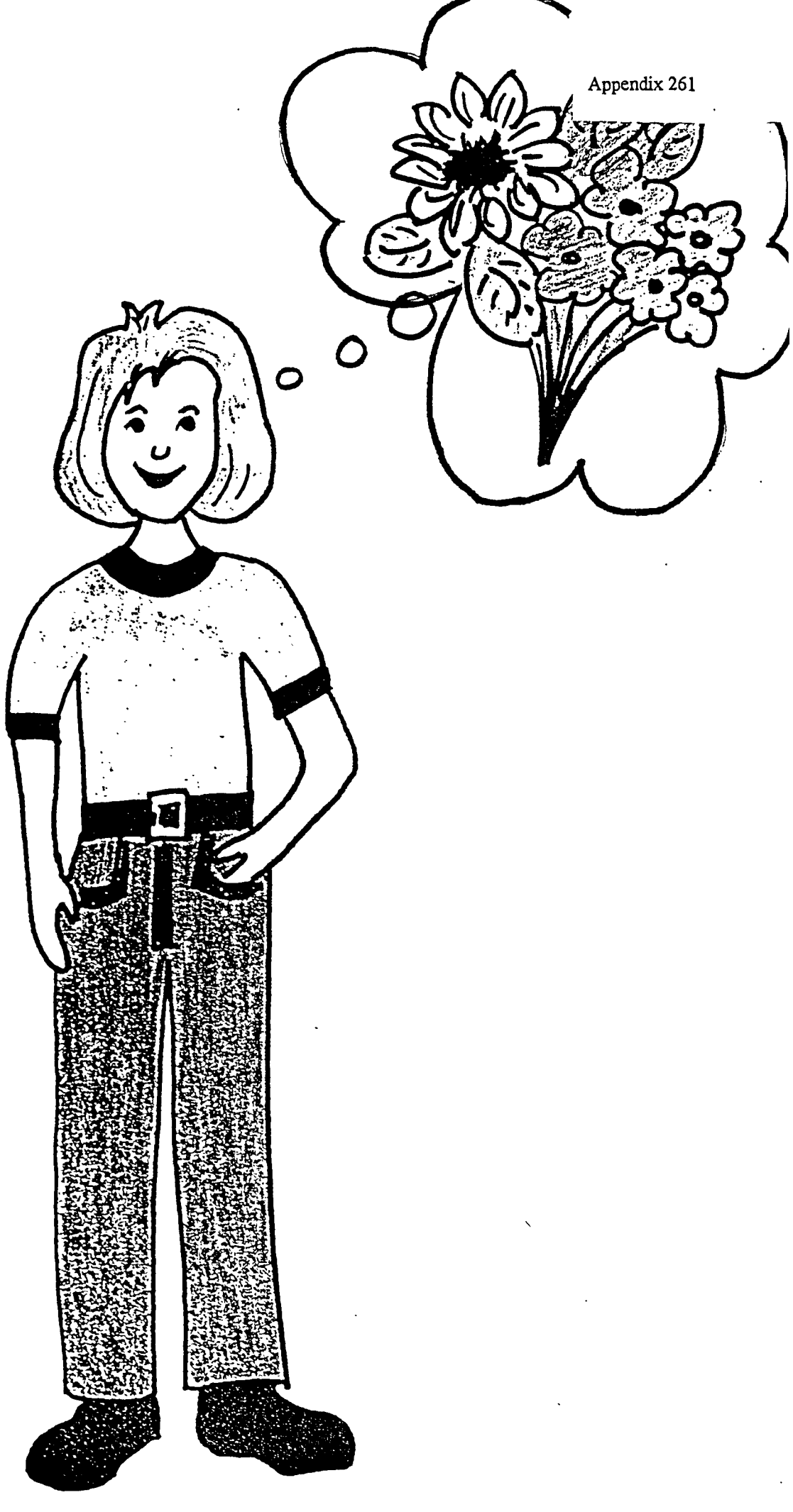
Wellman, H.M., Hollander, M. & Schult, C.A. (1996). Young children's understanding of thought bubbles and of thoughts. Child Development, 67, 768-788.

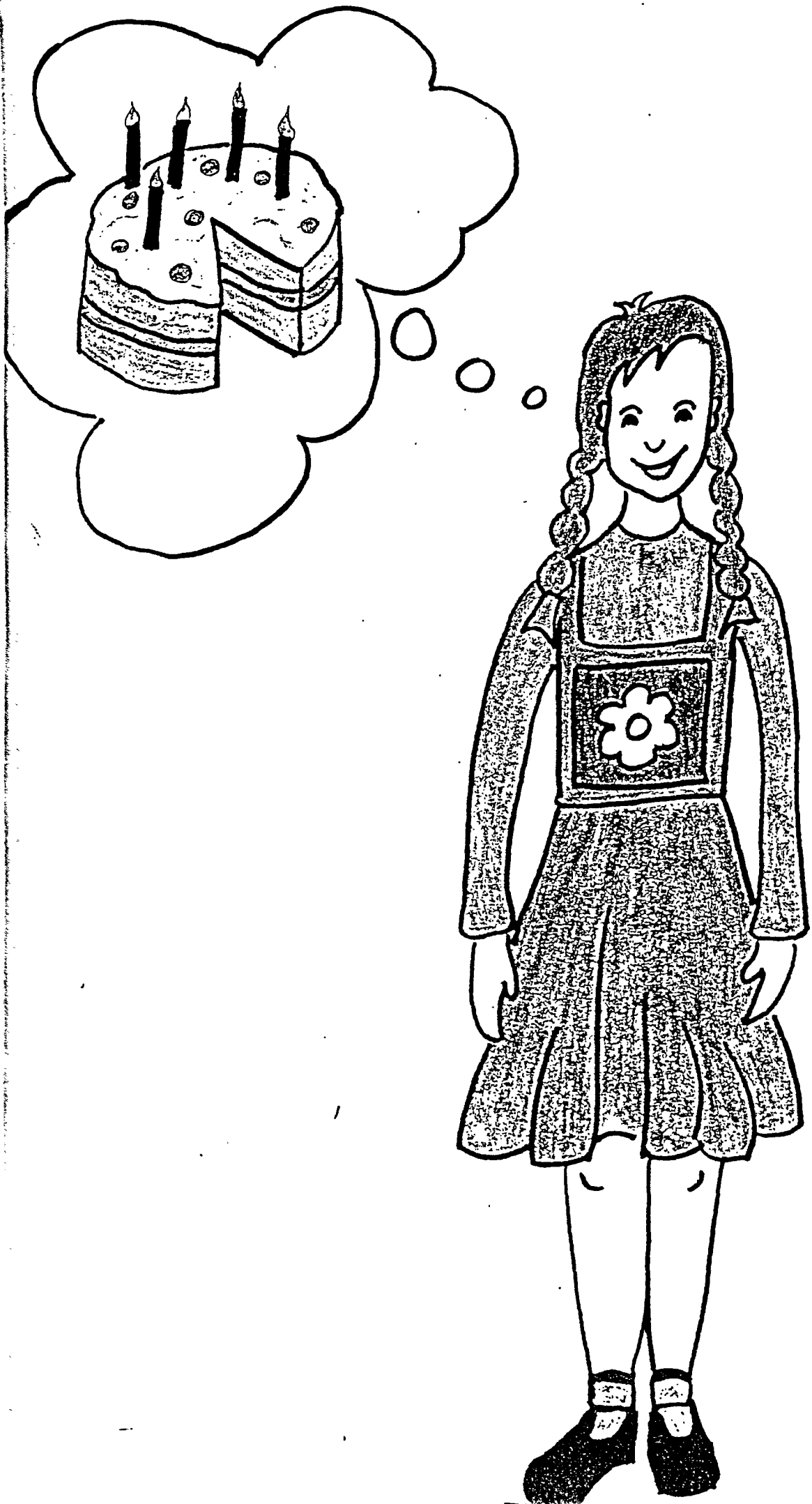
Whitehurst, G.J. (1976). The development of communication: Changes with age and modeling. Child Development, 47, 473-482.

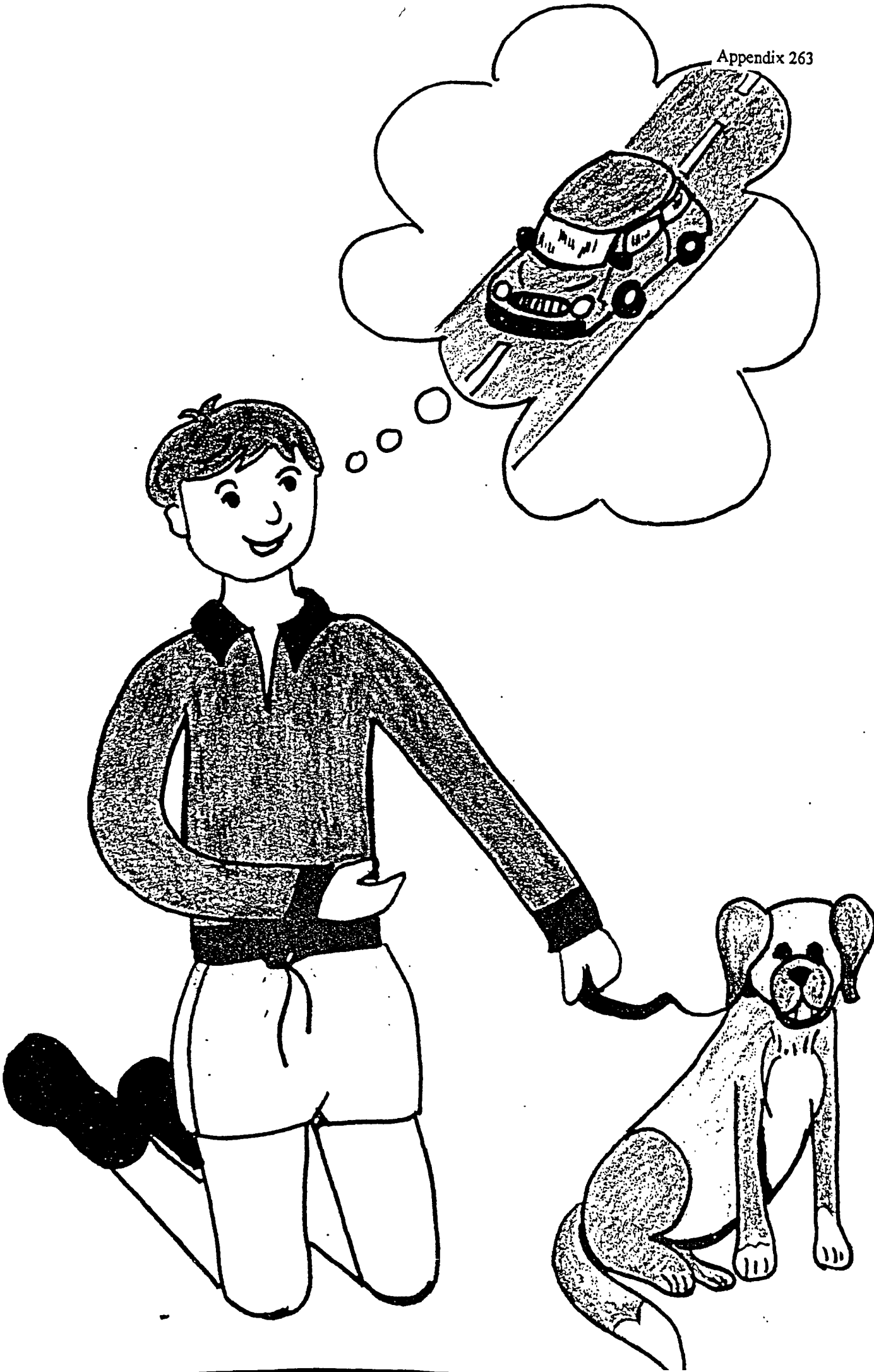
Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition, 13, 103-128.

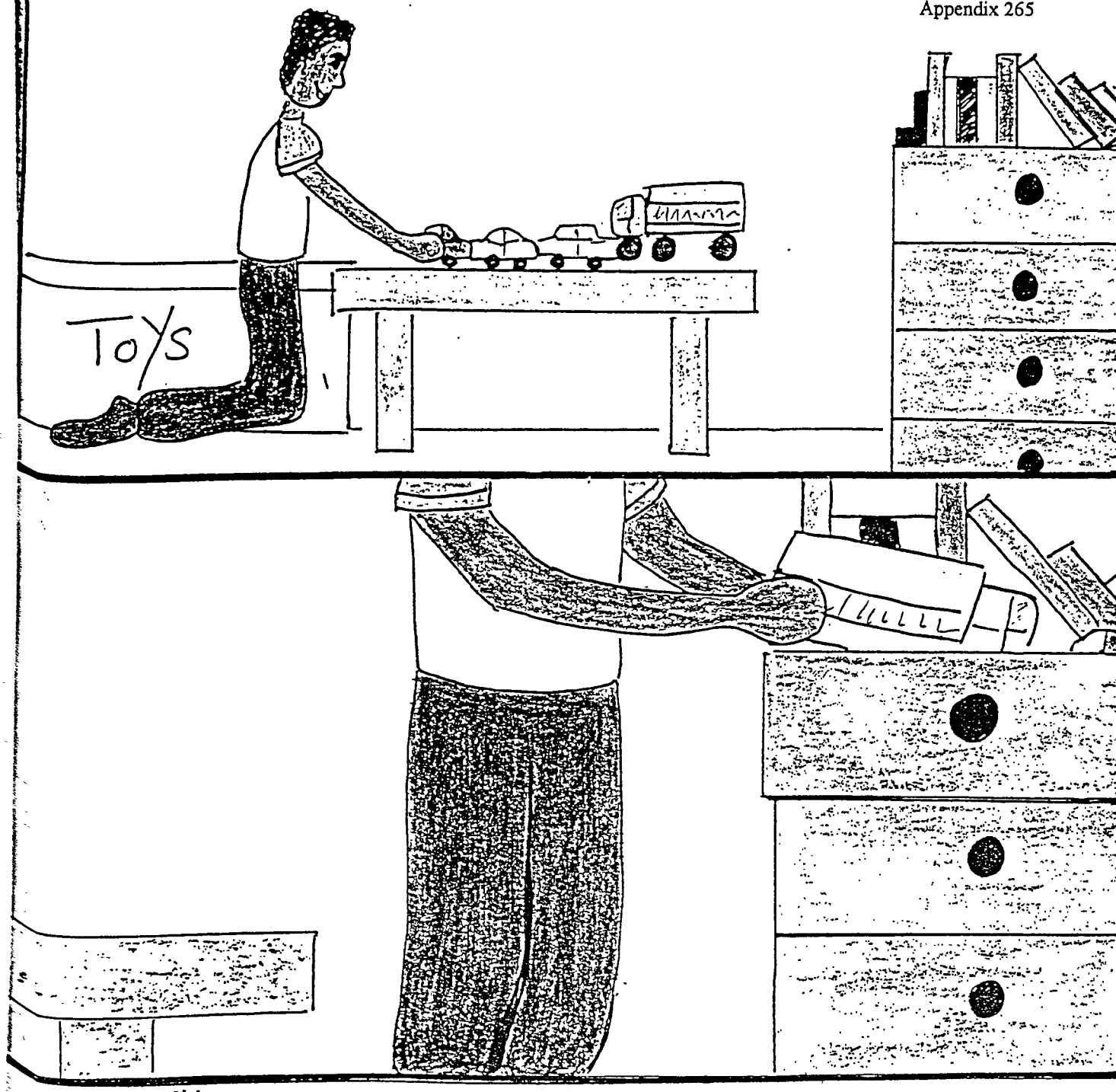
Yirmiya, N., Solomonica-Levi, D., Shulman, C. & Pilowsky, T. (1996). Theory of mind abilities in individuals with autism, Down Syndrome, and mental retardation of unknown etiology: The role of age and intelligence. Journal of Child Psychology and Psychiatry, 37, 1003-1014.

