



Appearance questions can be misleading: A discourse-based account of the appearance–reality problem

Mikkel B. Hansen*, Ellen M. Markman

*Department of Psychology, Stanford University, Jordan Hall, Building 420, Main quad,
Stanford, CA 94305-2130, United States*

Accepted 5 September 2004
Available online 21 November 2004

Abstract

Preschoolers' success on the appearance–reality task is a milestone in theory-of-mind development. On the standard task children see a deceptive object, such as a sponge that looks like a rock, and are asked, “What is this really?” and “What does this look like?” Children below 4½ years of age fail saying that the object not only is a sponge but also looks like a sponge. We propose that young children's difficulty stems from ambiguity in the meaning of “looks like.” This locution can refer to outward appearance (“Peter looks like Paul”) but in fact often refers to likely reality (“That looks like Jim”). We propose that “looks like” is taken to refer to likely reality unless the reality is already part of the common ground of the conversation. Because this joint knowledge is unclear to young children on the appearance–reality task, they mistakenly think the appearance question is about likely reality. Study 1 analyzed everyday conversations from the CHILDES database and documented that 2 and 3-year-olds are familiar with these two different uses of the locution. To disambiguate the meaning of “looks like,” Study 2 clarified that reality was shared knowledge as part of the appearance question, e.g., “What does the sponge look like?” Study 3 used a non-linguistic measure to emphasize the shared knowledge of the reality in the appearance question. Study 4 asked children on their own to articulate the contrast between appearance and reality. At 91%, 85%, and 81% correct responses, children were at near ceiling levels in each of our manipulations while they failed

* Corresponding author. Fax: +1 650 725 5699.

E-mail addresses: mikkel@psych.stanford.edu, mikkelh@ruc.dk (M.B. Hansen).

the standard versions of the tasks. Moreover, we show how this discourse-based explanation accounts for findings in the literature. Thus children master the appearance–reality distinction by the age of 3 but the standard task masks this understanding because of the discourse structure involved in talking about appearances.

© 2004 Elsevier Inc. All rights reserved.

1. Introduction

The need to distinguish appearances from reality is ubiquitous in children's lives. They may come across a cookie made out of play dough or a shadow on the wall that looks like a snake. They may have to explain misperceptions like going to the wrong car in the parking lot because it looks like the family car. They may find themselves the target of deliberately misleading appearances when people put on costumes, or find themselves on the receiving end of misleading actions and emotions when someone makes an ironic comment with an expressionless face.

Given its importance for understanding everyday perceptual, representational, social, and emotional phenomena, it is important to track the development of the ability to distinguish between appearances and reality. The appearance–reality task devised by Flavell and his colleagues (Flavell, Flavell, & Green, 1983; Flavell, Green, & Flavell, 1986; Flavell, Zhang, Zou, Dong, & Qi, 1983; Taylor & Flavell, 1984) has become the standard for evaluating whether children have mastered this distinction. The focus of their studies was on real and apparent object identities, such as the misleading outward appearances of deceptive objects; real and apparent object properties, such as the apparent size, color, and shape of various objects seen through distorting devices; and real and apparent action identities, such as an action viewed from two different angles. In the most common version of this task children see a deceptive object, such as a sponge that looks like a rock, and are asked what it is and what it looks like. To answer correctly children must understand that appearances can be misleading, but children below 4½ years of age typically fail saying that the object not only *is* a sponge but also *looks like* a sponge, thus making so-called realism errors (e.g., Flavell et al., 1986). Young children's failure is particularly striking because of the seeming ease of the task. Children are given the correct answer as part of the introduction to the objects just moments before the test question, e.g., "So this looks like a rock but it is really a sponge." When asked the test questions "What does this look like?" and "What is it really?" all children would have to do is repeat what they just heard.

The original finding that children below 4½ years of age fail these tasks has been replicated in numerous studies over the last 20 years in diverse cultures and languages (e.g., Flavell, 1993; Flavell et al., 1986; Flavell, Zhang et al., 1983; Gopnik & Astington, 1988; Melot & Houdé, 1998; Rice, Koinis, Sullivan, Tager-Flusberg, & Winner, 1997; Sapp, Lee, & Muir, 2000; Sodian, Huelsken, Ebner, & Thoermer, 1998). Outside of object identities and a wide range of object properties, the ability to distinguish between appearance and reality has been studied in the area of social understanding. Children's understanding that other children can act morally or

immorally while still looking mean or nice shows similar results to other appearance–reality tasks (Flavell, Lindberg, Green, & Flavell, 1992). Attempts to show that younger children appreciate the distinction have generally been unable to challenge the pattern of findings although with some exceptions that we will address in the discussion.

The general conclusion has been that children do not appreciate the difference between appearance and reality until late in their fourth year. A dominant theoretical claim is that children have not yet developed the cognitive foundations required to understand the distinction, namely that the human mind may represent objects in different ways (e.g., Flavell et al., 1986; Gopnik & Wellman, 1994; Perner, 1991). They are thus left unable to understand false beliefs, mistakes, or misleading appearances. In this work, we propose an alternative account of children’s problems on the appearance–reality task which focuses on the conversational requirements involved in talking about appearances. On our account, the discourse requirements specific to the key expression *looks like* cause children to fail these tasks rather than deficits in children’s understanding of the appearance–reality distinction.

It has been assumed in the literature that the use of *looks like* on the appearance–reality task refers unequivocally to the outward appearance of the deceptive object. However, we will argue that people do not only use *looks like* to refer to outward appearances; in fact, their default assumption is that it refers to likely reality. Consider the statement “It looks like the meeting will be on Saturday.” The speaker is using *looks like* to make a statement about what will most likely happen in the sense of “I think the meeting is going to be on Saturday.” Consider further a situation where a speaker sees another person from afar and says, “That looks like Paul.” Even if the speaker bases her statement on the outward appearances of the person, the force of the statement is what she thinks is the most probable identity of the person, in the sense of “I think that’s Paul.” Along the same lines, consider the situation where two speakers come across an unfamiliar object. If one of them asks, “What is this?” and the other person answers, “It looks like an accordion,” the focus of this comment is on what the object is likely to be more than on outward appearances.

Examples of *looks like* as referring to likely reality—roughly synonymously with “I think it is”—are easy to come by. But objects, people, and situations can have misleading appearances and it is still possible to refer to such false states of affairs with the phrase *looks like*. The challenge then is to explain which construal of *looks like* is intended. When speakers mean to comment on false appearances for example with the phrase “It looks like a rock,” how do they avoid being misunderstood as saying that the identity of the object is a rock? We suggest that speakers rely on a convention about the discourse used with *looks like*: the default assumption is that *looks like* refers to likely reality. The conversational context may however indicate that the reality of the referent is already known by the conversation partners. In such cases, this shared knowledge blocks the default assumption and marks that *looks like* refers to outward appearances.

New utterances in a conversation are generally made against the background of a set of assumptions about the other participants’ state of knowledge, often referred to

as the common ground of the speaker and listener (Clark, 1996). Common ground includes stated or unstated assumptions about several kinds of shared knowledge: the physical environment of the participants, what the participants are likely to be familiar with due to their membership of a given community, jointly perceived events before and during the conversation, and information exchanged verbally and non-verbally in the course of the conversation. Common ground is particularly important for interpreting ambiguous referential expressions (Clark, Schreuder, & Buttrick, 1983). If a speaker makes an ambiguous statement such as “Can you give me the jacket,” the listener’s interpretation may depend on whether one or more jackets are physically present, what kind of jacket a community member like the speaker might wear, whether both participants in the conversation had just witnessed an event like one jacket falling down, or whether the speaker had previously referred to a particular jacket, for example by saying, “The red jacket is my favorite.” Listeners are thus constantly gauging what is mutually known in order to interpret new possibly ambiguous utterances. Because of the ambiguity of utterances with *looks like*, we propose that accurately identifying what is common knowledge is the key to interpreting the utterance in the way the speaker intended.

More specifically, for *looks like* to refer to outward appearances language users rely on the presence of cues in the conversation that indicate that the identity of the object or situation is already common knowledge. Recall that the statement “That looks like Paul” in the absence of any other cues about the referent means roughly “I think that’s Paul,” in other words a statement about the probable identity of the referent. In contrast, the statement “Jack looks like Paul” includes a clear cue that the identity of the referent is already commonly known to be Jack.

The analysis that the phrase *look like* may have different interpretations depending on the context receives support from the philosopher J.L. Austin:

Clearly, if I say that petrol looks like water, I am simply commenting on the way petrol looks; I am under no temptation to think, nor do I imply, that perhaps petrol is water. Similarly with ‘A recorder sounds like a flute’. But ‘This looks like water’ (‘That sounds like a flute’) may be a different matter; if I don’t already know what ‘this’ is, then I may be taking the fact that it looks like water as a ground for thinking it is water. (Austin, 1962, p. 39)

In saying that petrol looks like water, the degree to which the speaker thinks outward appearances differ from reality may vary. The speaker may be implying that petrol shares certain properties, such as color or viscosity, with water in the sense of *resemble*, as in “John resembles uncle Bill.” The speaker may also be implying that it is easy to mistake petrol for water in the sense of *appears to be*, as in “John appeared to be uncle Bill for a moment.” For the purposes of the appearance–reality task, this difference in degree is not critical. Both indicate that the speaker knows reality and is talking about appearance.

Using the context to make it clear that the identity of a referent is common knowledge may undoubtedly take many forms, both verbal and non-verbal. One particularly clear way to verbally express knowledge of the identity on the part of the speaker is as the noun phrase in the sentence with *looks like*, e.g., “Petrol looks like water.”