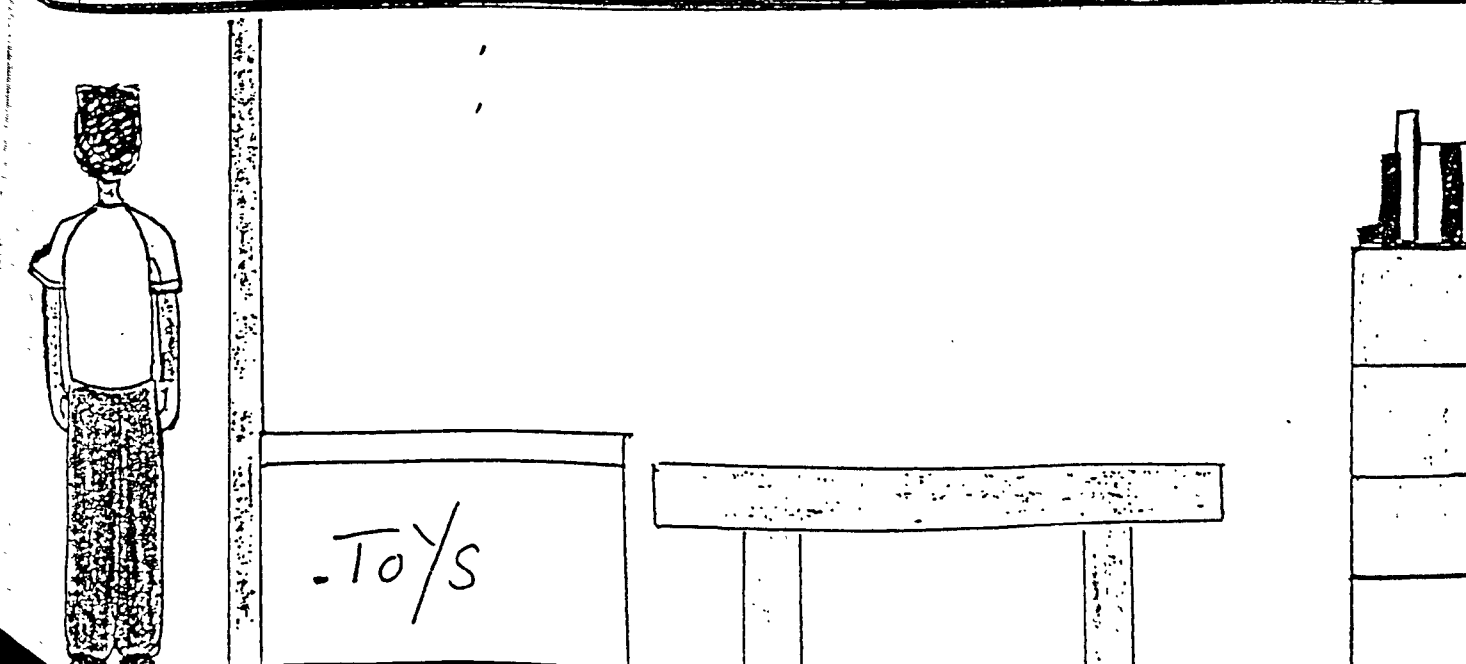
Zaitchik, D. (1990). When representations conflict with reality: The preschoolers' problem with false belief and 'false' photographs. Cognition, 35, 41-68.

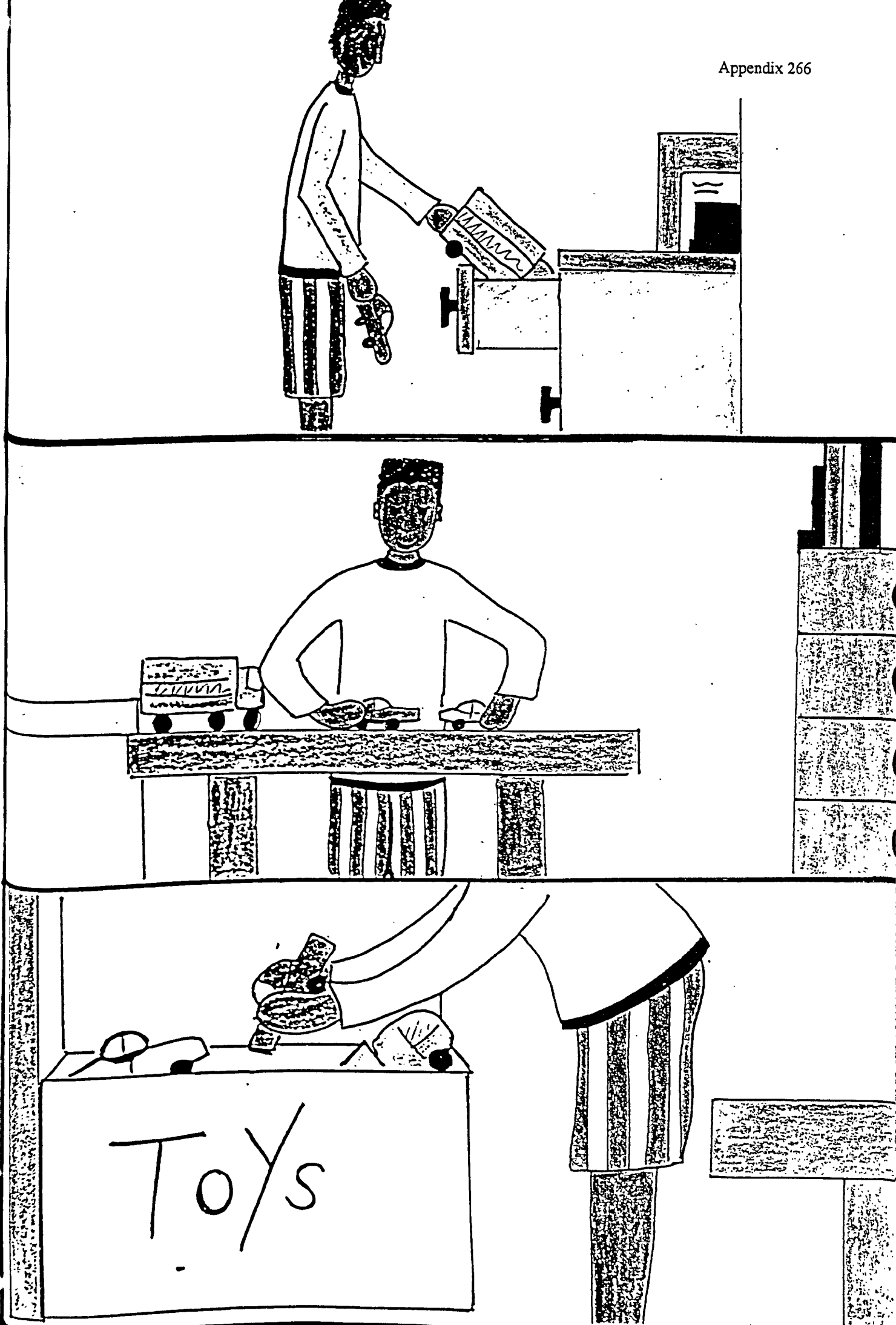


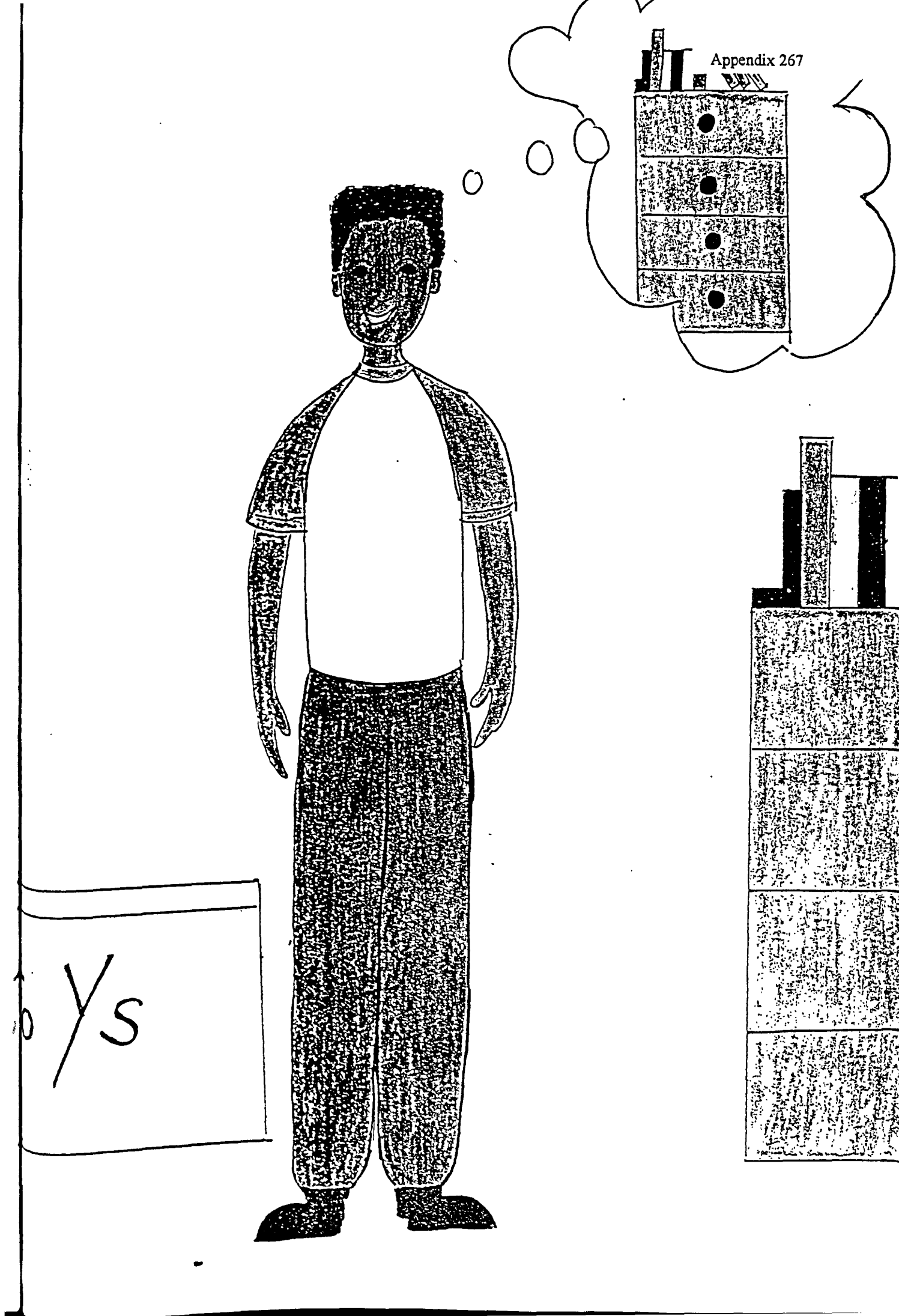


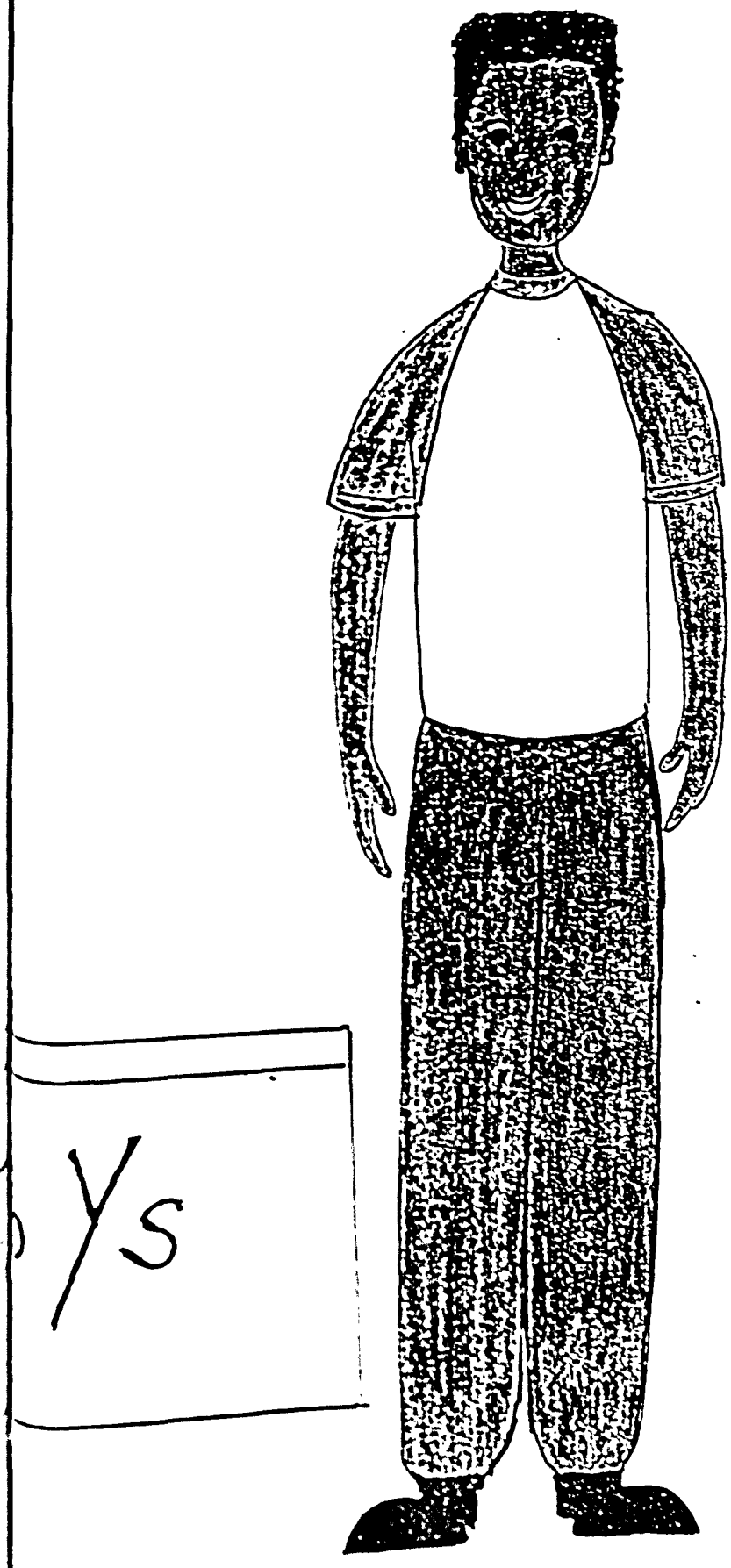


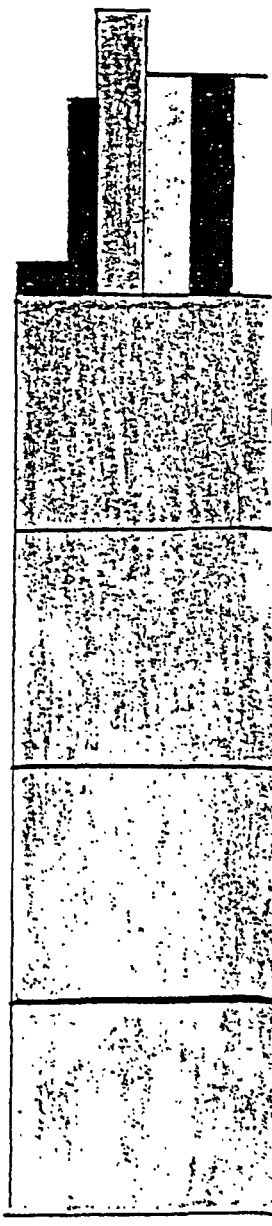

TOYS


TOYS


TOYS


Ys


Ys



Story scripts for Chapter 2

A. False belief: Change of location

(1) Paul and the chocolate.