The identity may also be clear from previous utterances in the ongoing conversation either by the speaker of the *looks like* phrase or other participants. Such turns of conversation can be characterized more precisely in terms of the conventions of discourse that have been laid out by Clark and Schaefer (1989). On their view conversations unfold in an orderly manner as a series of propositions that are either accepted or rejected as additions to the shared knowledge of the speakers and that may each build on previously accepted knowledge. If a speaker has previously stated with certainty what he or she assumes to be reality, e.g., “This is petrol,” and this was accepted as a true statement by the listener, then a later utterance like “It looks like water” builds on the assumption that the identity is commonly known, presuming that “it” refers to the same object.

The previous two examples depend on the identity being stated *prior* to the *looks like* phrase, but the identity may also be added as a qualification to the *looks like* phrase, e.g., “This looks like water, but it’s really petrol,” or “This looks like water but that’s not what it really is.”

Various situational aspects may also enhance the presence or absence of shared knowledge about reality. Frameworks like lack of visual access, deception, and costume play may enhance the notion that some but not other participants in a conversation share knowledge about the identity of an object that can be distinguished from alternative representations such as appearance.

Finally, there are undoubtedly many non-verbal means, e.g., pointing, gestures, and intonation which may emphasize which aspects of an object or situation are common knowledge. For example, Clark and Wong (2002) have suggested that the intonation of *looks like* prompts preschoolers to consider whether the expression is meant as a hint about the identity of an object or as an alert that the appearance of an object may contrast with the identity.

Of importance for the appearance–reality task is that the correct answer to the question “What does this look like?” relies on the assumption that the identity of the referent is already known. But if this knowledge is unclear to children, their default interpretation may well be to assume that the question means “What do you think this is?” If this is indeed 3-year-olds’ interpretation, then the correct answer “It looks like rock” would be an outright error because it would mean that they think the identity is a rock—and they were just shown that it is a sponge. It would be more accurate to say “This looks like a sponge” meaning “I think it’s a sponge,” which is in fact what 3-year-olds do.

On a typical appearance–reality task, it could be argued, the reality of the object has already been stated as part of the introduction to the deceptive object. The task typically begins by explaining to the child what an object really is and what it looks like, so at first sight it would seem that the identity of the object has already been clearly stated and thus is commonly known when the experimenter asks the test questions. There are, however, reasons to assume that 3-year-olds may not consider the identity common knowledge. First, on the standard task the test questions are separated from the introduction by statements such as “Here’s a different question” or “Now, when you look at this with your eyes.” These interspersed statements may serve to break the connection with the previously communicated reality. Second,

consider the pragmatic implication of asking questions. The implication of asking an honest question is that the answer is unknown by the speaker and perhaps known by the listener. In other words, there is no shared knowledge about whatever is being asked. It may be confusing to 3-year-olds that the presumed common knowledge about the deceptive object is contradicted by the pragmatics of asking questions. In other words, how is the 3-year-old to assume that the speaker knows the reality of the object when the speaker asks, “What is this?” or “What does this look like?” The very act of asking the question implies ignorance. Third, on the typical appearance–reality task the way the deceptiveness of the objects is treated may serve to confuse the children and make the common knowledge unclear. Recall that the experimenter traditionally goes to great lengths to avoid presenting the objects as deceptive even though they are designed to be. Children may be confused about what counts as shared knowledge when speakers avoid saying the objects are tricky or meant to be deceptive. Overall, 3-year-olds’ performance on the traditional tasks may be hindered if the fact that the identity is commonly known is obscured by any of these competing pragmatic factors.

Earlier accounts have emphasized the role of pragmatics in young children’s failure on classic developmental tasks, including the appearance–reality task. Siegal and Peterson (1994) have claimed that the format of such tasks confuses children because it violates the normal conversational rules noted by Grice (1975). They emphasize two problematic aspects of experimental questions: the device of asking questions when the speaker already knows the answer, and repeating questions even though children have already given an answer. If children do not understand that under such academic questioning it is acceptable to state what the experimenter already knows or to repeat an answer, they may interpret the question differently than intended and answer incorrectly, for example by changing the answer they already gave as in a conservation task. As Lillard (1999) has noted, their general account does not gracefully explain young children’s problem on the appearance–reality task. Instead of being led to change their answers on repeated questioning, children tend to give the same answer—e.g., *it is a sponge and it looks like a sponge*. On a different pragmatic account, Deák, Ray, and Brenneman (2003) note that young children tend to perseverate with the same answer in response to different questions in word learning tasks and they propose that children respond in the same way to appearance–reality tasks. They argue that children get thrown because every question offers the same answer options. Until children detect that a new question clearly offers different response options, they will choose one answer and stay with it. Deák et al. claim that children understand the difference between appearance and reality questions so they should get the first question right even if they end up repeating the same answer to the second question. Yet, in the standard appearance–reality task children consistently get the appearance question wrong even when it is asked first. Thus, their account may explain why children perseverate, but it does not explain well why they initially choose a particular answer. While the pragmatics of conversation also play a role in our account, the focus is not on the general demand characteristics of experimental questioning but on the consequences that questions may have specifically for the interpretation of the phrase *looks like*.

In sum, on our account young children's difficulty on the appearance–reality task stems from ambiguity in the meaning of *looks like*. The locution *looks like* can refer to outward appearance (“Peter looks like Paul”) but in fact often refers to likely reality (“It looks like the meeting will be on Saturday.” “That looks like Paul”). We propose that *looks like* is taken to refer to the likely real identity of an object unless the identity is commonly known in the conversation. On the traditional appearance–reality task what is common knowledge may be unclear to 3-year-olds at the time of the test questions for various pragmatic reasons. Once the identity is clearly expressed and thus is shared knowledge at the time of the appearance question, we predict that 3-year-olds will interpret *looks like* as referring to outward appearances.

In Studies 1–4, we test different facets of our account empirically. A prerequisite for our claim that *looks like* is ambiguous to children is that they have two different construals of *looks like* in the first place. It is conceivable that adults restrict the use of *looks like* to outward appearances in their speech to young children. The aim of study 1 was to document that 2- and 3-year-olds have experience with *looks like* in naturally occurring situations as referring to both likely reality and outward appearance as distinct from reality. We searched transcripts from the CHILDES database of conversations between children and adults for occurrences of both these uses of *looks like*.

Because of the ambiguity of *looks like* we claim that its interpretation depends on the context. If this claim is true, then we should be able to clarify the context of the appearance–reality task so that children give correct answers to the appearance question. In Study 2, we mention the real identity in the appearance question and compare it to a standard version. In Study 3, we use the non-verbal context to emphasize that the reality is known at the time the appearance question is asked and again compare this to a standard version. In both studies, we expect children to do better on the common knowledge version than the standard version of the appearance questions. Study 4 extends the results from these studies by asking both reality questions and appearance questions in a manner that allows children to assume that the real identity is common ground. We expect that the common knowledge version of the full task will allow children to demonstrate their appreciation of the appearance–reality distinction.

2. Study 1

The aim of Study 1 was to document that young children have the two widely different construals of the expression *looks like* as referring either to likely reality or surface appearances and thus have reasonable cause to find the appearance question ambiguous on the traditional appearance–reality task. To show this, we would need to find evidence that young children have conversational experience with at least these two different uses of *looks like*. For a source of children's conversations, we drew on the CHILDES database with its transcripts of many hours of conversations between young children and adults (MacWhinney, 2000).

Our first goal was to examine if children are exposed to the use of *looks like* as a likelihood statement about the reality of an object rather than a statement about

outward appearances that differ from reality. By our account, such uses occur when there is no shared knowledge about the identity of an object in a conversation and one speaker expresses an opinion about the identity of the object using the phrase *looks like*. The challenge was to reliably identify contexts where there was no shared knowledge about the identity of the object prior to the use of *looks like*. These transcript data have very little information about the nonverbal context so it can be difficult to infer the level of shared knowledge between the speakers in a conversation, but certain utterances make it clear that at least one of the conversation partners does not know the identity of the object. This occurs when one speaker asks, “What is this?” or similar questions. We thus focused on a very specific type of conversation where one of the speakers had asked about the identity of the referent by asking, “What is this?” or “Who is this?” including their derived forms immediately prior to the utterance that contained the *looks like* phrase. Examples in the transcripts were:

Child: “What’s those in there, Mom?”

Mother: “It looks like cherries.”

Child: “Huh?”

Mother: “It looks like cherries in a cherry pie.”

Child: “I don’t like cherry pie.” (Abe at 2;9, Kuczaj, 1976, 39;123)

Child: “What is this, huh?”

Child: “What?”

Mother: “It looks like a bottle opener.”

Child: “I going to have. . .”

Child: “You want to have some Seven-up, huh?” (Adam at 3;3, Brown, 1973, 27;1525)

The answer with *looks like* to a question about identity as in these examples relates the best estimate on the adult’s part about the likely identity of the object. It is highly unlikely that the adult is trying to comment exclusively on outward appearance and imply that he or she has no idea what the identity of the object is. If this were the case, there would be no relation between the question and the answer which would be a violation of Grice’s maxim of relevance (Grice, 1975). On a Gricean account, the adult is expected to stay relevant to the question asked which in this case would entail an answer that pertains to the identity of the object. We chose this readily identifiable and clear-cut type of conversation to substantiate the claim that children are exposed to *looks like* as referring to likely reality.