"This is Paul. Paul has been shopping and he has bought a large bar of chocolate. He puts his bar of chocolate in the kitchen drawer and remembers very carefully where he put it because he will want to eat some later."

Prompt: Where did Paul put the chocolate?

" Now Paul goes outside to play football. Whilst he is outside, Paul's Mum comes into the kitchen looking for some chocolate for a cake. She finds Paul's bar in the kitchen drawer, uses some of it for her cake and then puts the chocolate in the fridge instead of back in the drawer."

Check: Did Paul see Mum do that? No, he didn't!

"Paul is playing outside so he couldn't see where Mum put the chocolate. Paul has come back into the kitchen."

(2) John and the cars.

"This is John. John is playing with his toy cars but he wants to watch TV in the other room. He packs his cars away into the drawer and then leaves the room."

Prompt: Where did John put his cars?

" Whilst John is watching the TV, his brother Mark comes into the playroom looking for the toy cars. He finds them in the drawer, plays with then and then puts the cars in the toy box instead of back in the drawer."

Check: Did John see Mark do that? No, he didn't!

" John was watching the TV so he didn't see Mark put the cars in the toy box. John has come back into the room now."

B. False belief: change of contents

(1) Susie and her swimming bag.

"This is Susie. She decides to go swimming. Susie packs her swimsuit into her bag. Susie has left her money upstairs, so she leaves the room to get it."

Prompt: What did Susie put in her bag?

" Whilst Susie is gone, her Mum comes into the room. She's doing some laundry and is looking for some clothes to wash. She takes Susie's swimsuit out of her bag and puts a jacket in the bag instead – thinking that Susie might need it later."

Check: Did Susie see Mum do that? No, she didn't!

"Susie was upstairs so she didn't see Mum put the jacket in the bag. Now Susie has come back downstairs."

(2) Claire and her shoes.

"This is Claire. Claire has been shopping to buy a new pair of shoes. Here they are in the red box. Claire puts the shoe box on the kitchen table and goes to play outside."

Prompt: What did Claire have in the red box?

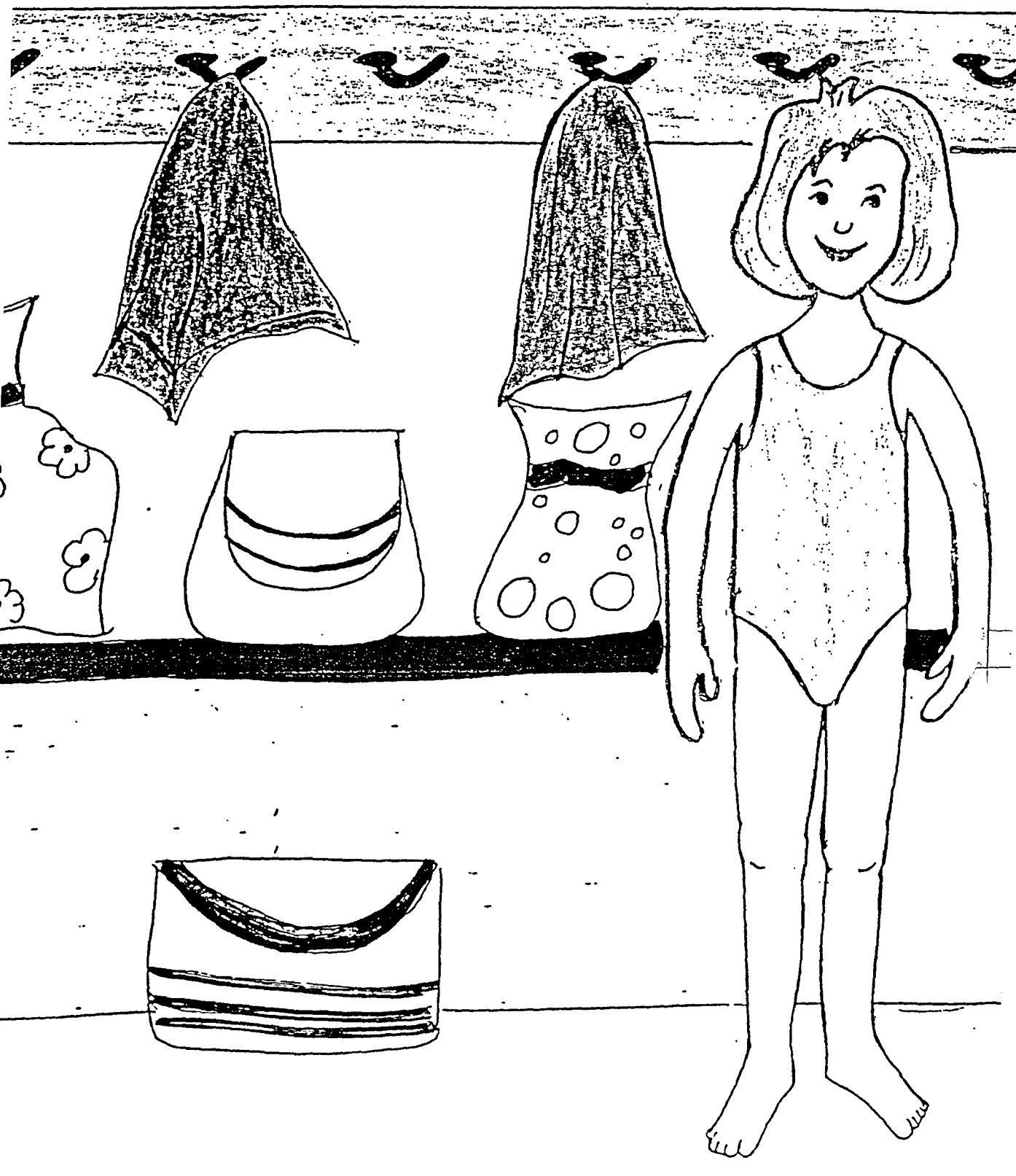
"Whilst she is outside, Claire's Mum comes into the kitchen. She's looking for a box to send a present to a friend. She's bought her friend a woolly jumper for her birthday. Mum takes Claire's shoes out of the box and puts them into a bag. Then she folds the jumper and puts it in the box. Mum puts the lid on the box and leaves the room."

Check: Did Claire see Mum do that? No, she didn't!

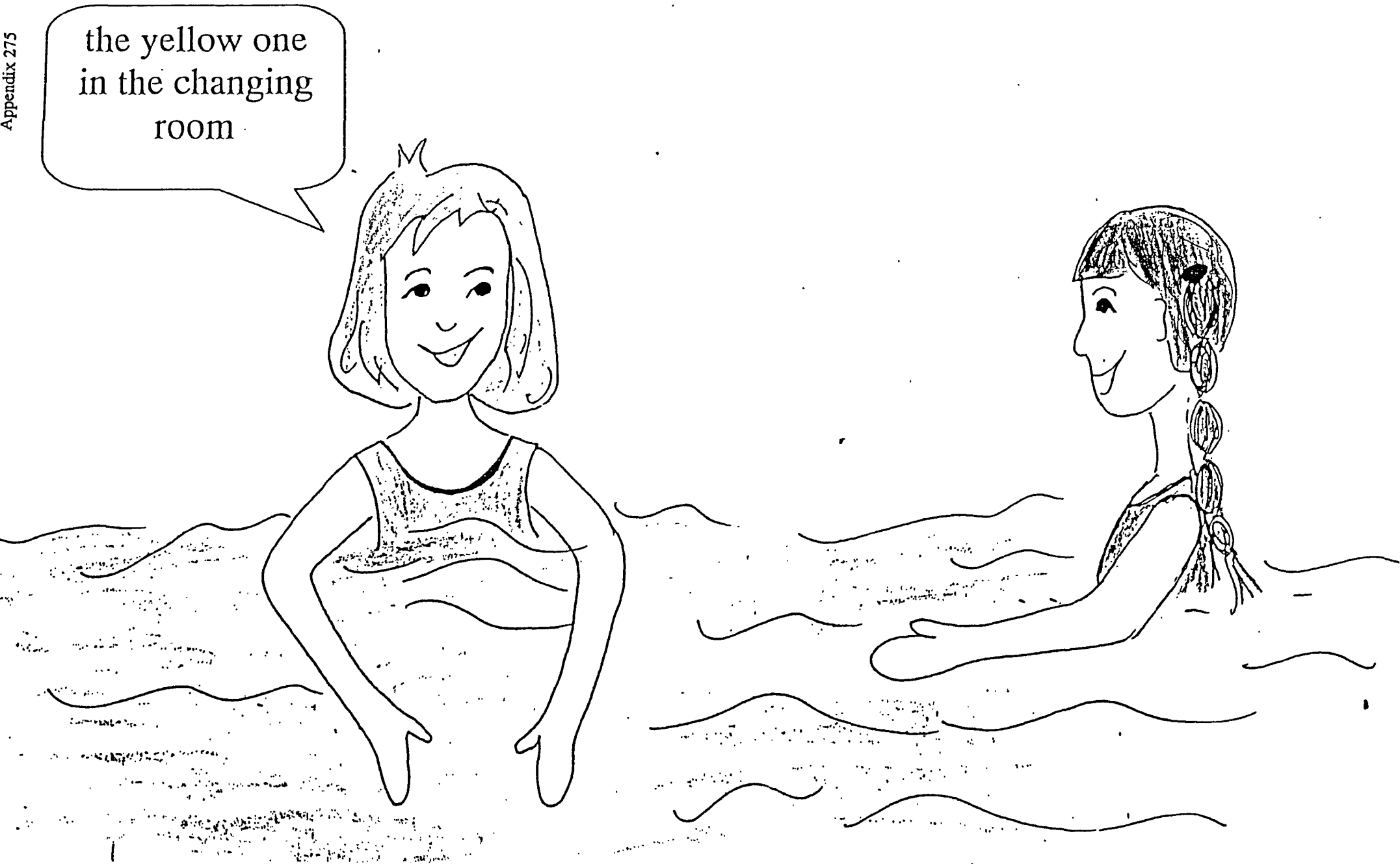
"Claire was playing outside so she couldn't see Mum put the jumper in the box. Now Claire's come back into the kitchen."

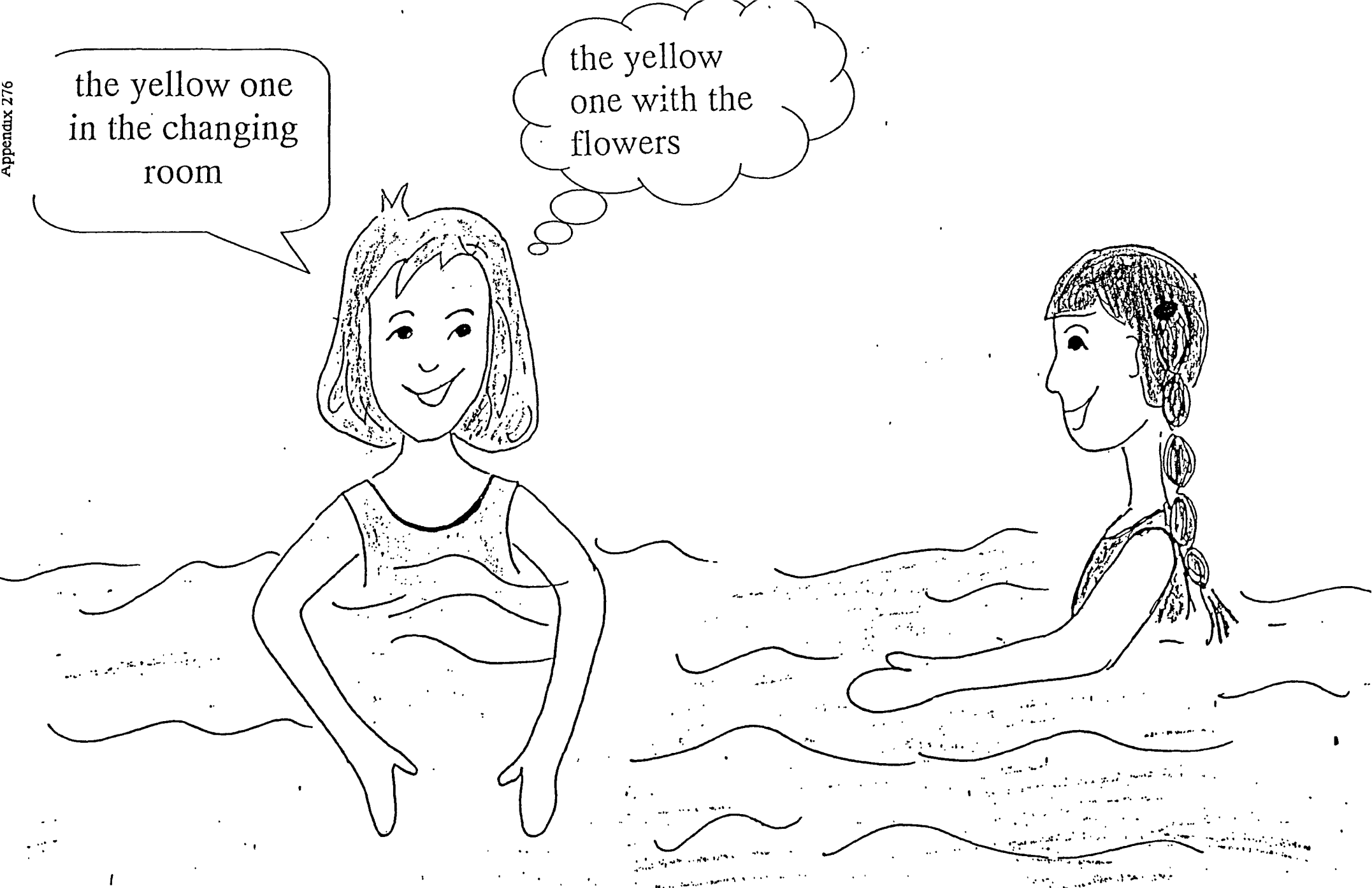

hello
drink

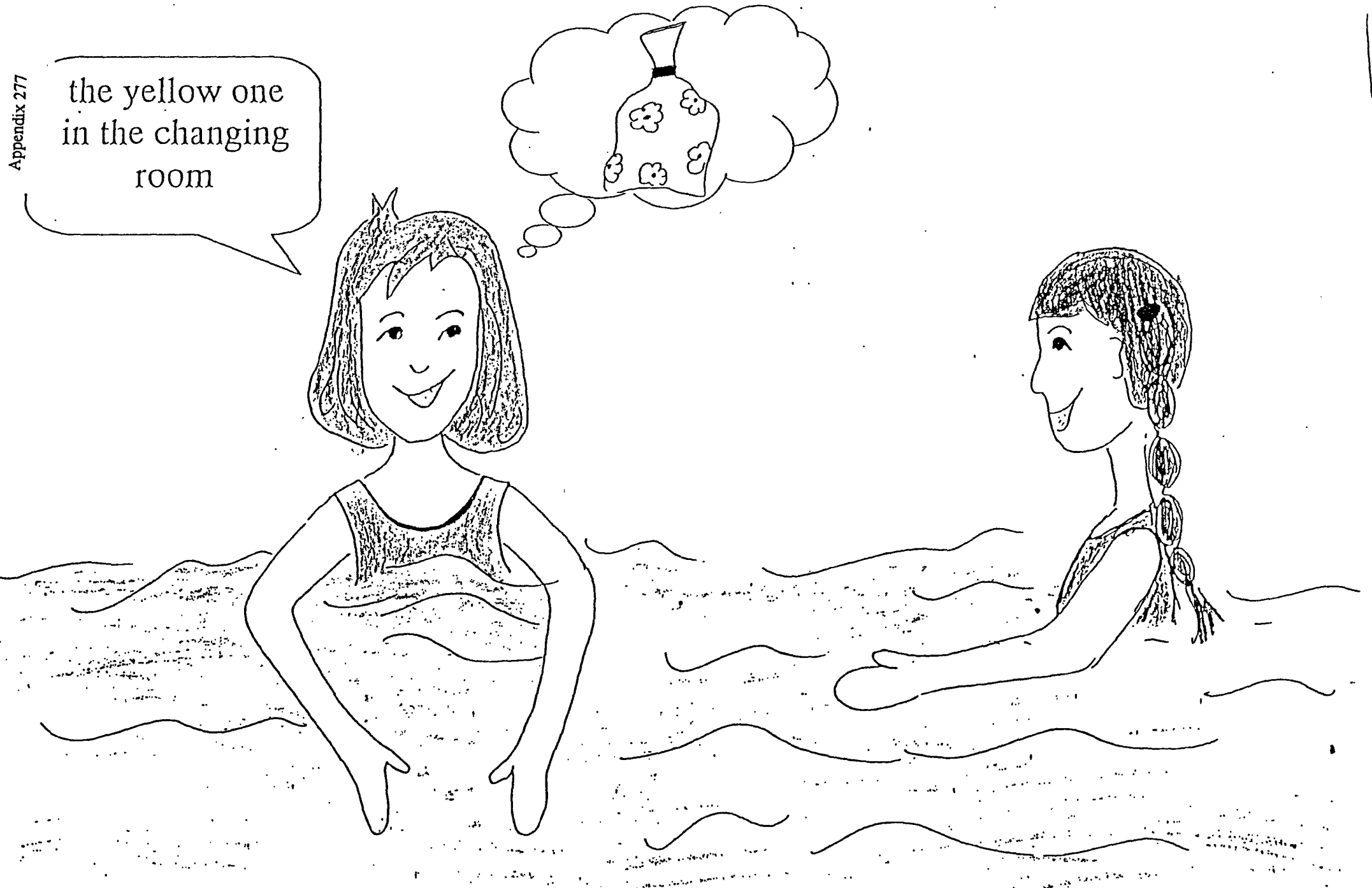

hello




the yellow one with the flowers


the yellow one in the changing room


the yellow one in the changing room
the yellow one with the flowers


the yellow one in the changing room

Story scripts for Chapter 3

Note that the narratives changed slightly between Experiments 3.1 and 3.2. The additional information included in the speech bubbles in Experiment 3.2 (and 3.3.) is displayed in italics after the original utterance included in Experiment 3.1.

(1) Sam goes to the shop to buy some sweets. He asks the man behind the counter to get one of the jars from the window. The man asks Sam which sweets he wants [Sam thinks "the orange ones with the stripes"] Sam says "the orange ones *in the jar*"

(2) Mrs. Edwards goes to the pet shop to buy a puppy. She tells the shop owner she would like one of the puppies in the front of the shop. He asks her which one she wants. [Mrs. Edwards thinks "the black one with brown ears"] Mrs. Edwards says "the black one *on the blanket*"

(3) Sophie has drawn a picture for the Parent's Evening. Later, she tells her Dad to make sure he looks at the picture of a horse. Dad asks Sophie which horse picture is hers. [Sophie thinks "the white one with the birds"] Sophie says "the white one *on the wall*"

(4) Rosie is getting ready to go swimming at the pool. In the pool, her friend Megan swims over to her and asks to borrow Rosie's swimming goggles. Rosie tells her they are in her bag in the changing rooms. Megan asks which bag. [Rosie thinks " the yellow one with the flowers"] Rosie says "the yellow one *in the changing rooms*"

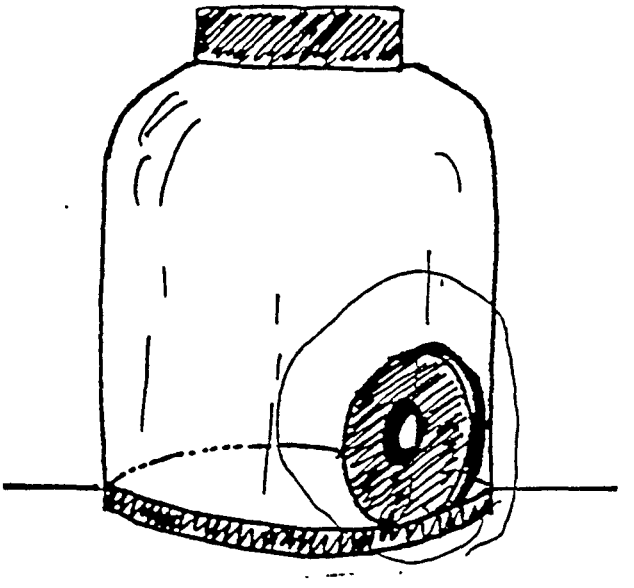
(5) Dad is getting ready to go outside gardening with Fiona. Later, he starts to get cold and asks Fiona to fetch him a jumper from his bedroom. Fiona asks Dad which jumper he wants.[Dad thinks "the green one with red sleeves"] Dad says "the green one *in the wardrobe*"

(6) Dave sees some new bikes in Marsh's shop window whilst he is out shopping. Later he tells his Mum about the bikes. She asks him which one he liked. [Dave thinks " the red one with the horn"] Dave says " the red one *in the window* "

(7) Claire sees some balls in her classroom before she goes out for break. At break, her friend Anna wants to play ball. Claire tells her there are some balls in her classroom. Anna asks her which ball she wants. [Claire thinks "the brown one with spots"] Claire says "the brown one *on the floor*"

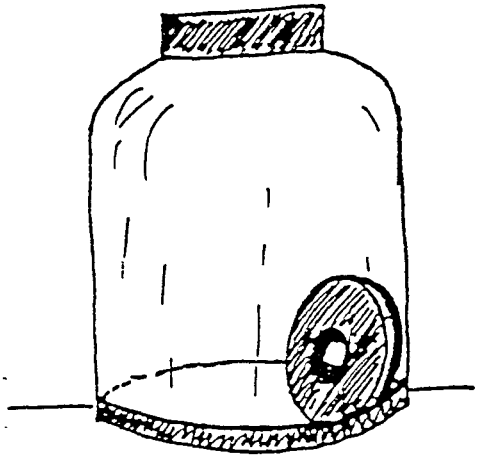
(8) At school, Mr. Riley hangs his coat up in the staff room. Later, Mr. Riley asks Nathan if he could fetch his coat. Nathan asks him which coat is his. [Mr.Riley thinks "the blue one with the hood"] Mr. Riley says "the blue one *on the peg*"

C/S/NS/t/gj



Adam comes home after school and is feeling hungry. He sees the glass biscuit jar on the shelf out of reach. There is only one biscuit left in the jar. It's a chocolate one which is Adam's favourite. Adam asks Dad if he can get a biscuit for him. Dad says "Yes, which biscuit would you like?" Adam says "I'd like the biscuit in the jar please".

Dad goes to get a biscuit for Adam.....



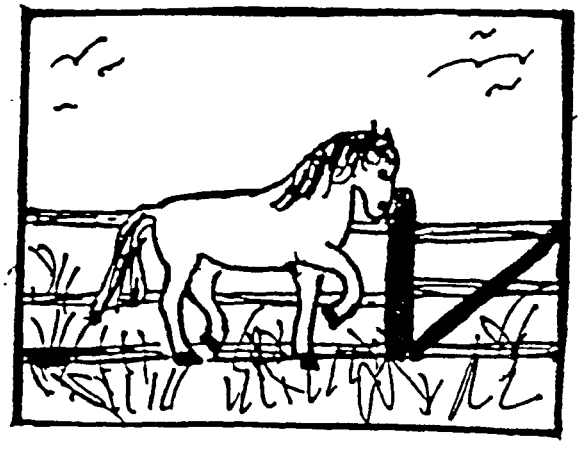
(1) Which biscuit will Dad think Adam wants?

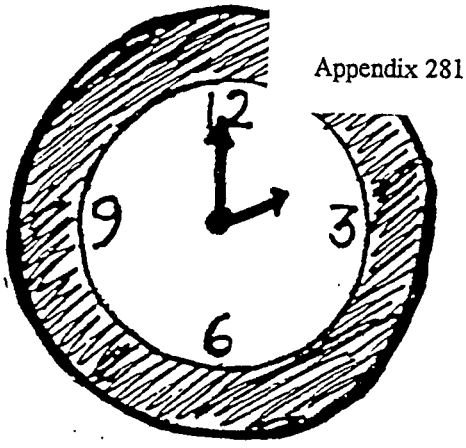
(Please circle one of the pictured biscuits or the '?' if you think Dad won't be able to tell)

(2) Did Adam do a good job of saying exactly which biscuit he wanted? Yes/[redacted]

(3) Why/why not? because theres only 1 biscuit.

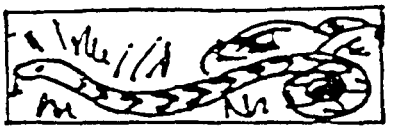
C/S/NS/t/gj

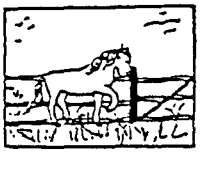


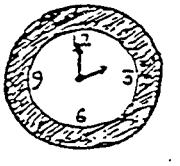


Sophie has drawn a picture for her school's open evening. It has a large white horse on it. She helps her teacher put the picture on the empty wall next to the classroom clock. The teacher then asks Sophie to help her put some more pictures around the clock. There are lots of pictures next to the clock now. Later at home, Dad asks Sophie which picture to look out for when he visits the school. Sophie says "My picture is the one by the clock."

Dad goes to the school to look at the pictures.....







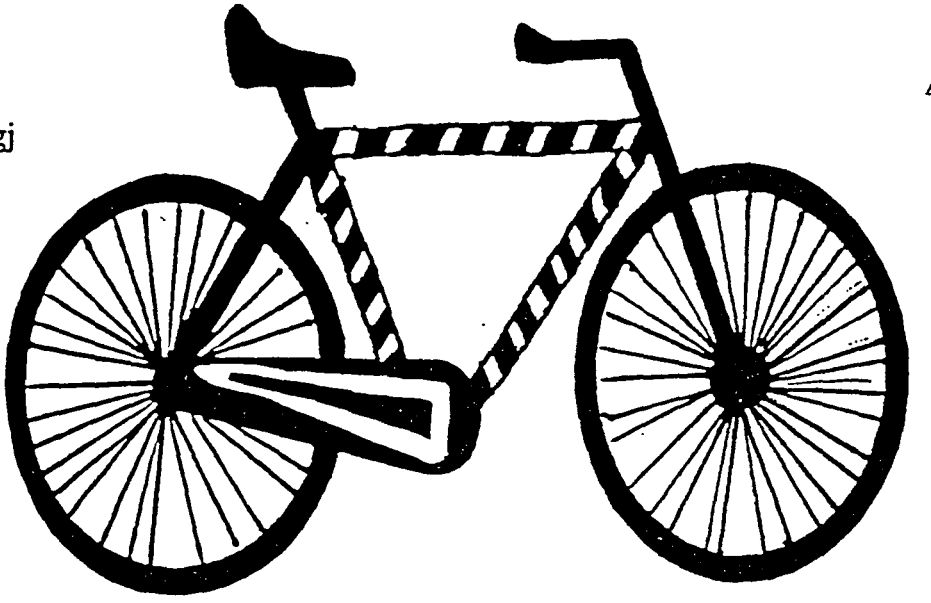
(1)Which picture will Dad think is Sophie's?

(Please circle one of the pictures or the '?' if you think Dad won't be able to tell)

(2)Did Sophie do a good job of saying exactly which picture was hers? Yes/No

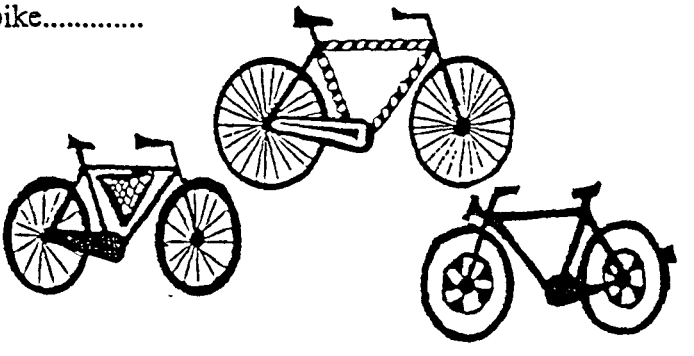
(3)Why/Why not? because there is more than 1 picture by the clock.

C/S/NS/t/gj



Dave is shopping in town with his friends when they walk past the toy shop. In the window is a new bike, which Dave stops to have a look at. Whilst Dave walks home, the shopkeeper brings out more bikes to display in the window. There are lots of bikes on display now. Later at home, Dave's Mum asks him if he has seen anything he would like to have for his birthday. Dave says "Yes, there was a bike in the toy shop window today. I'd like to have that."

The next day, Mum goes to buy the bike.....



(1) Which bike will Mum think Dave wants?

(Please circle one of the pictured bikes or the '?' if you think Mum won't be able to tell)

(2) Did Dave do a good job of saying exactly which bike he wants? ~~Yes~~/No

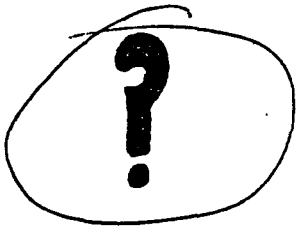
(3) Why/why not? because there are lots of bikes

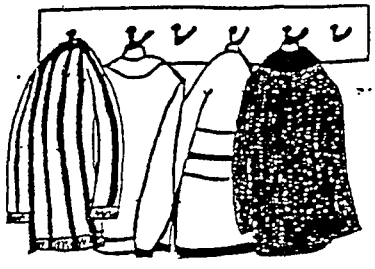
NS/C/S/gj/t

(1) Did Mr. Riley do a good job of saying exactly which coat was his? Yes/(No)

(2) Why/Why not? There a coat hanger with lots of hooks above the bookcase with more coats on it.

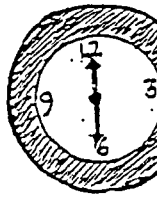
(3) Which coat will Nathan think Mr. Riley wants? (please circle one of the pictured coats or the '?' if you think Nathan won't be able to tell)



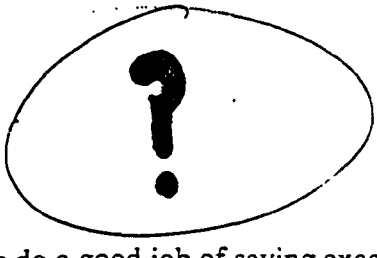


(1) Did Sophie do a good job of saying exactly which picture was hers? Yes/(No)

(2) Why/why not? should of described what the picture looked like incase other pictures are hung around the clocks.



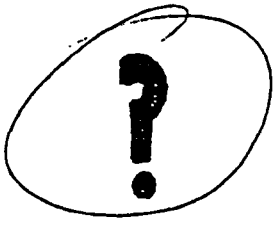
(3) Which picture will Dad think is Sophie's?