The second goal of the study was to establish that preschool children are exposed to a type of sentence construction with *looks like* where the speaker clearly intends the predicate to refer to outward appearances that differ from reality. It is less controversial that *looks like* can refer to outward appearances but we were also interested in identifying a naturalistically occurring construction that makes this unambiguous for use in Study 2. One type of such constructions uses the identity of the object as the noun phrase of the sentence, as in, “X looks like Y.” Examples from the transcripts were:

Father: “Yeah that rock does look like a beehive.” (Abe at 3;10, Kuczaj, 1976, 142;150)

Child: “A penguin looks like a boy.” (Naomi at 2;11, Sachs, 1983, 78;129)

As noted by Austin, the speaker in examples like these clearly means to comment on outward appearances and is not implying that the rock *is* a beehive or that the penguin *is* a boy. The presence of such constructions in the transcripts would corroborate that children have been exposed to the use of *looks like* as referring to outward appearances, and would provide us with a method for asking appearance questions that could not be mistaken for referring to likely reality in Study 2.

2.1. Methods

2.1.1. Participants

The CHILDES database at present contains hundreds of corpora from children recorded under different circumstances and in different languages. We decided to include all corpora that met a list of pre-determined criteria. We selected corpora of normally developing children beginning at the age of 2 years and ending at the age of 3 years and 11 months. These are the years leading up to important developments in children’s performance on the traditional appearance–reality task. The age of two years was chosen as the starting point because children’s speech in transcripts from younger children was often difficult to decipher. We were interested in corpora that were recorded mainly in naturalistic settings and across a variety of different everyday situations where adults and children have conversations. To be able to investigate developmental changes, we included only corpora that were evenly distributed across a recording period of a minimum of 10 months. Since adults may speak differently to children depending on their age, we used corpora where mainly one child was involved in the conversation at a given time. To maximize the likelihood that the expressions used by children and adults would be familiar to us we looked only at corpora in which US English was spoken by native US English speakers. Of the currently available corpora in CHILDES, eight children’s corpora fulfilled the selection criteria: Peter (Bloom, Hood, & Lightbown, 1974; Bloom, Lightbown, & Hood, 1975), Adam and Sarah (Brown, 1973), Shem (Clark, 1978), Trevor (Demetras, 1989), Abe (Kuczaj, 1976), Naomi (Sachs, 1983), and Nina (Suppes, 1974). The number of sessions for each child ranged from 15 to 152. Two children were recorded over a 1-year period and 6 children were recorded over almost the entire age span of 2–4 years. The entire corpora consisted of a total of 333,761 utterances, or 226,629 turns of conversation where a turn consists of a sequence of utterances spoken by a single speaker.

2.1.2. Procedure

The eight selected corpora were searched with the Computerized Language Analysis program CLAN (MacWhinney, 2000) for samples where *look* including its inflections and *like* occurred in the same utterance. This resulted in a

total of 984 samples from all eight children. Each target utterance was printed within a conversational window of 10 utterances before and 10 after the target utterance. We used the following exclusion criteria to derive a collection of samples (number excluded in parentheses): *Look* and *like* were unrelated (32); codes in the transcripts, “xxx”, “yyy”, “[?]”, “+...” indicated that the target utterance or the preceding utterance was unintelligible or separated from each other by a lull in the conversation (61); the target utterance was immediately followed by a change of conversation topic (46); the target utterance was a repetition of the preceding utterance (93; Bartsch & Wellman, 1989; Shatz, Wellman, & Silber, 1983); *looks like* was preceded by negations, e.g., “This doesn’t look like a dog” (23); the predicate of *looks like* was a verb or an adjective (165). The final collection of samples consisted of 478 utterances which contained the phrase *looks like* and a noun predicate.

The target utterance in each sample was always the utterance that contained *looks like* or its inflections. The samples were screened for occurrences of the two types of conversations that we were interested in: (1) As instances of *looks like* referring to reality we counted samples where a question about the identity of an object initiated with “wh-” words immediately preceded the target utterance without any other intervening utterances than generic requests for clarification and repetitions of the question, e.g., “What’s that? It looks like an X.” (2) As instances of *looks like* referring to outward appearances we counted samples where the identity of the object was followed by *looks like* and a noun predicate uttered in the same utterance, e.g., “X looks like a Y”. Personal pronouns were only accepted when they referred to a present person, animal, or character.

2.2. Results

Our first goal was to establish that children are exposed to uses of *looks like* where the predicate concerns the likely reality of an object rather than the outward appearance. The clearest cases of such use was when a speaker had asked about the identity of an object, e.g., “What is that?” and another speaker answered with a sentence with *looks like*, e.g., “It looks like a bottle opener.” Such turns of conversation occurred in 56 of the 478 samples (11.72%). Adults used this construction of *looks like* in 53 cases, and one child, Adam, used it in 3 cases. Seven of the 8 children heard this construction of *looks like* used at least 1 time and a maximum of 20 times, with a median of 7 exposures. The window we chose began at 2 years so any information about first exposure to this turn of phrase must be considered keeping in mind that there might have been earlier exposures. The earliest age that children in the selected corpora heard this construction for the first time was at 2 years and 10 days, the latest age was 3 years and 3 months. The median age of the first exposure was 2 years and 5 months. The one child that used this construction 3 times used it for the first time at 2;8, the second time at 3;4, and the last time at 3;8. The way he used it all three times was rhetorical. He posed himself an identity question and answered it with *looks like*, as in the following example: “What dat? Dat look like plate” (Adam, 2;8, Brown, 1973, 11;1422). These results show that 7 of the 8 children had heard a median of 7 very clear-cut cases where *looks like* was used to refer to the likely identity of an object.

Our second goal was to establish that children are familiar with the sentence construction “X looks like a Y” which is a clear instantiation of *looks like* referring to outward appearances. Of the 478 samples that contained the phrase *looks like*, this construction occurred in 71 samples (14.85%). Adults used it 55 times and children used it 16 times. All eight children heard *looks like* used this way by adults. The earliest age of exposure in our sampling window was at 2 years and 1 months, the latest at 3 years and 3 months. The median age of first exposure was 2 years and 8 months. Children also used this construction actively. Five of the 8 children (2 girls and 2 boys) used this construction in a total of 16 instances, or 22.5% of the occurrences of this sentence type. The first use by a child was at 2;5, the latest age of first use was 3;6. The median age of first use was 2;11. Although the median age of first use was later than first exposure, as one would expect, two children, Adam and Naomi, actually used the construction before there were any instances of exposure in the transcripts.

The two constructions accounted for 27% of the samples. Our impression is that many more occurrences would fall into one of the two categories of outward appearance or likely reality. However, the remaining occurrences in the transcripts are difficult to classify reliably because of the lack of information about context in the CHILDES database. Consider the following example, “This is an X. I’m going over here to play. It looks like a Y.” Without more information about the context, such as a video record that would reveal what the objects are and what the child is looking at, it is impossible to know whether the child is referring to the same object throughout. If she is, then *looks like* refers to outward appearance, if not, then *looks like* might refer to likely reality. Classifying such examples based on transcripts involves much guessing about context which is why we used a completely formulaic counting method in the present study, even if it underestimated the true number of exposures.

Despite the modest numbers, *looks like* is not a rare expression in the eight corpora. Children either heard or said the expression once in every 230 turns of conversation. At this rate, children will have had a substantial number of exposures to either construal before the age of 4. To the extent that the selected corpora from the CHILDES database make up a representative sample, we can document that at a minimum 2 and 3-year-olds have been exposed to the two very clear-cut uses of *looks like* which were unambiguously identifiable in the transcripts. Thus, when asked “What does this look like?” in the standard appearance–reality task, young children may well misconstrue the question.

3. Study 2

To accurately assess young children’s knowledge of the appearance–reality distinction one must ensure that children perceive the appearance questions as they were intended: as referring exclusively to outward appearances. Study 2 accomplished this by making it clear to children that the identity of the deceptive object is common knowledge at the time the appearance question is asked. The CHILDES analysis has shown that in everyday language children hear constructions of the form “X looks like Y.” Because it is a naturally occurring way to highlight that the identity is already

commonly known, we chose this format for the common knowledge condition of the study and changed it into a question, e.g., “What does the sponge look like?” We compared this modified version of the appearance question with the version that has been the standard in the literature, i.e., “What does this look like?” We expected that children would correctly answer with outward appearance in response to the modified question but give many reality answers in response to the standard question.

3.1. *Methods*

3.1.1. *Participants*

Thirty-two children from a preschool on the Stanford University campus participated in the study. There were 14 girls and 18 boys who all spoke English. Children were randomly assigned to one of two conditions with the limitation that gender and age were balanced across conditions. There were two conditions in the study: Common knowledge (mean age: 3;7, range: 2;10 to 4;0), and Standard (mean age: 3;7, range: 3;2 to 4;0).

3.1.2. *Materials*

The materials included four deceptive objects: a sponge that looked like a rock, a book that looked like a pig, an eraser that looked like a crayon, and a candle that looked like an apple. A blue cardboard box was used to keep the objects out of the children’s sight when required by the procedure.