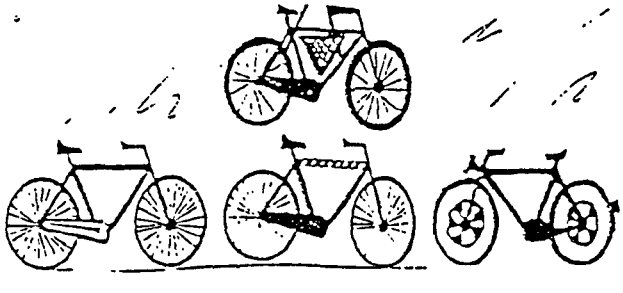


(1) Did Dave do a good job of saying exactly which bike he wants? Yes/(No)

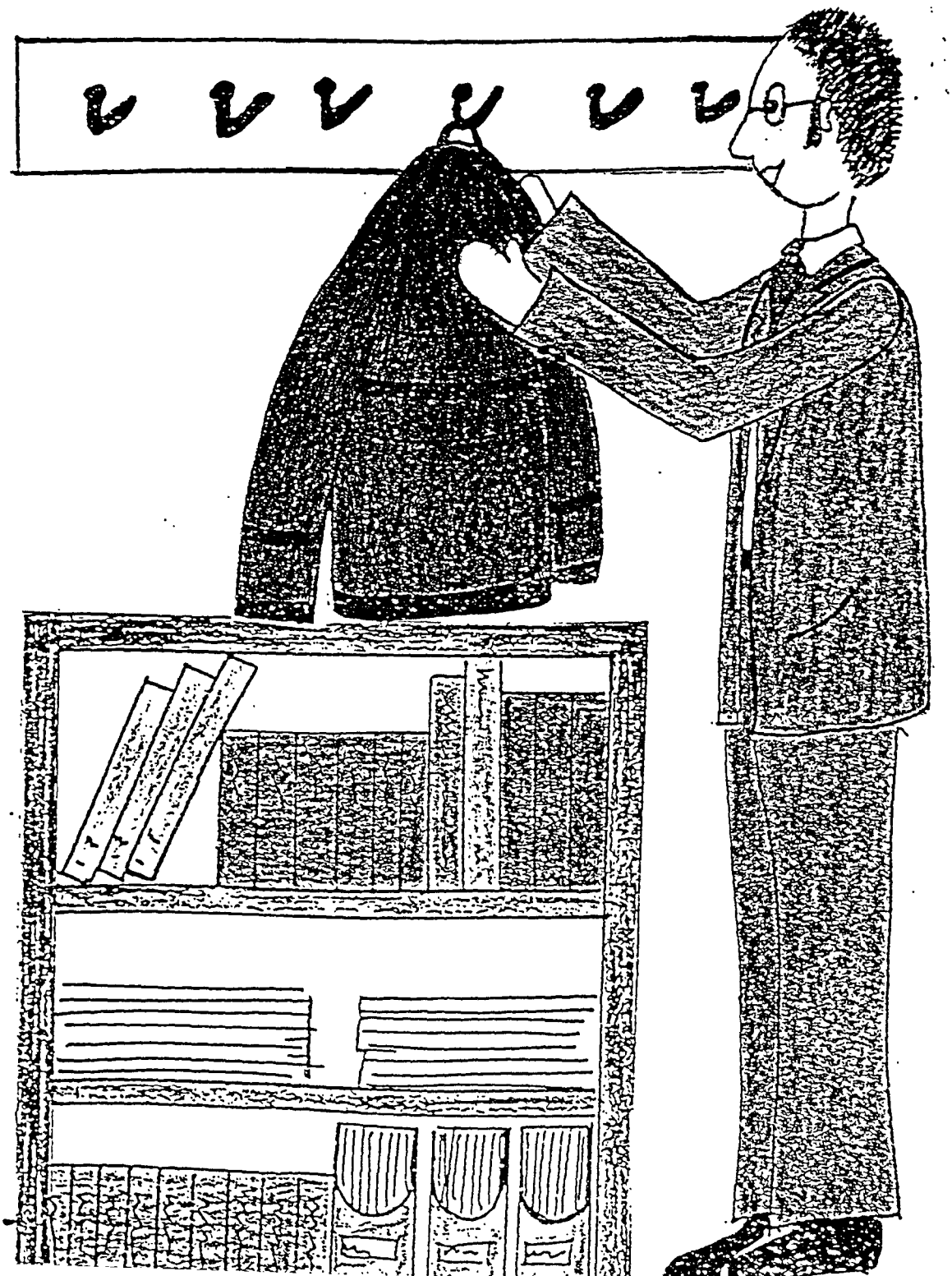
(2) Why/why not? Did not tell his mom which bike he wanted

(3) Which bike will Mum think Dave wants?

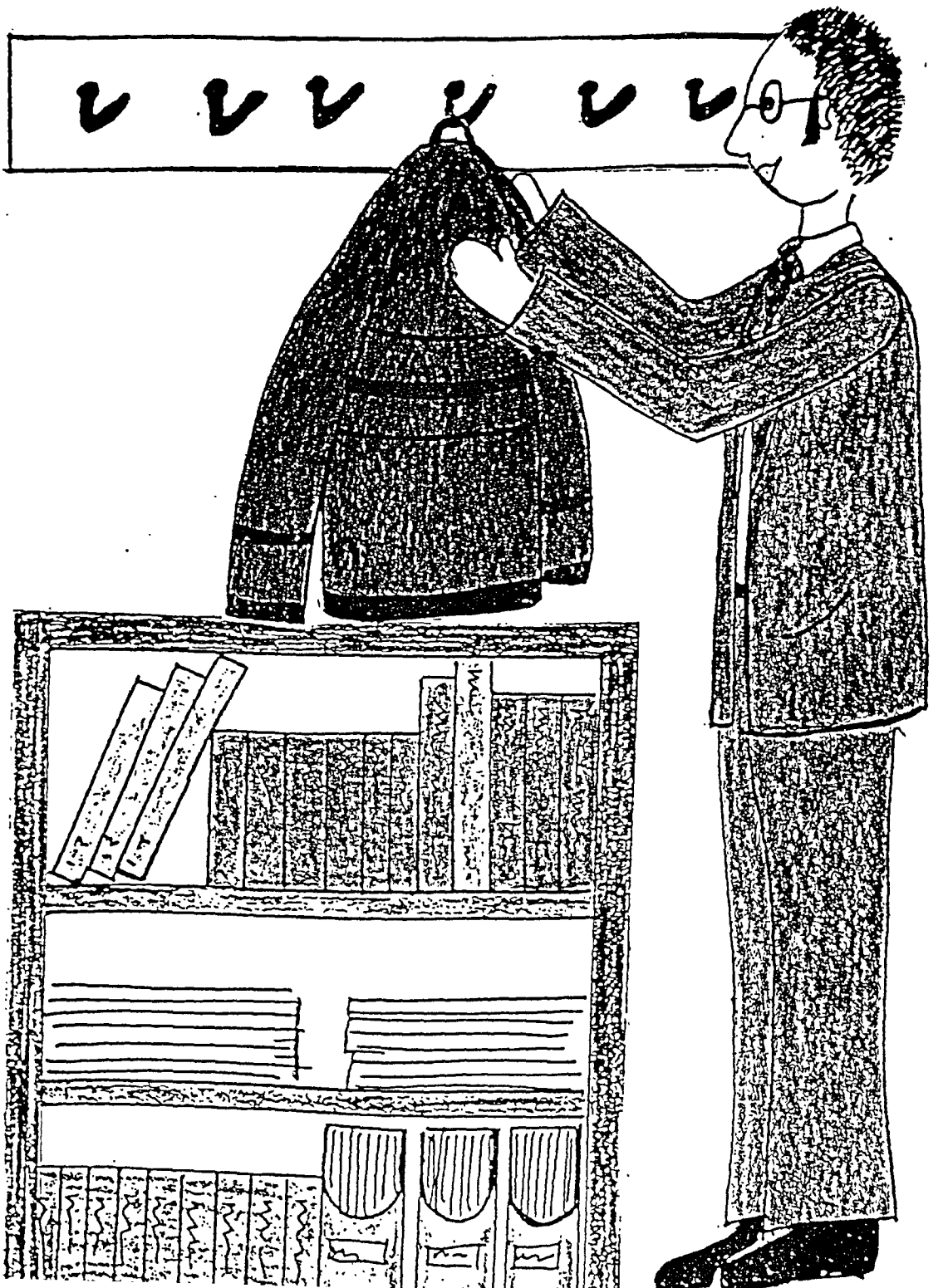




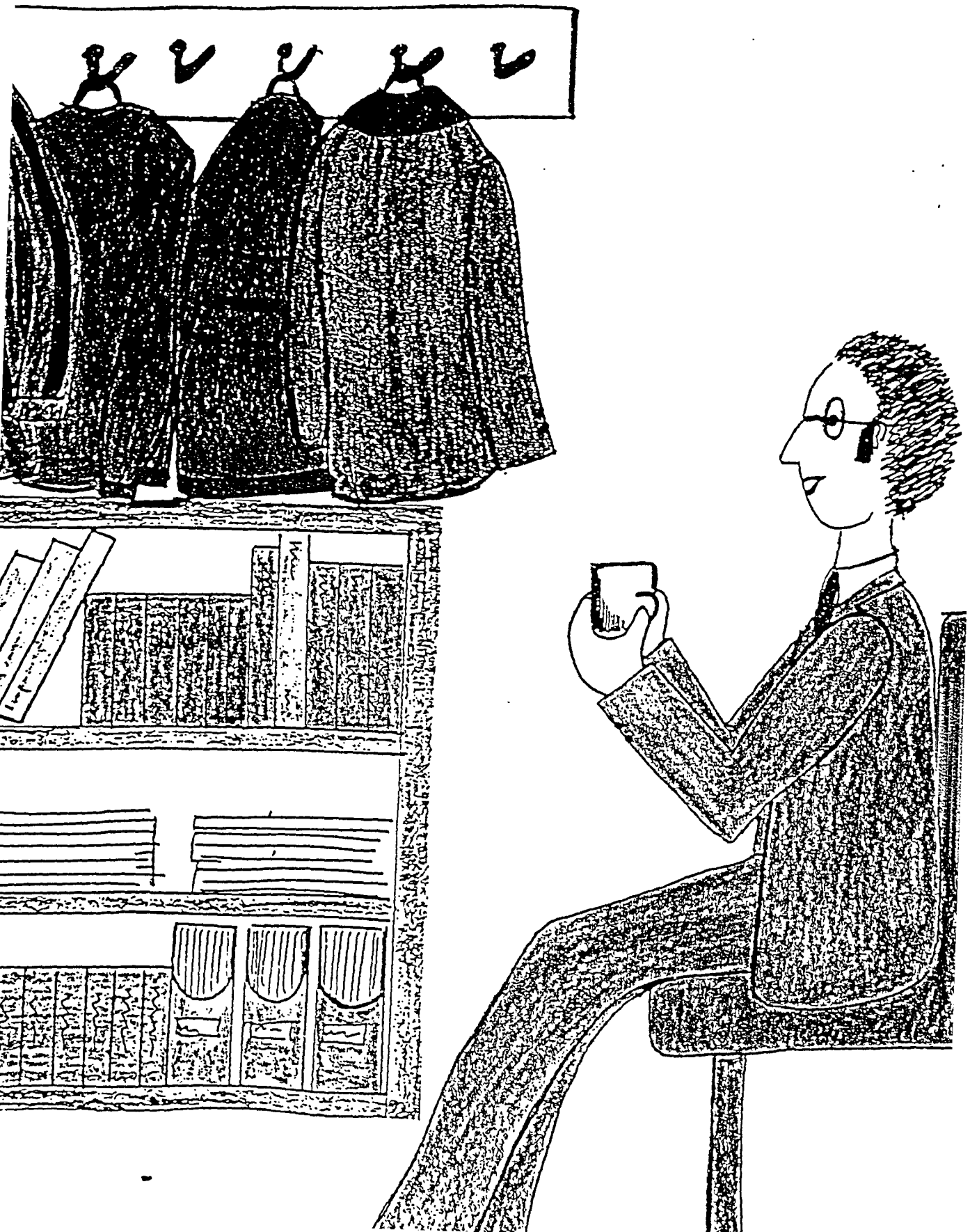
At school, Mr. Riley walks into the staffroom and hangs his coat above the bookshelves.



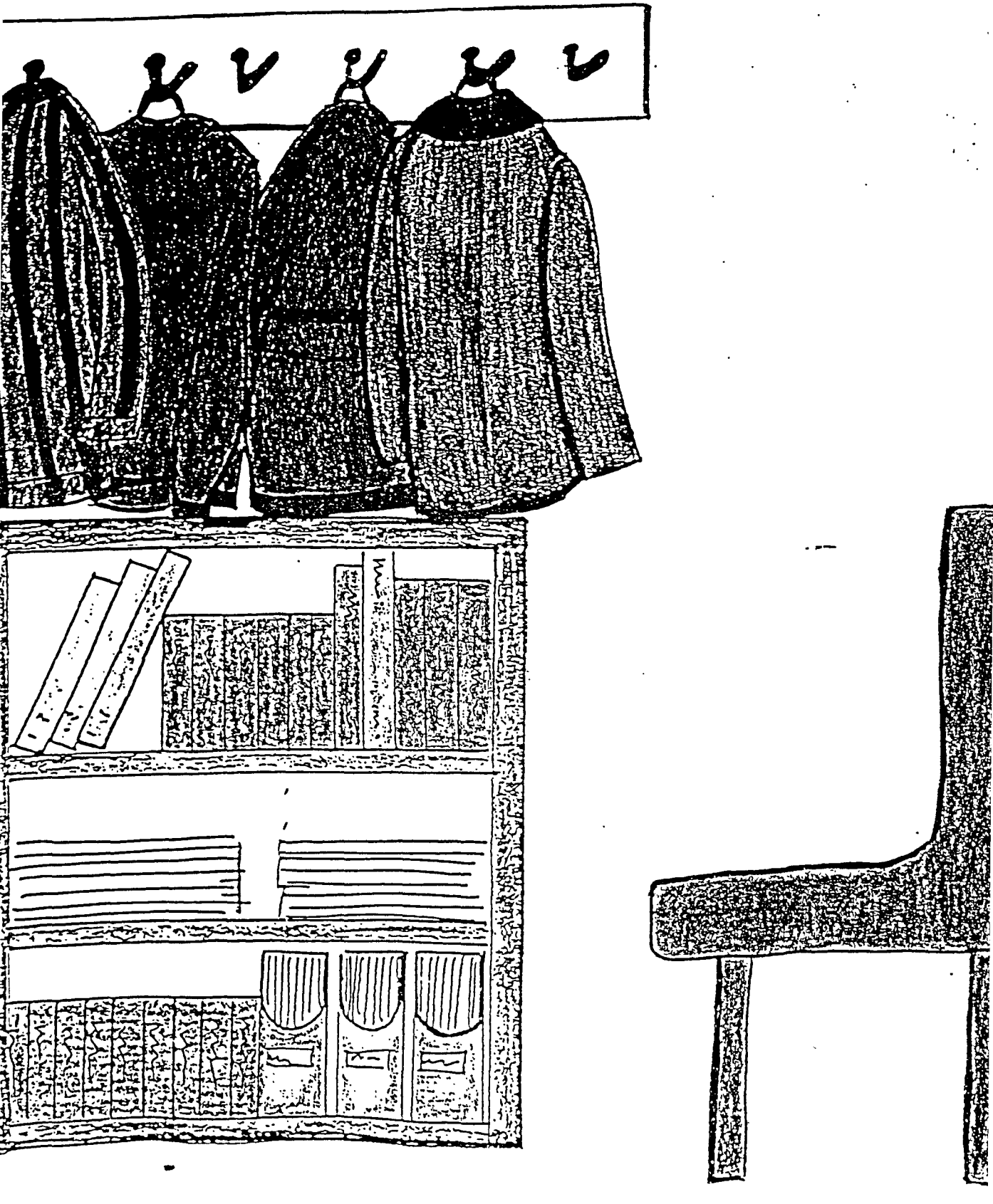
At school, Mr. Riley walks into the staffroom and hangs his coat above the bookshelves. So, there is only one coat hanging there.

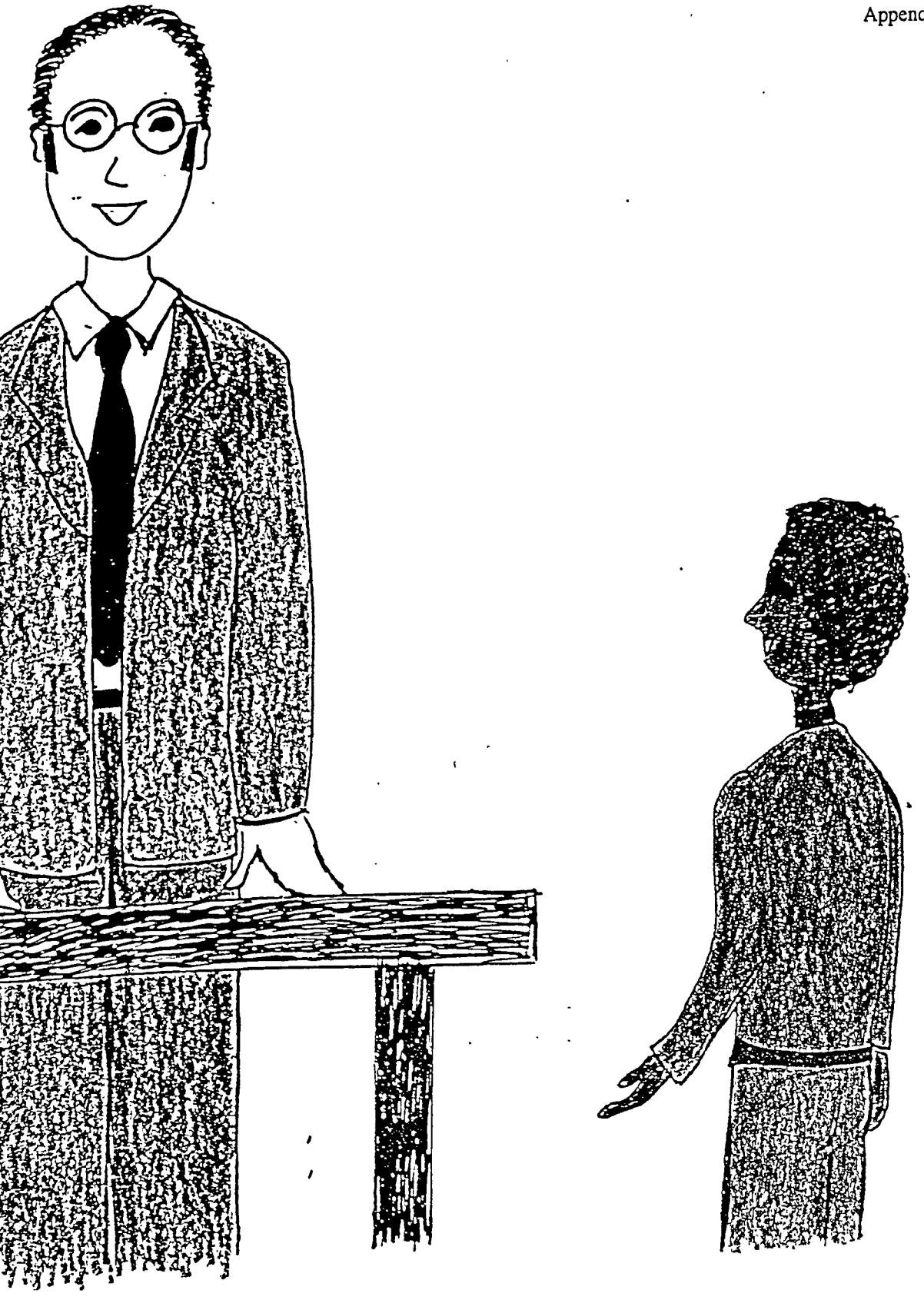


Whilst he drinks a cup of coffee, more teachers come in and hang their coats up. There are lots of coats above the bookshelves now.

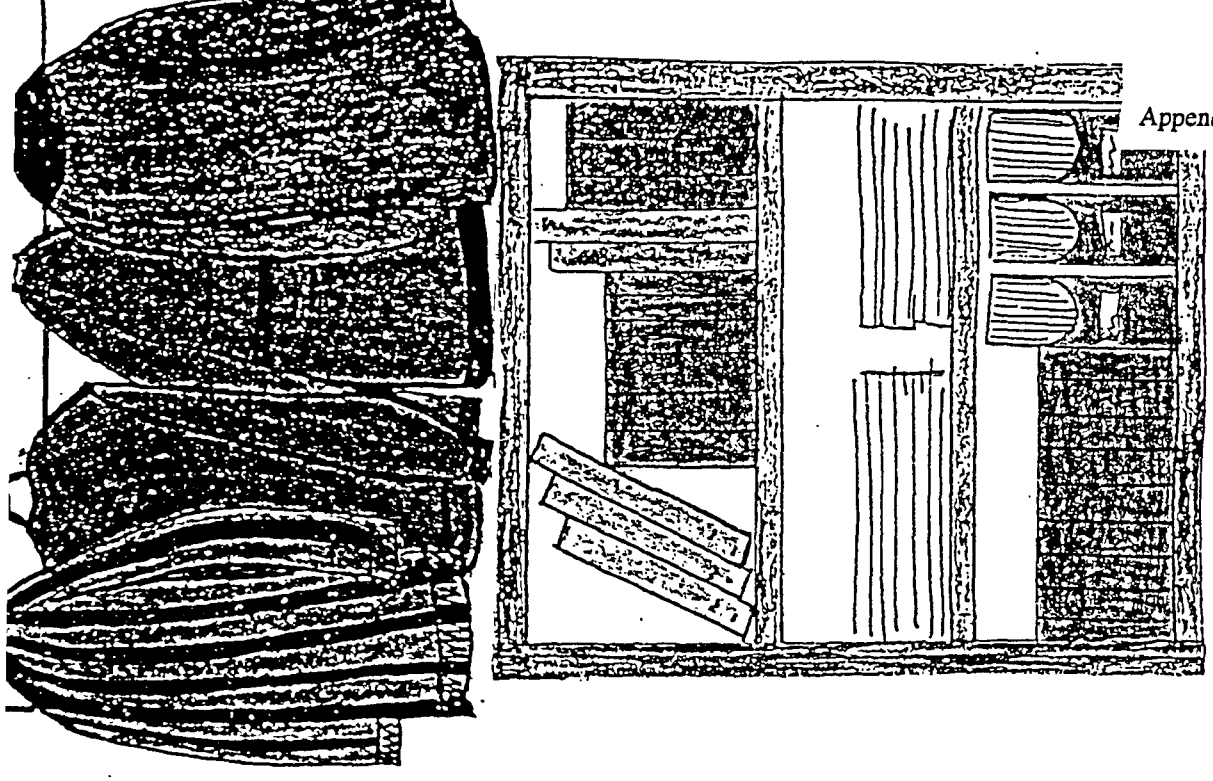


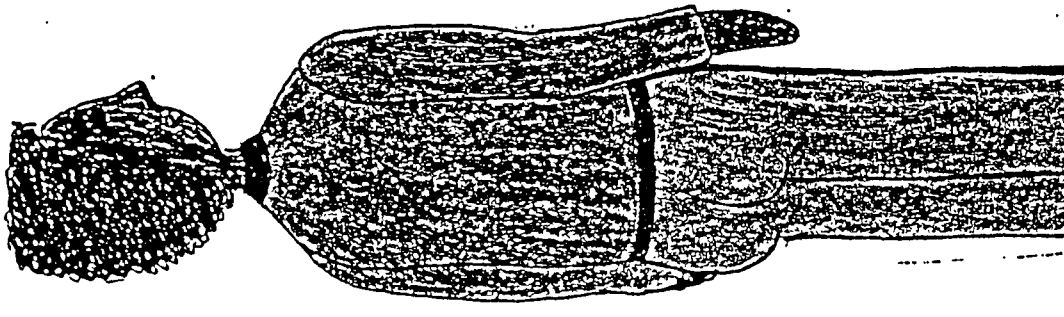
After Mr.Riley has left to go to his classroom, more teachers arrive and hang their coats up. There are lots of coats above the bookshelves now.

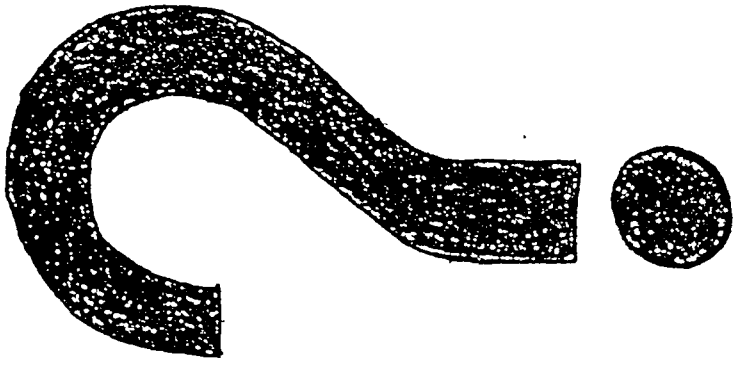


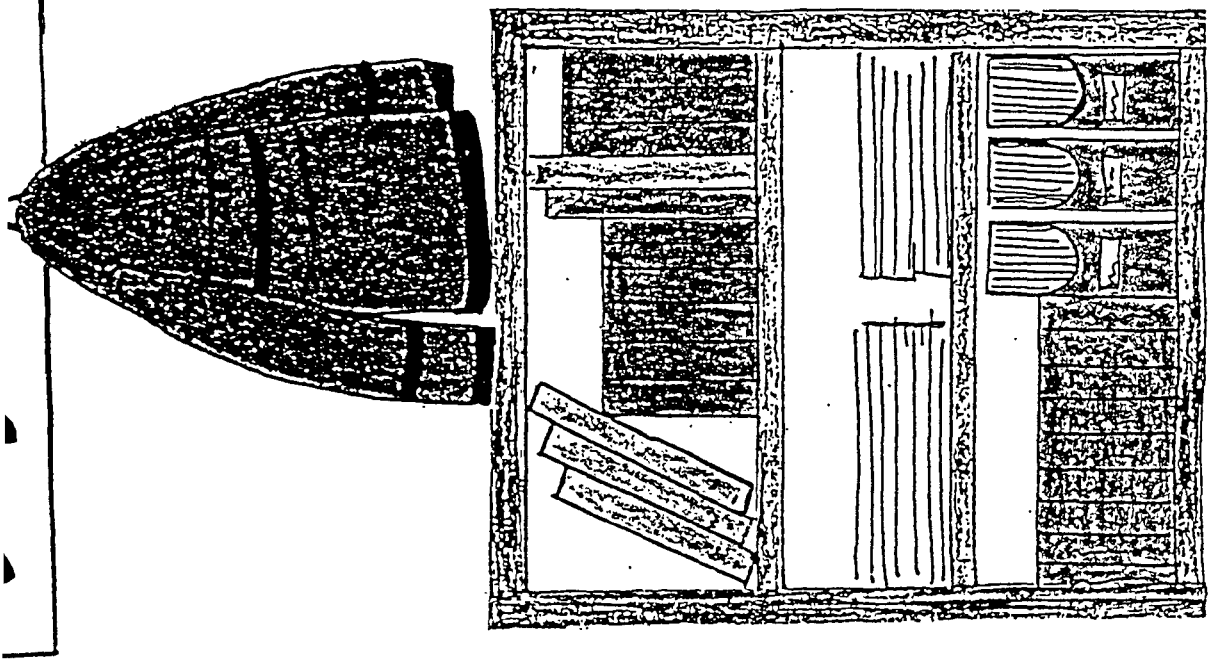


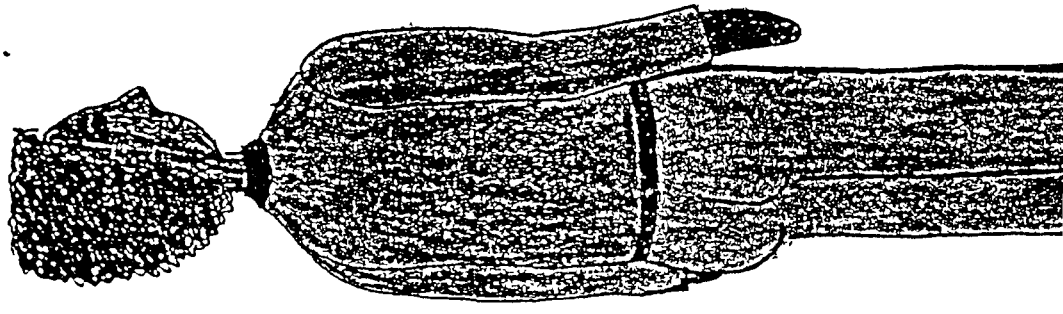
Later, Mr. Riley wants to show his class something outside and asks Nathan if he would fetch his coat from the staffroom. Nathan asks how he will know which coat is Mr. Riley's. "It's the one above the bookshelves" says Mr. Riley.