3.1.3. *Procedure*

Each child participated in one session in a separate room at the preschool. The sessions were recorded on videotape and had an introduction phase and a test phase. Children were first introduced to the deceptive objects. The introduction was the same for both conditions. The experimenter produced one object at a time from the box and asked the child, “What does this look like?” In the rare instance where a child did not immediately answer with the apparent identity, the experimenter would state what the object looked like. The experimenter then told the child about the identity of the object, e.g., “It looks like a rock. But it’s really a sponge.” The identity of the object was then explained and demonstrated and the child was encouraged to briefly examine the object. For the rock-sponge, the experimenter explained that it was soft and squishy and that one could wipe things up with it. For the apple-candle, the experimenter explained that it had a wick and that one could light it and blow it out. For the crayon-eraser, the experimenter explained that one could erase pencil markings on a piece of paper with it. For the pig-book, the experimenter explained that one could open the book and read it. In the case of the rock-sponge and the apple-candle, the experimenter demonstrated the identity by pretending to carry out the described actions. For the crayon-eraser and the pig-book, the explanations were illustrated with actual demonstrations. Following each demonstration, the experimenter took the object back and set it aside on the table. After all the objects had been introduced, the experimenter put them back in the empty box and told the child, “Now we’re going to play a game with all these

things. I'm going to take them all and put them in this box and then we are going to see if you can remember what I'm holding in my hand." The deceptive objects were then produced one at a time from the box and the test question asked. The objects were then set aside on the table. The order of the introduction of the objects and the subsequent test questions for each child was specified in a set of eight unique random orders. The set was cycled through twice within each condition.

The aim of the study was to ensure that children interpreted the appearance question as unambiguously referring to outward appearance by emphasizing that the identity of the object in the question was already known, forming the *common knowledge* condition of the test phase. The questions were "What does the sponge look like? What does the candle look like? What does the eraser look like? What does the book look like?" In the *standard* condition of the test phase, children were asked the appearance question as it is typically asked in the literature, i.e., "What does this look like?" In the few cases where children hesitated in either of the conditions, the question was first repeated without answer options, and then with answer options in the format "Does the sponge/this look like a rock or like a sponge?"

3.2. Results

Children's responses to the appearance questions were scored from the videotapes. Almost all children without prompting answered with either the label for appearance or the label for reality that was given in the introduction. A correct answer consisted of answering with the appearance of the deceptive object. A few children used appropriate synonyms for the appearance, such as mountain for rock, which were also scored as correct answers. Because the initial analyses of variance showed no main effects of or interactions with gender and experimenter, these factors were dropped in the presented analyses. The data are presented in Table 1.

The crucial question in this study was if making the common knowledge of the identity explicit in the appearance question improved children's performance over the standard way that appearance questions have been asked. As predicted, children did significantly better in the common knowledge condition, $t(30) = 3.78, p < .001$. In this condition, children gave an average of 3.63 correct answers out of 4 possible (91%). In the standard condition, children gave an average of 1.88 correct answers of 4 possible (47%). Compared to chance, which was 2 correct answers out of 4, children performed significantly better in the common knowledge condition, $t(15) = 8.06, p < .001$. The standard condition did not differ from chance.

Table 1
Mean percent correct answers to appearance questions in Studies 2 and 3

Condition	Study 2		Study 3	
	Mean	SD	Mean	SD
Common knowledge	91	20	85	30
Standard	47	42	50	42

Thus, asking the appearance question in an unambiguous manner enabled 3-year-olds to reveal their knowledge of the outward appearances of deceptive objects. They readily answered near ceiling levels of performance in marked contrast to the typical failure to answer appearance–reality questions.

One alternative explanation for these results might be raised: because there are only two salient representations of each of the deceptive objects, children may reason that since the experimenter explicitly gave one of the possible two answers as part of the question, the expected answer must be the other available option. If children were using such an “opposites” strategy, we believe they would also be using it on the traditional task in the literature. On that task, children give one answer to the first question, then hear, “Here’s a different question for you,” and are then asked the second question. On an “opposites” argument, this frame seems to call for opposite answers as much as or more than the procedure in Study 2. Nevertheless, to rule out the alternative explanation, the appearance question must be asked in ways that do not explicitly state one of the two answer options as part of the question. This was part of the motivation for Study 3.

4. Study 3

In the previous study, the question “What does the sponge look like?” enabled 3-year-olds to readily answer the appearance question presumably by emphasizing in the question that the identity was already common knowledge. In everyday conversations, however, the common knowledge may not always be expressed verbally. The first goal of Study 3 was to examine if the non-verbal context can be arranged to clearly signal that the identity of an object is already known. The second goal was to avoid giving children the opportunity to use the verbally stated identity as the basis for an alternative response strategy, as could have been the case in Study 2. To accomplish these goals, we manipulated the non-verbal context in which the objects was presented. A deceptive object such as a candle that looks like an apple can be presented along with other candles with various deceptive appearances, such as a candle that looks like a dump truck and a candle that looks like a butterfly, or it can be presented along with random deceptive objects like a rock-sponge and a crayon-eraser. Deliberately grouping objects with the common theme of candles with different appearances should highlight the common ground that the identity of the candles is known. We expected the emphasis on the common knowledge of the identity to lead children to interpret the appearance question as referring to outward appearances that differed from reality. In contrast, grouping deceptive objects without any apparent commonalities should leave the common ground less clear. Without clearly shared knowledge of the identity, as on the standard task in the literature, we expected many children to interpret the question of what the object looked like as referring to its identity. Thus, the simple grouping of objects at the time of the appearance question should lead to differences in children’s ability to reveal their knowledge of the outward appearances.

4.1. Methods

4.1.1. Participants

Sixteen children from a preschool on the Stanford University campus participated in the study. There were 9 girls and 7 boys who all spoke English (mean age: 3;4, range: 2;11 to 3;11). Children participated in both of the two conditions in the study: Same identity and different identity. They were randomly assigned to one order of conditions with the limitation that gender and age were balanced across orders.

4.1.2. Materials

To ensure that children saw a variety of objects, we created two different sets of 6 objects which all had different appearances but whose identities differed from each other: a candle set and an eraser set. The base of the sets was the same four deceptive objects used in the previous studies: a sponge that looked like a rock, a book that looked like a pig, an eraser that looked like a crayon, and a candle that looked like an apple. These objects were supplemented with 2 objects for each larger set. Two candles that looked like a dump truck and a butterfly, respectively, for the candle set, and 2 erasers that looked like a cat and a dinosaur, respectively, for the eraser set. A blue cardboard box was used to keep the objects out of the children's sight when required by the procedure.

4.1.3. Procedure

Each child was seen individually in a separate room at the preschool. There was one session with two conditions per child: Same identity and different identity. The sessions were recorded on videotape and had an introduction phase and a test phase. Children were introduced to one of the two sets of six deceptive objects. The procedure was the same as in Study 2. The experimenter produced one object at a time from the box and said, e.g., "What does this look like? It looks like a rock. But it's really a sponge." The identity of the object was then explained and demonstrated and the child was encouraged to briefly examine the object. The additional two candles and two erasers were explained and demonstrated like the candle and the eraser of the original set.

In the test phase, children were asked one appearance question per object. Out of the larger set of six objects in the box, the experimenter drew two subsets of 3, one subset for each condition. The key manipulation of the study was whether objects were grouped with objects of the same identity in the *same identity* condition, or with objects of different identity in the *different identity* condition. For example, in the case of the eraser set, children were first shown a subset of erasers that looked like a crayon, a cat, and a dinosaur. Next they were shown a subset of a rock-sponge, a pig-book, and an apple-candle. The order of conditions was counterbalanced across children. For each condition, the experimenter pointed to each of the three objects in the subset one at a time and asked, "What does this look like?" Piloting had shown that most children would give their answers immediately, but if this was not the case, the question was first repeated, and second repeated with answer

options. The order of the introduction of the objects and the subsequent test questions was specified in a set of eight lists of unique random orders with the provision that two objects with the same identity never occurred back to back in the introduction phase. The set was cycled through twice.

4.2. Results

Children's responses were scored from the videotapes. With few exceptions, children's responses consisted of either the reality or the appearance of the deceptive objects as they were labeled during the introduction. Answers were counted as correct if the child mentioned the appearance of the deceptive objects or appropriate synonyms. Two children in each condition spontaneously gave compound nouns as answers, e.g., "Dump truck-candle. Apple-candle. Butterfly-candle." Because these answers expressed awareness of the different appearances for each object, they were counted as correct. The initial analyses of variance showed no main effects of or interactions with gender, order of conditions, and set of objects so these factors were dropped in the presented analyses.

The question of interest was whether the context in which the objects were presented made children interpret the appearance question differently. Table 1 presents the results. As predicted, in direct response to the appearance question "What does this look like?" children changed their answers depending on the context. Children gave significantly more correct answers when the non-verbal context highlighted the shared identity of the objects, paired $t(15) = 3.06, p < .01$. In the same identity condition, children gave an average of 2.56 correct answers out of 3 (85%), compared with 1.50 out of 3 in the different identity condition (50%). Thus, when the non-verbal context emphasized shared knowledge of the identity of a deceptive object it thereby freed young children to correctly treat *looks like* as referring to the outward appearance of the object. The nonverbal way of expressing that the identity was commonly known precluded the alternative verbal response strategy discussed in Study 2.

We also compared children's answers to chance. Chance performance was equal to 1.5 correct answers out of 3. Children in the same identity condition did significantly better than chance, at 85% correct, $t(15) = 4.76, p < .001$. Children's performance in the different identity condition was exactly at chance, at 50% correct.

Although most of the objects by virtue of the design were used in either the same identity condition or the different identity condition, the crayon-eraser and the apple-candle occurred in both conditions, e.g., half of the children saw the crayon-eraser in the same identity condition, the other half in the different identity condition. This allowed for a direct comparison of how children responded to the same item in the two different contexts. The correct responses to these two objects were analyzed by condition with an independent samples t test. The analysis showed a significant difference between the two conditions, $t(30) = 2.03, p < .05$, and thus mirrored the results obtained with the larger sets above. Children gave 88% correct responses in the same identity condition and 56% correct responses in the different identity con-

dition. Thus, the results held up when responses to the same objects in the two different contexts were compared.