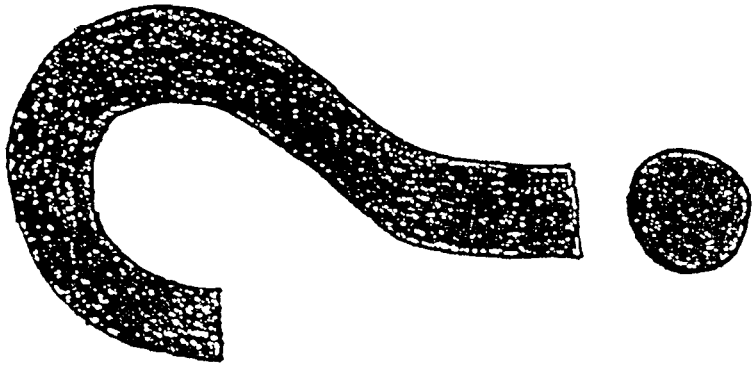




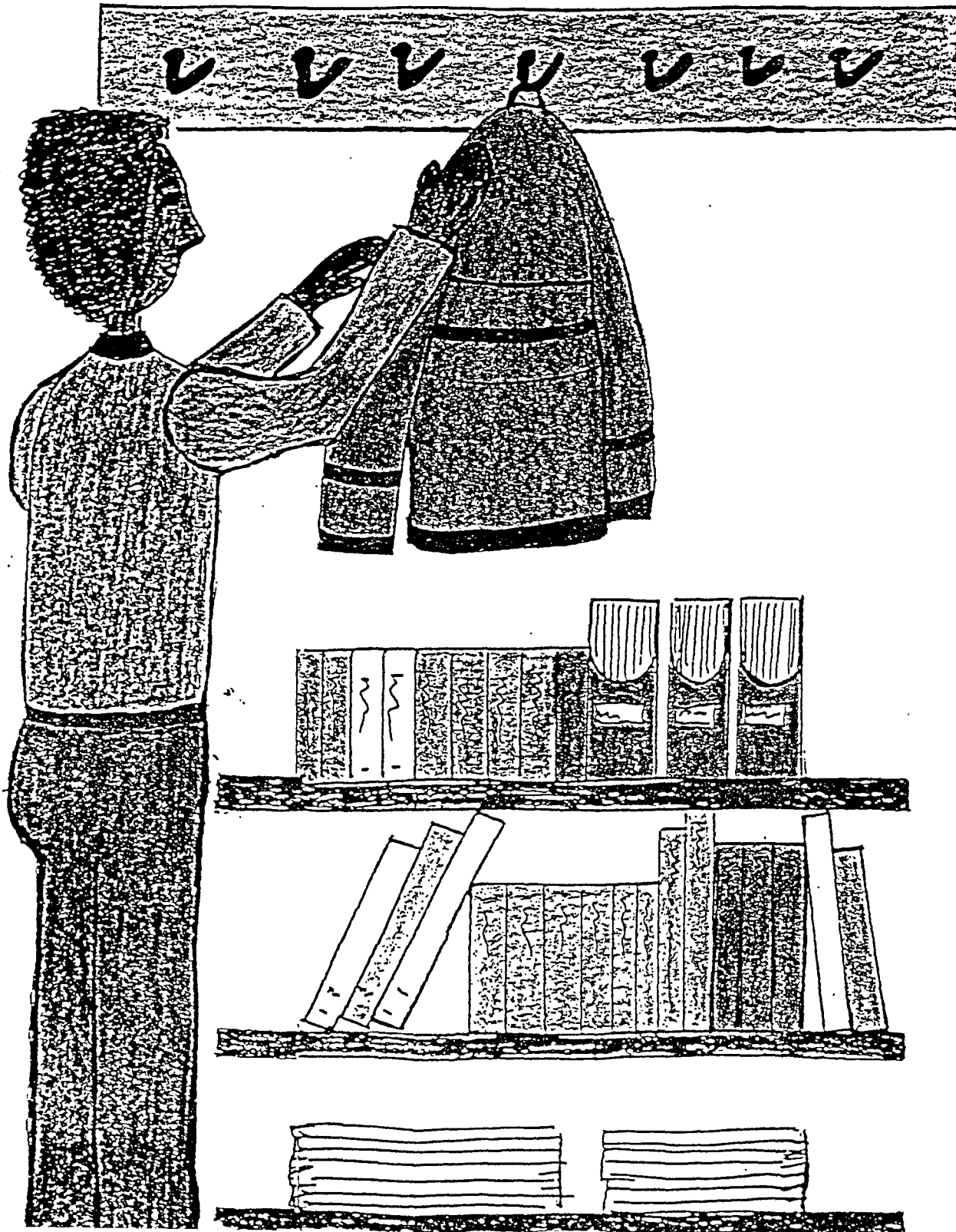






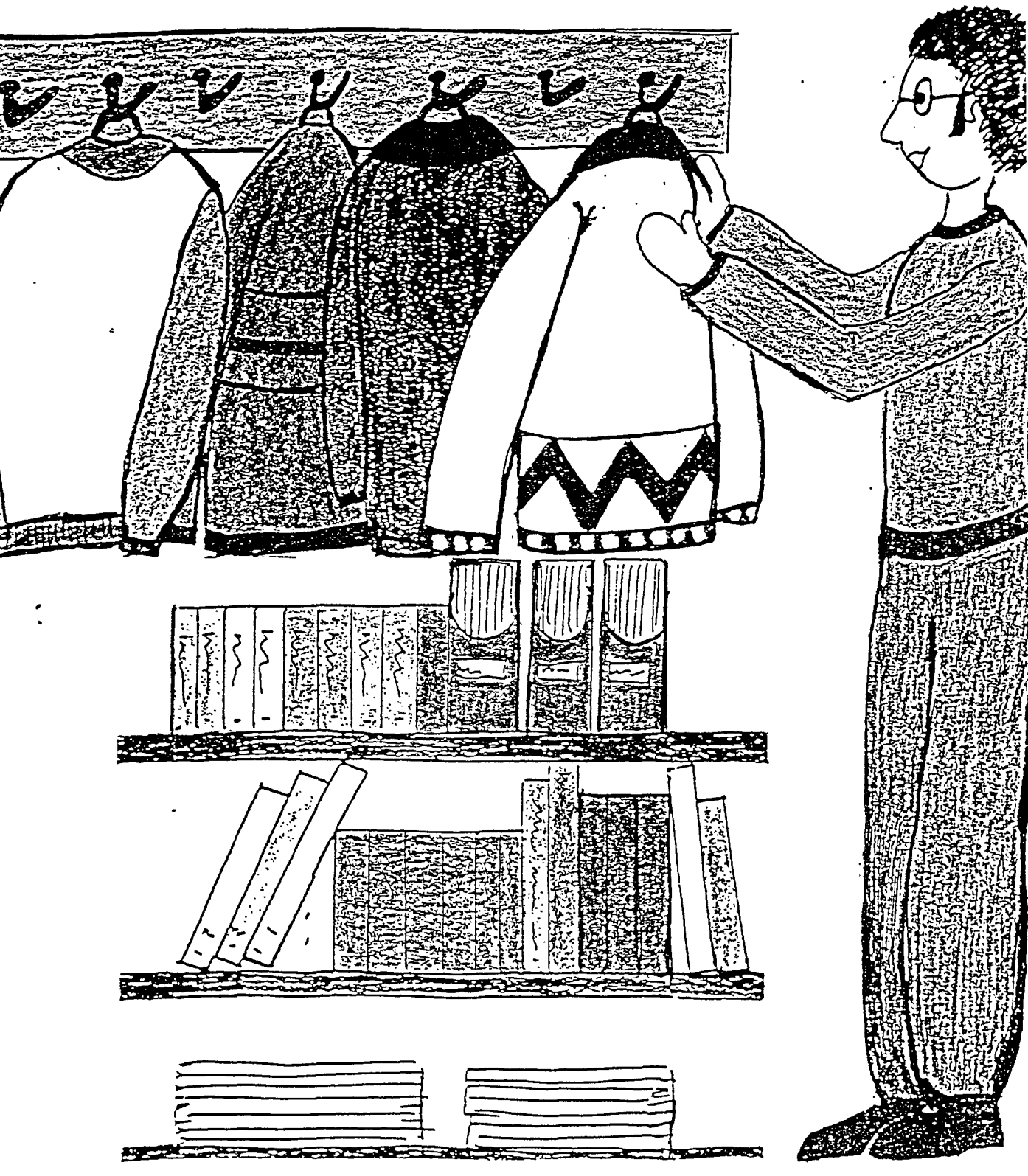


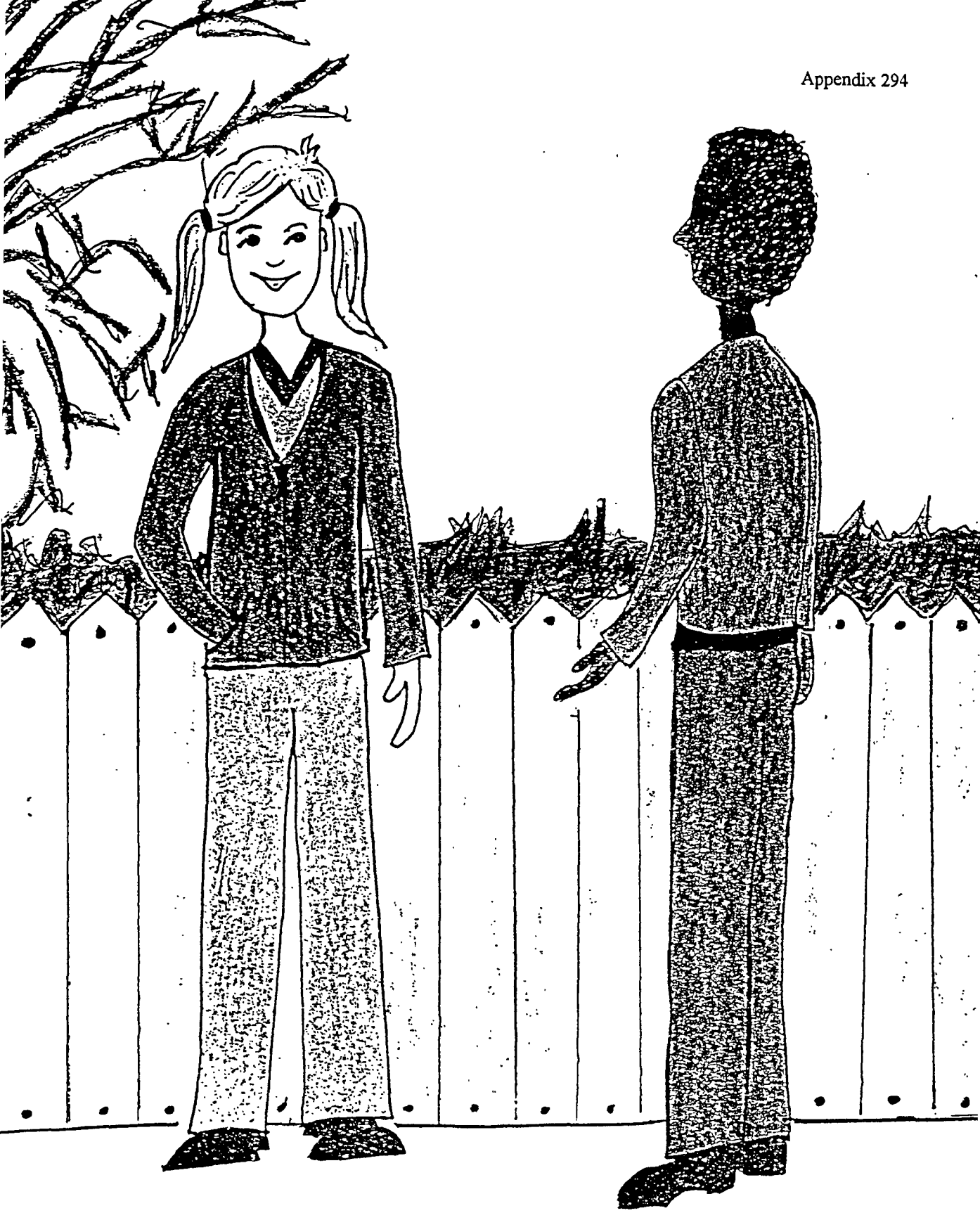
Nathan arrives at the Youth Club and hangs his coat above the bookshelves. His is the only coat hanging there.



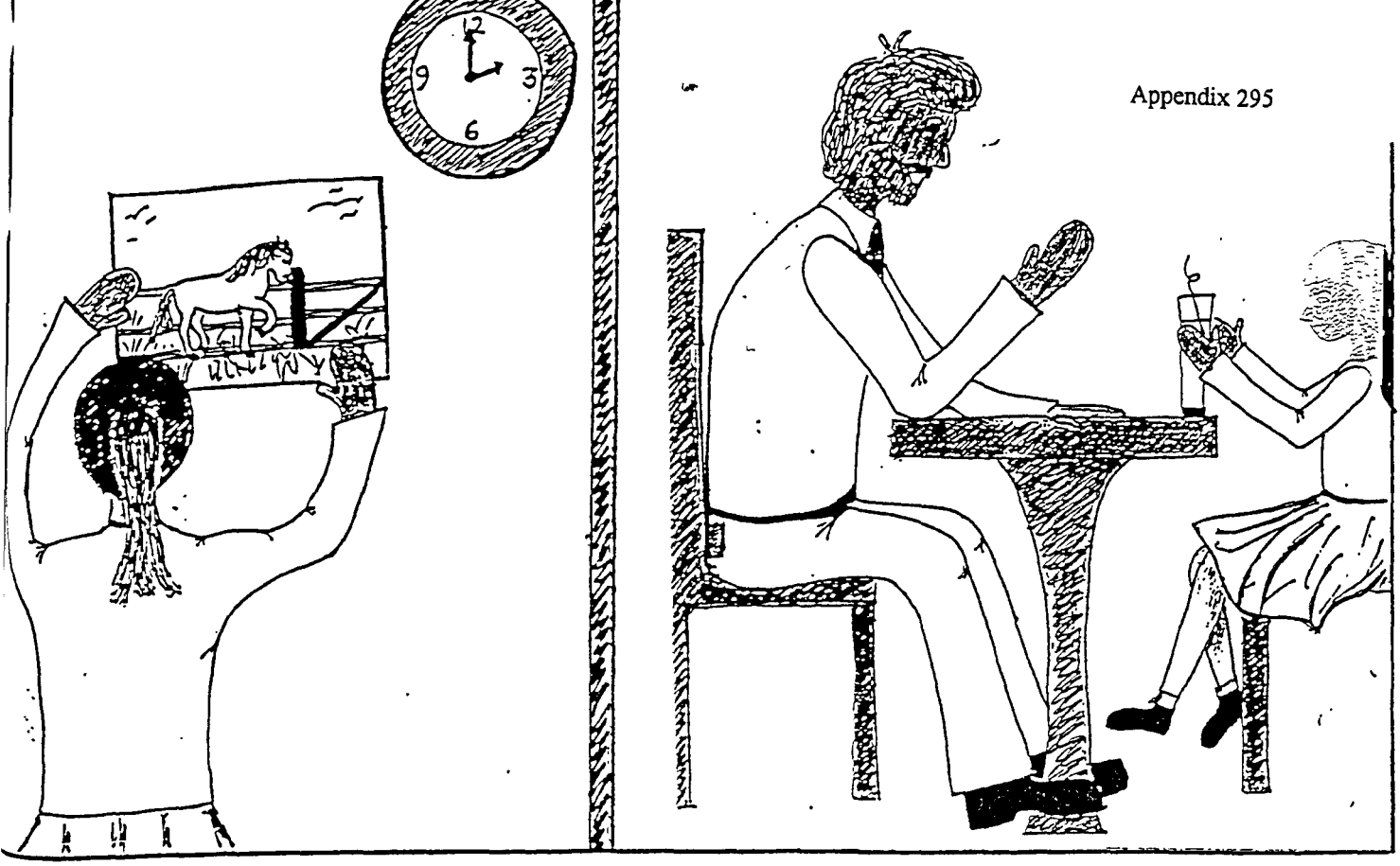
Whilst Nathan waits for his friend, the youth leader moves some coats from a chair and hangs them up next to Nathan's. So there are lots of coats above the bookshelves now.

After Nathan has gone outside, the club leader moves some coats from a chair and hangs them next to Nathan's. So there are lots of coats above the bookshelves now.





Whilst Sam and Nathan are outside playing, Sam decides to fetch a ball for their game. Nathan asks her if she could get his coat because he is cold. Sam asks how she will know which coat is Nathan's. He says "It's the one above the bookshelves."



ophie has drawn a picture for her school's open
ening. It has a large white horse on it. She puts
r picture on the empty wall next to the classroom
ock. So there is only one picture on the wall.

Later at home, Dad asks Sophie which picture to look out for when he visits the school. Sophie says "My picture is the one by the clock."

Did Sophie do a good job of saying exactly which picture was hers? (Yes) / No

Why/why not?

Because mee might be more pictures by the clock, which had been put up after she left.

Do you think Dad asks Sophie any more about the picture, OR does he just head off to the school?

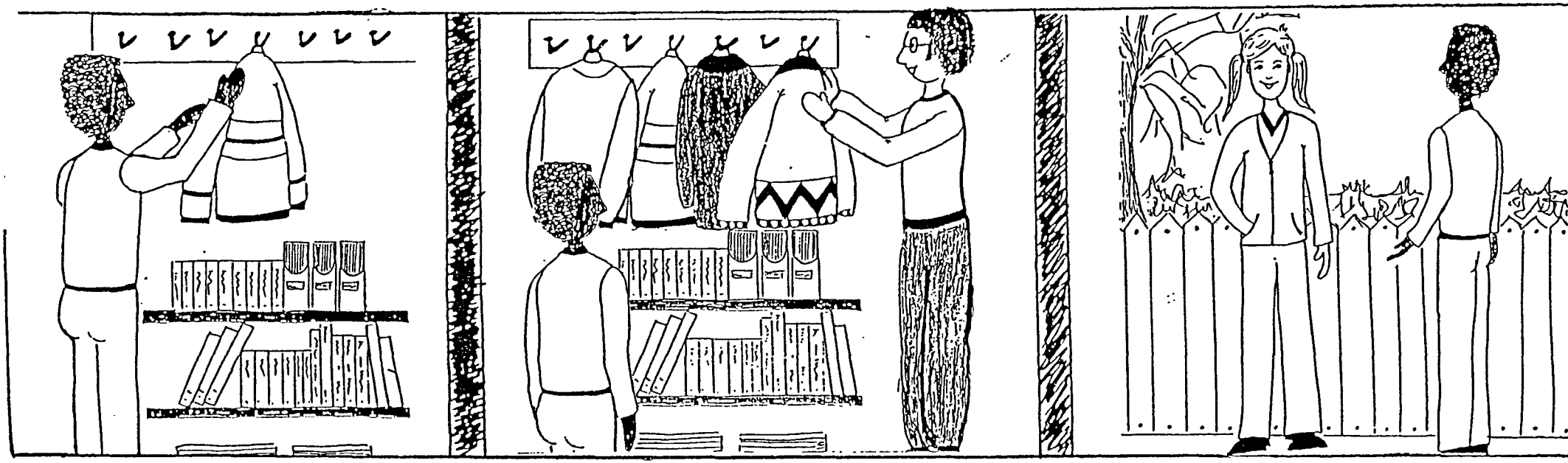
yes, what is it of?