5. Study 4

One very strong source of evidence for children understanding the distinction between appearances and reality would come from their ability to articulate the distinction on their own. A casual observation motivated the method of Study 4. Despite just having failed a standard appearance reality task, one child went on to spontaneously explain to his teacher that he had seen a sponge that looked like a rock. The child's comment was surprising because constructing a full-sentence comment is seemingly more difficult than answering the piecemeal questions of the standard task, which are intended to make the task easier for children. One explanation is that the single sentence makes the contrast between appearances and reality clear. We decided to capitalize on this observation and ask children to describe both the appearance and reality of the deceptive objects in a single sentence. Note that a contrastive question reinforces that the reality of the object is known by the speaker, thus making clear that *looks like* is intended to refer to outward appearances.

In other studies, 3-year-olds on their own have been shown to make statements that contrast mental states and reality (Bartsch & Wellman, 1989; Shatz et al., 1983). An example of a thought-reality contrastive could be "the people thought Dracula was mean, but he was nice" (Child of 3;3, quoted in Bartsch & Wellman, p. 56). Such contrastives are considered particularly convincing evidence that children understand mental representation because the explicit contrast of thoughts with reality makes it clear that children consider thoughts and beliefs as true mental conceptions (Bartsch & Wellman, 1989; Perner, 1991, note 8.2; Shatz et al., 1983). On the same logic, contrastives between appearance and reality would show that children consider representations of appearance to be separate from reality.

Children were asked to tell a puppet about the deceptive objects in one question, "Can you tell Ernie what this looks like and what it really is." An example of a correct answer would be, "This looks like a rock but it's really a sponge." We compared their performance to a standard version of the appearance–reality task where we asked, "What does this look like" and "What is this really and truly" in two separate questions. Three-year-olds were expected to do better in the contrastives condition than in the standard condition.

5.1. Methods

5.1.1. Participants

Sixteen children from a preschool on the Stanford University campus participated in the study. There were 7 girls and 9 boys who all spoke English (mean age: 3;7, range: 3;4–3;9). Children participated in both of the two conditions in the study: Contrastive and Standard. They were randomly assigned to one order of conditions with the limitation that gender and age were balanced across orders.

5.2. Materials

Four deceptive objects: a sponge that looked like a rock, a book that looked like a pig, an eraser that looked like a crayon, and a candle that looked like an apple. A blue cardboard box was used to keep the objects out of the children's sight when required by the procedure. Children described the deceptive objects to a puppet of the Ernie character from Sesame Street which had a Band-Aid on one knee.

5.2.1. Procedure

Our intention was to compare children's performance on the exact same 4 objects with the contrastive or standard framing being the only difference between the conditions. Since it would be tedious for children to go through the same set of objects twice in a row, children participated in two sessions 2–5 days apart with one condition in each: contrastive or standard. The sessions were conducted by the same experimenter in a separate room at the preschool and were recorded on videotape. Regardless of which condition came first, children were introduced to the deceptive objects in detail on their first visit. The procedure for explanations and the demonstrations of each object was the same as in Study 2, e.g., "What does this look like? It looks like a rock but it's really a sponge." After all 4 objects had been introduced, the experimenter put the objects back in the box while reminding children verbally about each object, e.g., "This looks like a rock but it's really a sponge." On their second visit, the detailed introduction to the object was skipped. Instead the experimenter put all four objects on the table and then gave children the reminder about each object as they were put back in the box.

The test phase differed in the two conditions. In the *Contrastives* condition, children were told that they were going to tell a puppet about the objects in the box. The experimenter then introduced the Ernie puppet, pointed to the Band-Aid on the puppet's leg, explained that Ernie was feeling sad, and that it was the child's task to cheer Ernie up by telling him about the objects in the box. The experimenter then placed Ernie on a chair and moved to the child's side of the table. The experimenter asked Ernie if he was ready and then reached over and made Ernie's mouth move while saying a muffled, "Yes, I'm ready." The experimenter then prompted the child to tell Ernie about the first object. Half of the children heard, "Can you tell Ernie what this looks like and what it really is?" The other half of the children heard the opposite order of mention of appearance and reality: "Can you tell Ernie what this really is and what it looks like." After each trial the experimenter reached over and moved the puppet's mouth and said a muffled, "Thank you." If a child hesitated to answer, the question was repeated. At the end of the procedure, the experimenter told the child that Ernie was much happier now that the child had told him about the objects.

In the *Standard* condition children were asked the test questions in a way that mirrored the standard task in the literature (e.g., Flavell et al., 1986). They were seated across from the experimenter and told, "Now, here's a question for you. What does this look like?" When they had answered, they were told, "Here's a different question. What it is this really and truly?" If a child hesitated to answer one of the questions, the question was first repeated and second repeated with answer options.

The order of the introduction of the objects, and the subsequent test questions was specified in a set of 8 unique random orders. The set was cycled through twice within each condition. The order of condition was counterbalanced across children. The order of mention of appearance and reality within the reminder was counterbalanced across children, so some children would hear appearance first and then reality, e.g., “This looks like a rock but it’s really a sponge,” and other children would hear reality first and then appearance, e.g., “This is really a sponge but it looks like a rock.” This was done so that the order of mention of appearance and reality in the reminder matched the order in the test questions which were counterbalanced in the same manner. Although the order of mention of appearance and reality within the reminder and the test questions was always the same for all four objects for each child, it was crossed between conditions. Half of the children that received a given order of mention on the first visit received the opposite order on their second visit. The other half received the same order on both visits.

5.3. Results

Children’s responses to the contrastive questions and to the separately asked appearance and reality questions were scored from the videotapes. The format of the questions in the contrastive condition invited children to give more of a free-format answer. The most frequent types were, “Hey Ernie, this looks like a rock but it’s really a sponge,” or “This really looks like a rock but it really is a sponge.” In 2 cases children would say, “This looks like a rock and a sponge.” Two children answered with compound nouns, e.g., “This is an apple-candle.” Although it could be argued that these examples imply knowledge that there are two representations of the same object, we chose a conservative criterion for what counted as a correct answer. In addition, the strict criteria were intended to rule out that children simply mentioned two different labels in response to a question that called for two labels. A correct response consisted of explicitly contrasting the two labels by using the phrases “looks like” or “really” correctly with either or both labels. The following examples were counted as correct answers, “This looks like a rock but it’s really a sponge,” “This looks like a rock but it’s a sponge,” “It’s a pig but it’s really a book.” In the standard condition, children were asked the test questions as in the appearance–reality literature with two separate questions for appearance and reality. A correct response consisted of correct answers to both questions. However, if children spontaneously provided the full contrast as in the contrastive condition in response to the first question this was counted as a correct answer and the experimenter refrained from asking the second question.

The main question in this study was whether children were able to demonstrate knowledge of the appearance–reality distinction when they were asked to produce a full contrastive statement compared to the separate questions in the standard task. Because the initial data analyses showed no main effects of or interactions with gender and the order of the appearance and reality *within* a question, these factors were dropped in the presented analyses. Table 2 presents the results. The results were analyzed with a Condition (Contrastive vs. Standard) \times Order (Contrastive First vs. Standard First) ANOVA, with condition as a within-subjects variable. There was

Table 2

Mean percent correct answers to appearance–reality questions by order in Study 4

Condition	Order			
	Standard first		Contrastive first	
	Mean	<i>SD</i>	Mean	<i>SD</i>
Contrastive	88	35	81	35
Standard	38	40	78	37

a significant effect of condition, $F(1,14) = 8.19$, $p < .01$. There was no main effect of order, $F(1,14) = 1.18$, but there was a Condition \times Order interaction, $F(1,14) = 6.38$, $p < .02$. The interaction with order was that the results in the standard condition were influenced by whether children had already participated in the contrastive condition or not. If the standard condition was children's first task, they gave fewer correct answers than if it came second and they had already participated in the contrastive condition, $t(14) = 2.12$, $p < .05$. In the contrastive condition, the results were not influenced by the order of the conditions.

We first examined children's performance completely unbiased by order of participation. This allowed for the most direct comparison with previous studies in the literature. We used a between-subjects test to compare children's performance on their first task only. Eight of the children participated in the standard task first and gave 1.5 (38%) correct answers. The performance on the standard task mirrors that found in the literature. The other 8 children participated in the contrastive task first and gave 3.25 (81%) correct answers. This was a significant difference, $t(14) = 2.33$, $p < .04$. Thus, as predicted asking a contrastive test question led to superior performance relative to asking the test questions separately.

The Order \times Condition interaction was examined first in the order where children participated in the standard condition first and the contrastive condition second. When children were asked to explicitly contrast reality with appearance they gave an average of 3.5 correct answers (88%) compared with 1.5 correct answers (38%) on the standard version of the task, revealing the predicted improvement of the contrastive question over the standard questions, paired- $t(7) = 2.83$, $p < .03$. The pattern was different, however, when children participated in the contrastive condition first. Their correct responses to the contrastive question remained high at 3.25 (81%), but now they also gave almost as many correct answers to the standard task, at 3.12 (78%). For the standard task this represents an improvement of 40% between the giving it as the first and second task. Children's individual responses reveal why this might be. When the standard task followed the contrastive task, 5 of the 8 children spontaneously provided a total of 16 contrastive answers rather than the separate answers to appearance or reality that they had been asked for. In contrast, when the standard task came first only one child spontaneously provided two contrastive answers. The difference of 14 contrastive answers between the two tasks (a nearly 44% difference between the total of possible answers in each task) can account for the difference between the standard tasks when given first and second. This indicates that having answered the contrastive question first led children to understand the

experimenter's question in a contrastive way even on the separate questions on the standard task.

Even though there were significant differences between the conditions, it is important to know whether children did better than chance in the contrastives condition. Because one correct response required children to give two correct answers per object, chance was considered 25%, or 1 out of 4 responses. While children in the contrastive conditions responded above chance levels regardless of order, at 88% correct when given first, $t(7) = 5.00, p < .005$ and at 81% correct when given last, $t(7) = 4.58, p < .005$, children's responses in the standard condition varied depending on the order. When the standard condition was given first children's responses did not differ from chance at 38% correct, $t(7) = .89, p < .41$. When the standard condition was given in the second session, children's responses were above chance at 78% correct, $t(7) = 4.12, p < .005$.

It could be argued that children's performance in this study was susceptible to the task demand that children were asked for two answers. But if children had no understanding of the appearance–reality distinction, they should have generated the two answers randomly and their performance should have been close to chance in the contrastive condition, which was not the case. A related possible task demand is that children were simply imitating the way the objects had been introduced by the experimenter. In response to this possibility, we note that children did not seem constrained by the way the objects were introduced nor by the order within the question, which they would have if this possibility were at play. Rather, in some cases children flexibly reversed their answers relative to what they previously heard, so in response to the question, “Can you tell Ernie what this really is and what it looks like”, they answered, “This looks like a rock but it's really a sponge.”

To compare children's responses in this study with previous studies in the literature, we looked at their performance in terms of the traditional set criterion for passing the task of 3 or more correct out of 4. The data were analyzed with a categorical repeated measures ANOVA using the SAS *CATMOD* procedure. A Condition (Contrastive vs. Standard) \times Order (Standard First vs. Contrastive First) ANOVA with condition as a repeated measure was conducted to examine whether there were differences in the frequency with which children passed the tasks. There was a significant main effect of condition, $\chi^2(1, N = 16) = 8.70, p < .005$. There was no main effect of order, $\chi^2(1, N = 16) = 1.22, p < .27$. Order interacted with condition, although slightly less strongly than in the previous ANOVA, $\chi^2(1, N = 16) = 3.13, p < .08$. Overall, by this criterion 14 of 16 children (87%) passed the appearance–reality task in the contrastive condition, 7 of 8 children in each order of presentation. Overall, 9 of 16 children (56%) passed the standard task, but as indicated by the interaction the number of children who passed varied by condition. When the standard condition was given first, only 3 of 8 children (37%) passed. When the standard condition was given second, 6 of 8 children (75%) passed.

Finally, we note that the three youngest children in this study were 3 years and 4 months old. Given that it should be harder to answer two questions at once rather than one at a time and that generating a complex contrastive statement on request is no small matter, it is noteworthy that they all gave correct responses in the contrastive condition.

5.4. Discussion

On our account young children's problems on the traditional appearance–reality task arise because of ambiguity in the appearance question, “What does this look like?” The ambiguity stems from a subtlety in the expression *looks like*. Generally, if the reality of an object is already commonly known among the speakers in a conversation, (“Jack looks like Paul,”) then the predicate of *looks like* is taken to refer to surface appearances. If there is no shared knowledge of reality, then *looks like* is taken to refer to likely reality (“That looks like Jim”). We have claimed that the common knowledge of reality is unclear to 3-year-olds on the standard appearance–reality task which leads them to fail the task.

A prerequisite for our claim that *looks like* is ambiguous to children is that they have experienced the two uses in the first place. In Study 1 we examined extensive corpora from the CHILDES database of 8 children's conversations with adults when they were 2 and 3 years old. Conversations where *looks like* referred to likely reality constituted roughly 12% of our samples. Constructions where the predicate clearly refers to surface appearances occurred in roughly 15% of the samples. Thus, 3-year-olds are likely to have two different construals of the expression.

The remaining three studies investigated 3-year-olds' answers to appearance questions when the common knowledge of reality was emphasized. One prediction is that 3-year-olds will answer appearance questions correctly when the immediate verbal context clearly expresses that reality can be assumed to be known, e.g., “What does the sponge look like?” When asked about appearances in this format, children gave 91% correct answers to the appearance question compared to 47% when asked in the traditional manner with “What does this look like?” It should also be possible to manipulate the nonverbal context such that the reality of the deceptive objects can be assumed to be common ground at the time of the appearance question. In Study 3, this was accomplished by presenting children with three deceptive objects of the same hidden identity but with different surface appearances, e.g., three erasers that looked like a crayon, cat, and dinosaur. When the experimenter singled the objects out one at a time and asked what they looked like, children gave 85% correct answers, compared to a standard version of the task where they answered randomly, at 50% correct answers.

Because our account was centered on the discourse surrounding the expression *looks like*, Studies 2 and 3 had looked only at the appearance question of the appearance–reality task. Eliminating the ambiguity of *looks like* should also lead 3-year-olds to do better on a common knowledge version of the full appearance–reality task relative to a standard version of the task. In Study 4, children were asked to contrast the appearances with the reality of the deceptive objects on their own, e.g., “Can you tell Ernie what this looks like and what it really is.” Asking for such a contrast is another way to emphasize that the speaker knows about the real identity that differs from appearances. Even though this required children to generate a complicated verbal response on their own, children still gave more correct answers, at an average of 85%, to the contrastive questions than to the standard version of the task where they were asked separate questions for appearances and reality. These results were

moderated by an effect of order of presentation. Children's good performance on the contrastive condition was unaffected by order, but on the standard task children spontaneously gave full contrastive answers to the separate questions when they had first participated in the contrastive condition.

In sum, the evidence supported the key components of our account of children's difficulties on the traditional appearance–reality task. In assessing the results it is important to rule out that features of the three tasks might allow children to pass the tasks by other means than understanding the difference between appearances and reality. On the original appearance–reality task children are shown a deceptive object, such as a sponge that looks like rock, questioned about its appearance, and then shown and told explicitly that while the object looks like a sponge it really and truly is a rock. Children are then asked two separate questions: One question which states, “Now, here's a question for you. What it is this really and truly?” And another question which states, “Here's a different question. What does this look like (to your eyes right now)?” One reason why the results of the standard task are so striking is that children do not take any of the many opportunities to use alternative response strategies. They are literally given the correct answer immediately before the test questions and they would only have to repeat the correct answer. They are also told that they will be given a *different* question which invites simply giving the opposite answer of what they previously said. Yet children do not use either of these strategies on the standard task. Overall, our modifications to the standard task were so minor that we doubt that they made alternative response strategies available that were not present in the standard task to begin with. In Study 2 we simply reworded the appearance question to include reality, e.g., “What does the sponge look like?” compared with “What does this look like?” If an opposites strategy was at the tip of children's tongue in our modified task, we would expect them to use this strategy in many cases on the standard task, but an opposites answer is neither a typical correct answer nor a typical error on the standard task (e.g., Deák et al., 2003). In Study 4, our modification consisted of simply combining the two separate questions of the standard task into one question, e.g., “Can you tell Ernie what this looks like and what it really is?” Children might here reason that they were asked for two answers and so feel compelled to randomly combine two different responses. However, children did not mindlessly give two different responses, but flexibly and appropriately juggled the way they combined the answers with the phrases *looks like* and *really is* into complete sentences. Regardless of the plausibility of the opposites strategy, part of the motivation for Study 3 was to empirically eliminate any information in the question as a basis for an alternative strategy by nonverbally emphasizing the common knowledge of reality. When an alternative response strategy was eliminated children nevertheless succeeded, supporting our overall conclusion that children's success derives from understanding the appearance question in the correct context.

It is also important to assess the strength of the results. In a large meta-analysis of hundreds of conditions from false belief and appearance–reality studies, Wellman, Cross, and Watson, 2001 note that on the handful of manipulations that led to a lower age of correct performance, children went from below-chance performance to chance performance. It is difficult to interpret whether chance performance dem-

onstrates a better grasp of the concept of interest relative to a more difficult task or whether it reflects that children are guessing. To truly demonstrate that they understand the concept being studied, children must perform above chance. In all of our three studies, children performed well above chance in the common knowledge conditions.

A related indicator to examine is the breadth of the findings. Compared to the existing range of standard tasks, it is possible that a modified task reveals incipient knowledge that may be brought out by one particularly easy task rather than reveal robust earlier knowledge. We have examined 3-year-olds' knowledge of the appearance–reality distinction with three different types of manipulations that were motivated by a coherent theoretical framework to emphasize the common knowledge of reality at the time of the test questions. In all cases we found near-ceiling performance. Thus, children's performance in our studies was robust which precludes the possibility of having stumbled upon one task modification that succeeded in bringing out incipient knowledge. The strength and breadth of the findings from Studies 2 to 4 provide support for our account that emphasizing that both experimenter and children already know reality allows children to demonstrate knowledge of the appearance–reality distinction on the standard task.

We turn now to show how this discourse-based explanation can account for successes and failures of other modifications in the literature on the appearance–reality task. A number of training studies have failed to improve the performance of 3-year-olds whether the training consisted of helping them distinguish between real and apparent color (Flavell et al., 1986), better understand the expressions “looks like” and “really is” (Taylor & Hort, 1990), or provide corrective feedback (Braine & Shanks, 1965a, 1965b). On our view, the training and the feedback take place well before the appearance question is asked and thus do not serve to make the common knowledge of reality any clearer in immediate connection with the appearance question. Thus children are unlikely to be helped by this kind of training or feedback.