DON'T FORGET.....

Age (years and months):
19

Sex: ~~Male~~ / Female

THANKS VERY MUCH!



Nathan arrives at the Youth Club and hangs his coat above the bookshelves. His is the only coat hanging there.

Whilst Nathan waits for his friend, the youth leader moves some coats from a chair and hangs them up next to Nathan's. So there are lots of coats above the bookshelves now.

Whilst Sam and Nathan are outside playing, Sam decides to fetch a ball for their game. Nathan asks her if she could get his coat because he is cold Sam asks how she will know which coat is Nathan' He says "It's the one above the bookshelves."

Did Nathan do a good job of saying exactly which coat was his? ~~Yes~~/ No

~~Why~~/why not? Because he could've guessed that more people would've put their coats on the coat racks above the bookshelves.

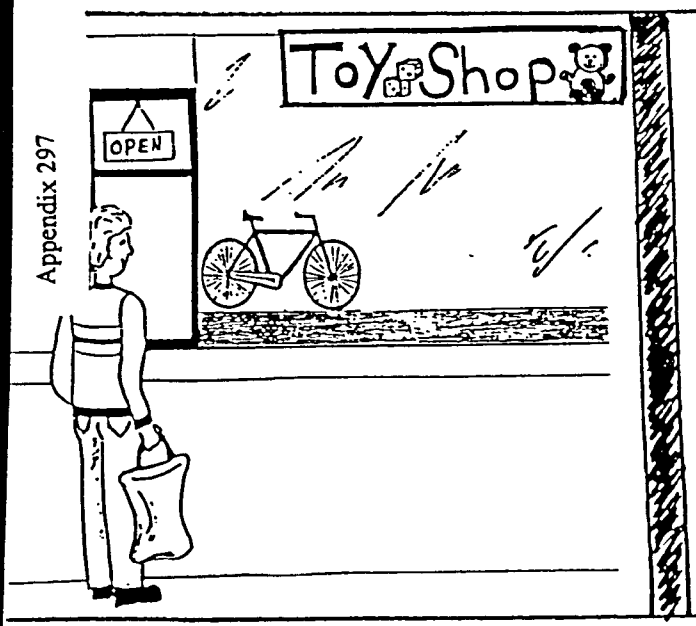
Do you think Sam just goes to fetch the coat, OR does she ask Nathan any more about it?

sam would probably ask some more details (what it looks like) so that her journey wouldn't be wasted.

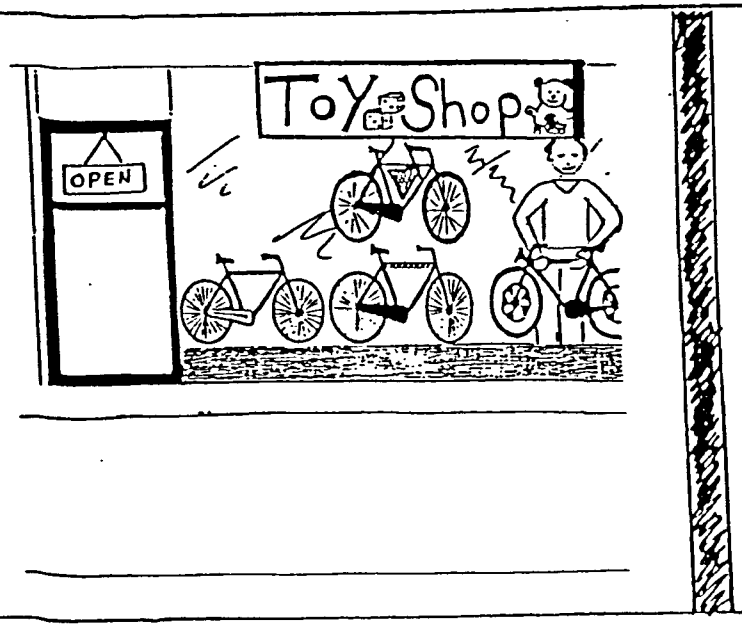
DON'T FORGET.....

Age (years and months): 19yrs 7mths.

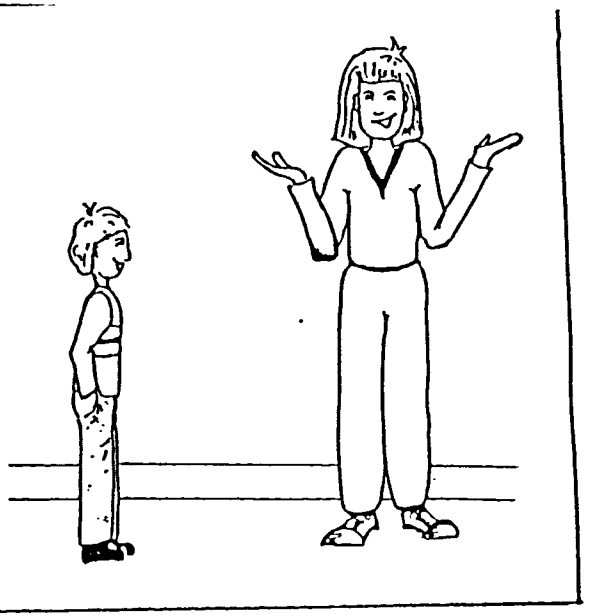
Sex: ~~Male~~/ Female



)ave is shopping in town when he walks past
ıe toy shop. In the window is a new bike which
)ave stops to have a look at. So, there is only
ıne bike on display.



Whilst Dave walks home, the shopkeeper brings out more bikes to display in the window. There are lots of bikes on display now.



Later at home, Dave's Mum asks him if he has seen anything he would like to have for his birthday. Dave says "Yes, there was a bike in the toy shop window today. I'd really like that."

)o you think Mum just heads off to the shops, OR does she ask Dave any more about the bike?

She'd ask him what it looked like

Did Dave do a good job of saying exactly which bike he wanted? Yes / ~~No~~

Why/why not? There was only one when he walked past.

DON'T FORGET.....

Age (years and months): 19 yrs 3 month

Sex: ~~Male~~ / Female

THANKS VERY MUCH!!

1. Uas/AB/jumper first

Please give your age 21 years; 6 months. Please circle: Male (Female)

Rachel went shopping for some groceries. On the way back, she passed Stevenson's clothes shop and saw a really stylish jumper in the window. It was the only jumper on display. The rest of the clothes were dresses and coats. Rachel then went straight home. No more jumpers were added to the display in Stevenson's shop. When Rachel arrived home, her friend Susan greeted her and asked if there was anything she would like for her birthday, which was tomorrow. Rachel said: **"There's a really nice jumper on display in the window in Stevenson's clothes shop. You'll spot it easily because it's the only one there."** Susan was glad to discover Rachel had seen something specific and decided to get the jumper. Susan was very eager to set out to buy the jumper as soon as possible, especially as it was nearing closing time.

Question: What do you think happened next? (circle either A or B)

(A) Do you think Susan asked Rachel for more information about the appearance of the jumper?
OR
(B) Do you think Susan went out right away to Stevenson's clothes shop?

Julie went to the night-club bar to get a round of drinks. There was a really good-looking guy behind the bar. He was the only guy there. All of the other bartenders were women. Julie couldn't carry all the drinks herself so she asked the barman if he could look after them. Then Julie walked back to her friends. No other guys were working behind the bar that night. Claire offered to fetch the remaining drinks from the bar and wanted to know who she should ask. Julie said: **"There is a really cute guy serving the drinks. You'll spot him easily because he is the only guy there."** Claire was glad to discover that the drinks were waiting for her at the bar. She was eager to get the drinks as soon as possible so that she could see the guy Julie was talking about.

Question: What do you think happened next? (please circle either A or B)

(A) Do you think Claire asked Julie any more about the appearance of the guy behind the bar?
OR
(B) Do you think Claire went straight away to fetch the drinks?

7. As/AB/bar first

Please give your age 41 years; 5 months. Please circle: Male (Female)

Julie went to the night-club bar to get a round of drinks. There was a really good-looking guy behind the bar. He was the only guy there. All of the other bartenders were women. Julie couldn't carry all the drinks herself so she asked the barman if he could look after them. Then Julie walked back to her friends. Meanwhile, more male bar staff arrived for work so there were a few guys behind the bar now. Claire offered to fetch the remaining drinks from the bar and wanted to know who she should ask. Julie said: **"There is a really cute guy serving the drinks. You'll spot him easily because he is the only guy there."** Claire was glad to discover that the drinks were waiting for her at the bar. She was eager to get the drinks as soon as possible so that she could see the guy Julie was talking about.

Question: What do you think happened next? (please circle **either** A **or** B)

(A) Do you think Claire asked Julie any more about the appearance of the guy behind the bar?
OR
(B) Do you think Claire went straight away to fetch the drinks?

Rachel went shopping for some groceries. On the way back, she passed Stevenson's clothes shop and saw a really stylish jumper in the window. It was the only jumper on display. The rest of the clothes were dresses and coats. Rachel then went straight home. Meanwhile, the owner of Stevenson's shop rearranged the window. He removed the dresses and coats but added several jumpers to accompany the one already on display. When Rachel arrived home, her friend Susan greeted her and asked if there was anything she would like for her birthday, which was tomorrow. Rachel said: **"There's a really nice jumper on display in the window in Stevenson's clothes shop. You'll spot it easily because it's the only one there."** Susan was glad to discover Rachel had seen something specific and decided to get the jumper. Susan was very eager to set out to buy the jumper as soon as possible, especially as it was nearing closing time.

Question: What do you think happened next? (circle **either** A **or** B)

(A) Do you think Susan asked Rachel for more information about the appearance of the jumper?
OR
(B) Do you think Susan went out right away to Stevenson's clothes shop?

Please give your age 21 years; 9 months. Please circle: Male (Female)

Mark is reading a book at desk 31 in the busy college library. It is the only book on his desk. He decides he needs some more information and goes to the bookshelves on the next floor. Whilst Mark is walking to the shelves, Sue is on her way out of the library. She has finished reading her books and puts them down on desk 31, next to the book that Mark was reading. At the bookshelves, he sees his friend Jim, who is reading about a similar topic. Jim asks Mark if there is a book he can recommend. Mark says "Yes. I have a great one downstairs on my desk. You'll spot it easily because it's the only one there." Mark tells Jim he is sat at desk 31, near the exit.

(1) What do you think happened next? (circle either A or B)

A) Do you think Jim asked Mark for more information about the book?

OR

(B) Do you think Jim went straight away to fetch the book from the desk?

(2) How often do extra books end up on people's desks like this? 78

[If 0% represents Very Rarely, 50% represents Sometimes and 100% represents Very Frequently, give a percentage value anywhere between 0-100 that estimates the frequency with which you would expect this to happen.]

(3) Indicate on the list below the description that most closely represents the percentage value that you have stated above.

Every day
✓Every 2-3 days
Once a week
Once a fortnight
Once a month
Less than once a month
Less than once every six months
Less than once a year
Almost never

Please give your age 21 years; 8 months. Please circle: Male (Female)

Mark is reading a book at desk 31 in the busy college library. It is the only book on his desk. He decides he needs some more information and goes to the bookshelves on the next floor. Whilst Mark is walking to the shelves, Sue the Librarian struggles past desk 31 with a heavy over-filled box. She is transferring some books to a new room in the library. Some of the books fall onto Mark's desk, next to the book that Mark was reading. At the bookshelves, he sees his friend Jim, who is reading about a similar topic. Jim asks Mark if there is a book he can recommend. Mark says "Yes. I have a great one downstairs on my desk. You'll spot it easily because it's the only one there." Mark tells Jim he is sat at desk 31, near the exit.

(1) What do you think happened next? (circle either A or B)

A) Do you think Jim asked Mark for more information about the book?

OR

(B) Do you think Jim went straight away to fetch the book from the desk?

(2) How often do extra books end up on people's desks like this?

[If 0% represents Very Rarely, 50% represents Sometimes and 100% represents Very Frequently, give a percentage value anywhere between 0-100 that estimates the frequency with which you would expect this to happen.]

2.1.

(3) Indicate on the list below the description that most closely represents the percentage value that you have stated above.

Every day
Every 2-3 days
Once a week
Once a fortnight
Once a month
Less than once a month
Less than once every six months
~~Less than once~~ a year
(Almost never)

2.unp/am/b

Please give your age 21 years; 1 months. Please circle: Male (Female)

Mark is reading a book at desk 31 in the busy college library. It is the only book on his desk. He decides he needs some more information and goes to the bookshelves on the next floor. Whilst Mark is walking to the shelves, Sue is on her way out of the library. She has finished reading her books and puts them down on desk 32, the desk next to Mark's. At the bookshelves, he sees his friend Jim, who is reading about a similar topic. Jim asks Mark if there is a book he can recommend. Mark says "Yes. I have a great one downstairs on my desk. You'll spot it easily because it's the only one there." Mark tells Jim he is sat at desk 31, near the exit.

(1) What do you think happened next? (circle either A or B)

(A) Do you think Jim went straight away to fetch the book from the desk?

OR

B) Do you think Jim asked Mark for more information about the book?

(2) How often do extra books end up on people's desks like this?

[If 0% represents Very Rarely, 50% represents Sometimes and 100% represents Very Frequently, give a percentage value anywhere between 0-100 that estimates the frequency with which you would expect this to happen.]

60%

(3) Indicate on the list below the description that most closely represents the percentage value that you have stated above.

Every day
Every 2-3 days ✓
Once a week
Once a fortnight
Once a month
Less than once a month
Less than once every six months
Less than once a year
Almost never

3. pre/unam/a

Please give your age 19 years; 7 months. Please circle: Male (Female)

Mark is reading a book at desk 31 in the busy college library. It is the only book on his desk. He decides he needs some more information and goes to the bookshelves on the next floor. Whilst Mark is walking to the shelves, a Librarian struggles past desk 32 with a heavy over-filled box. She is transferring some books to a new room in the library. Some of the books fall onto desk 32, the desk next to Mark's. At the bookshelves, he sees his friend Jim, who is reading about a similar topic. Jim asks Mark if there is a book he can recommend. Mark says "Yes. I have a great one downstairs on my desk. You'll spot it easily because it's the only one there." Mark tells Jim he is sat at desk 31, near the exit.

(1) What do you think happened next? (circle either A or B)

A) Do you think Jim asked Mark for more information about the book?

OR

(B) Do you think Jim went straight away to fetch the book from the desk?

(2) How often do extra books end up on people's desks like this? 10%

[If 0% represents Very Rarely, 50% represents Sometimes and 100% represents Very Frequently, give a percentage value anywhere between 0-100 that estimates the frequency with which you would expect this to happen.]

(3) Indicate on the list below the description that most closely represents the percentage value that you have stated above.

Every day
Every 2-3 days
Once a week
Once a fortnight
Once a month
Less than once a month
Less than once every six months
Less than once a year
Almost never

4. unpr/unam/b