A handful of studies have led to success because of modifications to the original task. One successful manipulation involved setting up the appearance–reality task with the goal of deceiving a second experimenter (Rice et al., 1997). On our account, the frame of deception includes a very strong emphasis on a jointly known reality between two conversation partners to which a third partner is not privy. Thus deception may enhance the fact that reality is commonly known and help carry this fact over to the test question more clearly than on the traditional task.

A second successful manipulation has been to couch the task in terms of pretense, for example pretending that the rock-sponge is a rock and then asking if the rock-sponge is a pretend-rock or a real rock at the time of test (Flavell et al., 1986, Study 1; Flavell, Flavell, & Green, 1987; Woolley & Wellman, 1990). Note that the phrase *looks like* was not used on this type of task, but we still believe that our account is relevant. The concept of joint pretense reinforces that there is a commonly known reality which is mutually suspended as a function of the pretense act. Thus, a pretense framing of the appearance–reality task on our account serves to clarify to children that the experimenter intends to refer to the surface appearance as distinct from reality when asking the test question.

In a third successful manipulation, Rice et al. (1997) presented a deceptive object along with two objects that corresponded to the appearance and the real identity, for example a rock-sponge along with a real rock and a real sponge (Rice et al., 1997). The idea was that the two real objects would reduce the information processing load by serving as reminders of the two possible representations of the deceptive object. This improved 3-year-olds' performance relative to the standard task. On our account, the presence of the real object together with the deceptive object is parallel to the Same identity condition in our Study 3. When two sponges with different appearances are present and the experimenter singles one out and asks what it looks like, the child takes the identity of the objects as already known and thus interprets *looks like* to refer to surface appearances.

In a fourth successful manipulation, Sapp and colleagues (2000) showed 3-year-old children an array of deceptive and non-deceptive objects, for example a sponge that looked like a rock presented along with a real rock among other objects. For the appearance question, children were requested twice to pick an object so that a teddy bear could be photographed with something that looked like a rock. For the reality question children were requested twice to give the experimenter "something to wipe off dirt with." A majority of children picked the real rock in response to the first appearance request and the rock-sponge in response to the second appearance request. For the reality questions, a majority of children picked the real sponge in response to the first reality request and the rock-sponge in response to the second. Children's response to the appearance and reality questions was scored as correct if they picked the appropriate object as either their first or second choice. This yielded a high overall passing rate relative to a standard task and is evidence that children do not indiscriminately map two labels on to the deceptive objects but clearly know that one refers to the surface and the other to the identity. The authors attributed the improvement to the nonverbal nature of the response. Framing the task in terms of taking a photograph does not make clear that reality is commonly known so we could not account for the finding in this way. However, there may well be other ways that the context can make it clear to children that the experimenter is asking for appearances per se. In so far as children understand that photographs are a type of pictorial representation of surfaces, requesting to take a photograph of an object makes it clear that the context is about surface appearance and not hidden reality. Similarly, if children were asked to draw a picture of a rock-sponge, they would presumably try to draw a rock and not a sponge.

The focus of the present work is on the object identity version of the appearance–reality task, but other types of tasks exist, most notably tasks involving object properties such as apparent color, shape and size. An example of a typical properties task could be showing children a glass of white milk placed behind a red filter and asking about its apparent and real color (Flavell et al., 1986). There is a striking difference in the pattern of children's errors between the object identity and object properties types of tasks. Whereas 3-year-olds make predominantly realism errors on the object identity task, saying that the rock-sponge *is* a sponge and *looks like* a sponge, they make mainly phenomism errors on the properties task, i.e., by insisting that the milk *looks* red and *is* red. The point of departure of our analysis of the object identity task

was the fact that children make realism errors, which means that their failure stems from giving reality answers to appearance questions. Thus the problem must be in children's understanding of the appearance question which contains the phrase *looks like*. Applying the same logic to the object properties task, children here make phenomism errors, which means that their failure stems from giving appearance answers to the reality questions. Thus, the problem must be in children's understanding of the reality question which contains the phrase *really is*. An account of the subtleties of the phrase *really is* and its role in the object properties task does not fall out of our analysis. Analyses of children's errors in studies of real and apparent moral character (Flavell et al., 1992) and real and apparent emotion (Gross & Harris, 1988) showed that realism and phenomism errors were about equally common. Therefore our account may not readily apply to appearance–reality studies in the area of social understanding. Any conclusions about children's appearance–reality abilities in general must be tempered accordingly.

The appearance–reality task has been conducted in diverse cultures and languages with the same results, suggesting a certain universality to the development of the appearance–reality distinction, e.g., China (Flavell, Flavell et al., 1983), Germany (Sodian et al., 1998), and France (Melot & Houdé, 1998). We consider it most likely that the equivalents of *looks like* on the appearance–reality task in other languages—certainly Mandarin, German, and French—have the potential for confusion between likely reality and surface appearance just as in English, and that some version of the pragmatic principles we have outlined are needed to avoid misunderstandings. Although it is an empirical question, we would initially expect that our modifications to the appearance–reality task would allow 3-year-olds to pass the task in other languages. We cannot rule out, however, that some languages may have expressions that uniquely refer to surface appearances without any implication whatsoever that this is also the likely identity. If such expressions existed, they might make the meaning of the appearance question immediately clear to 3-year-olds by eliminating the reliance on pragmatic devices that makes the task difficult in English.

On a broader level, it has been noted that a number of tasks used for studying children's understanding of false beliefs show nearly the same developmental sequence as the object identity version of the appearance–reality task (see Wellman et al.'s (2001) meta-analysis for a comprehensive list). Can our account explain 3-year-olds' failure on these tasks? On a typical false belief task children are asked to comment on a character's false belief about the location of an object relative to what the child and the experimenter know to be the true location. Children are then queried with questions that include references to the mental state of the character, such as "What does she think" or "What did you think," in other words third-person or past-tense references to mental states. Our account may be extended to other mental state references than *looks like*. As was the case for the expression *looks like* on the appearance–reality task, the expression *thinks* may also be subject to a discourse structure that without context makes it ambiguous whether it refers to likely reality or a false mental state and thus whether it should be added to the common ground of the conversation as true or false. The necessity to mark true from false contributions to conversations may be what compels children to spontaneously

add a contrast with reality when reporting on third persons' thoughts as in, "the people thought Dracula was mean, but he was nice" (Bartsch & Wellman, 1989, p. 56). A complete treatment of the false belief task will bring us too far a field, but the construal of *thinks* and other references to mental states may be as determined by the epistemic context as we have shown to be the case for *looks like*.

With the exception of the object properties and social understanding tasks, then, our account can explain the pattern of results of the training studies and modifications to the appearance–reality task in the existing literature. What still remains to be accounted for is the development that enables 4½-year-olds and not 3-year-olds to master the standard task in the literature, if their mastery does not result from conceptual development of the appearance–reality distinction. The point of departure of our framework was the ambiguity of the expression *looks like* which requires shared knowledge of reality to refer to outward appearances. Why would the pragmatic information of shared knowledge of reality on the standard task be clear to older children but not to 3-year-olds? We have suggested that the likely cause of 3-year-olds' problem is competing pragmatic information in the setup of the standard task. As mentioned in the introduction, there may be at least three sources of competing pragmatic information: (1) The experimenter's emphasis that the question is new or "different" obscures the connection to previous events; (2) The pragmatics of asking a question imply ignorance on the part of the speaker; (3) Deliberately ignoring the deceptiveness of an object muddles what is jointly known about reality. From younger children's point of view, these three factors alone or in combination are tantamount to the experimenter asking a fresh question, breaking the connection with previously provided pragmatic information. Older children, however, may realize that previously shared knowledge is not negated by asking a new question. So whereas younger children require that pragmatics such as the commonly known reality must be very clear locally, older children realize that this knowledge is sustained throughout the conversation and even dominates subsequent infelicitous pragmatic moves.

Two relevant sources of evidence speak to young children's ability to sustain preceding pragmatic information when presented with ambiguous statements. Avrutin and Coopmans (2000) examined how 3- to 5-year-old Dutch, Russian, and English-speaking children made bridging inferences to link ambiguous statements about pictures to preceding sentences. In one example, they showed children a picture of a boy with red pants eating and a girl with green pants playing. A puppet would then utter the statement, "There is a boy eating. The pants are green." Children's task was to judge whether this was a true or false statement. Presumably, adults would judge that since the boy's pants were actually red, the puppet's statement would be false. Adults would be inferring that the focus on the referent of the initial sentence, the boy, was already in the common ground. Consequently, they would bridge the definite reference to the pants back to the boy. Three-year-olds were at chance at identifying such statements as false whereas older children called most of such statements as false. In subsequent studies, the authors found that younger children's judgments were easily changed by increasing the salience of the referent. When the topic of the initial phrase was moved to the

beginning of the sentence, e.g., “A boy is eating. The pants are green,” 3-year-olds identified the statements as false most of the time.

A related set of studies investigated if the implicit pragmatic suggestions of the discourse setting would influence children’s responses to subsequent ambiguous questions (Allen, 1991; Bacharach & Luszcz, 1979; Luszcz & Bacharach, 1983; Shatz, 1978; see also Haryu, 1991, related in Haryu & Imai, 1999). In a representative study, Bacharach and Luszcz (1979) showed 3 to 5-year-olds pictures that allowed children to refer to either an action or an object in the picture. Each ambiguous request for comments was prefaced by discourse that implicitly suggested that the topic of the conversation was either the action or the object, e.g., “Horses can run. Horses can jump fences. Horses can eat hay. Here’s a picture, tell me about it” or, “A cow is an animal. A rabbit is an animal. A cat is an animal. Here’s a picture, tell me about it.” Children’s answers showed that 5-year-olds, but not 3-year-olds, matched the action setting with action answers and the information setting with information answers. While the studies of referential ambiguity are not direct analogues to our construal of the standard appearance–reality task, the results give reason to think that younger children do in fact struggle with making the required bridging inferences to implicit pragmatic information which older children more readily accomplish.

The specific developmental change that enables the improved pragmatic abilities in children above $4\frac{1}{2}$ years is difficult to formulate in exact terms at this point. The change could be described in theory of mind terms in that pragmatics is a form of theory of mind knowledge. Children need to understand: how and when to connect current and preceding discourse, the rhetorical devices involved in asking questions, and which information is decisive when different pragmatic principles conflict. This pragmatic development may also be aided by more basic cognitive achievements such as general information processing abilities, including the ability to inhibit the knowledge about certain kinds of information in favor of other more distant ones. Achievements in these areas have previously been found to correlate with the standard appearance–reality task (Andrews, Halford, Bunch, Bowden, & Jones, 2003; Carlson & Moses, 2001).

In conclusion, 3-year-olds mastered the appearance–reality task when the pragmatics related to the key expression *looks like* were made clear. Study 1 analyzed data from the CHILDES database to document that very young children are familiar with two different uses of the locution *looks like*. In Studies 2 to 4 we found simple ways to avoid the ambiguity of this locution by making sure that both speaker and listener could assume that the reality was commonly known at the time of the appearance question and in each study found marked improvement over the standard appearance–reality task. The readiness and ease with which 3-year-olds answered the questions, their near ceiling performance, and the robustness of the results across a range of different manipulations suggest that the results cannot be attributed to an emerging, fragile ability. Rather when young children know what they are being asked, they clearly understand the appearance–reality distinction one year earlier than previously thought. The results also suggest that between the ages of 3 and 4 years children are becoming better able to understand the pragmatics

of discourse about mental states. Thus contrary to current theories (e.g., Flavell et al., 1986; Gopnik & Wellman, 1994; Perner, 1991), 3-year-olds have the conceptual foundation needed to distinguish reality from mental representations such as appearances.

Acknowledgments

We are grateful to the parents, children, and staff at Bing Nursery School, the Children's Center of the Stanford Community, and the Stanford Arboretum Children's Center for their participation. Additionally, we thank Shelly Gonzales, Felicity Miao, Megan Corty, and Maurica Carlucci for their assistance with this research, and John Flavell, Susan Johnson, and Herb Clark for helpful comments. This work is based in part on a doctoral dissertation submitted to Stanford University and was supported by a grant to Mikkel Hansen from the Danish Research Academy. Portions of this research were presented at the Society for Research in Child Development Biennial Meeting, Tampa, Florida, 2003.

References

- Allen, R. (1991). Integration of communicational cues by very young children. *Journal of Psycholinguistic Research*, 20, 389–402.
- Andrews, G., Halford, G. S., Bunch, K. M., Bowden, D., & Jones, T. (2003). Theory of mind and relational complexity. *Child Development*, 74, 1476–1499.
- Austin, J. L. (1962). *Sense and sensibilia*. New York: Oxford University Press.
- Avrutin, S., & Coopmans, P. (2000). Children who build bridges. In S. Catherine Howell et al. (Eds.), *BUCLD 24 Proceedings* (pp. 80–91). Somerville, MA: Cascadia Press.
- Bacharach, V. R., & Luszcz, M. A. (1979). Communicative competence in young children: The use of implicit linguistic information. *Child Development*, 50, 260–263.
- Bartsch, K., & Wellman, H. (1989). Young children's attribution of action to beliefs and desires. *Child Development*, 60, 946–964.
- Bloom, L., Hood, L., & Lightbown, P. (1974). Imitation in language development: If, when and why. *Cognitive Psychology*, 6, 380–420.
- Bloom, L., Lightbown, P., & Hood, L. (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development*, 40, (Serial No. 160).
- Braine, M. D. S., & Shanks, B. L. (1965a). The conservation of a shape property and a proposal about the origin of the conservations. *Canadian Journal of Psychology*, 19, 197–207.
- Braine, M. D. S., & Shanks, B. L. (1965b). The development of conservation of size. *Journal of Verbal Learning and Verbal Behavior*, 4, 227–242.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, 72, 1032–1053.
- Clark, E. V. (1978). Awareness of language: Some evidence from what children say and do. In R. J. A. Sinclair & W. Levelt (Eds.), *The child's conception of language*. Berlin: Springer.
- Clark, E. V., & Wong, A. D.-W. (2002). Pragmatic directions about language use: Offers of words and relations. *Language in Society*, 31, 181–212.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior*, 22, 245–258.

- Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*, 259–294.
- Deák, G. O., Ray, S. D., & Brenneman, K. (2003). Children's perseverative appearance–reality errors are related to emerging language skills. *Child Development*, *74*, 944–964.
- Demetras, M. (1989). *Working parents conversational responses to their two-year-old sons*. Working paper. University of Arizona.
- Flavell, J. H. (1993). The development of children's understanding of false belief and the appearance–reality distinction. *International Journal of Psychology*, *28*, 595–604.
- Flavell, J. H., Flavell, E. R., & Green, F. L. (1983). Development of the appearance–reality distinction. *Cognitive Psychology*, *15*, 95–120.
- Flavell, J. H., Flavell, E. R., & Green, F. L. (1987). Young children's knowledge about the apparent-real and pretend-real distinctions. *Developmental Psychology*, *23*, 816–822.
- Flavell, J. H., Green, F. L., & Flavell, E. R. (1986). Development of knowledge about the appearance–reality distinction. *Monographs of the Society for Research in Child Development*, *51*, 1–68.
- Flavell, J. H., Lindberg, N. A., Green, F. L., & Flavell, E. R. (1992). The development of children's understanding of the appearance–reality distinction between how people look and what they are really like. *Merrill-Palmer Quarterly*, *38*, 513–524.
- Flavell, J. H., Zhang, X.-D., Zou, H., Dong, Q., & Qi, S. (1983). A comparison between the development of the appearance–reality distinction in the People's Republic of China and the United States. *Cognitive Psychology*, *15*, 459–466.
- Gopnik, A., & Wellman, H. M. (1994). The 'theory' theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–293). Cambridge, UK: Cambridge Univ. Press.
- Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance–reality distinction. *Child Development*, *59*, 26–37.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: speech acts*. New York: Seminar press.
- Gross, D., & Harris, P. L. (1988). False beliefs about emotion: Children's understanding of misleading emotional displays. *International Journal of Behavioral Development*, *11*, 475–488.
- Haryu, E. (1991). A developmental study of children's use of "mutual exclusivity" and context to interpret novel words. *Japanese Journal of Educational Psychology*, *39*, 11–20.
- Haryu, E., & Imai, M. (1999). Controlling the application of the mutual exclusivity assumption in the acquisition of lexical hierarchies. *Japanese Psychological Research*, *41*, 21–34.
- Kuczaj, S. (1976). *-ing, -s and -ed: A study of the acquisition of certain verb inflections*. Unpublished doctoral dissertation, University of Minnesota.
- Lillard, A. S. (1999). Siegal on Piaget's legacy: Gricean child meets blundering experimenter. *Developmental Science*, *2*, 18–21.
- Luszcz, M. A., & Bacharach, V. R. (1983). The emergence of communicative competence: Detection of conversational topics. *Journal of Child Language*, *10*, 623–637.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. 3rd ed. 2: The Database*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Melot, A., & Houdé, O. (1998). Categorization and theories of mind: The case of the appearance–reality distinction. *Cahiers de Psychologie Cognitive*, *17*, 71–93.
- Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.
- Rice, C., Koinis, D., Sullivan, K., Tager-Flusberg, H., & Winner, H. (1997). When 3-year-olds pass the appearance–reality test. *Developmental Psychology*, *33*, 54–61.
- Sachs, J. (1983). *Talking about the there and then: The emergence of displaced reference in parent–child discourse*. In K. E. Nelson (Ed.), *Children's language* (Vol. 4). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sapp, F., Lee, K., & Muir, D. (2000). Three-year-olds' difficulty with the appearance–reality distinction: Is it real or is it apparent?. *Developmental Psychology*, *36*, 547–560.
- Shatz, M. (1978). On the development of communicative understandings: An early strategy for interpreting and responding to messages. *Cognitive Psychology*, *10*, 271–301.

- Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition*, *14*, 301–321.
- Siegal, M., & Peterson, C. (1994). Children's theory of mind and the conversational territory of cognitive development. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sodian, B., Huelsken, C., Ebner, C., & Thoermer, C. (1998). Die begriffliche unterscheidung von mentalitaet und realitaet im kindlichen symbolspiel-vorlaeuf einer theory of mind? Children's differentiation of mentality and reality in pretense-precursor to a theory of mind?. *Sprache & Kognition*, *17*, 199–213.
- Suppes, P. (1974). The semantics of children's language. *American Psychologist*, *29*, 103–114.
- Taylor, M., & Flavell, J. H. (1984). Seeing and believing: Children's understanding of the distinction between appearance and reality. *Child Development*, *55*, 1710–1720.
- Taylor, M., & Hort, B. (1990). Can children be trained in making the distinction between appearance and reality?. *Cognitive Development*, *5*, 89–99.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*, 655–684.
- Woolley, J. D., & Wellman, H. M. (1990). Young children's understanding of realities, nonrealities, and appearances. *Child Development*, *61*, 946–961.