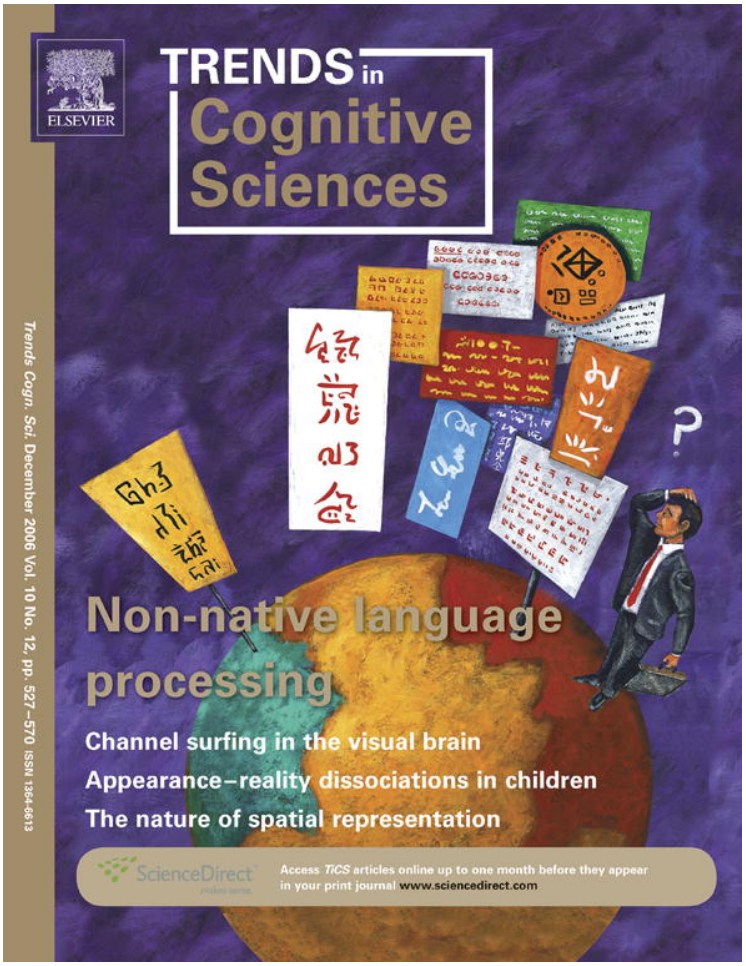

Opinion TRENDS in Cognitive Sciences Vol.10 No.12

Do children really confuse appearance and reality?

Gedeon O. Deák

Department of Cognitive Science, University of California–San Diego, La Jolla, CA 92093-0515, USA

Our understanding of many mental, social and physical phenomena hinges on a general understanding that appearances can differ from reality. Yet young children sometimes seem unable to understand appearance–reality dissociations. In a standard test, children are shown a deceptive object and asked what it really is and what it looks like. Many preschool children give the same answer to both questions. This error has been attributed to children's inflexible conceptual representations or inflexibility in representing their own changing beliefs. However, evidence fails to support either hypothesis: new tests show that young children generally understand appearance–reality discrepancies as well as fantasy–reality distinctions. These tests instead implicate children's failure to understand the unfamiliar discourse format of the standard test. This misunderstanding might reveal a subtler difficulty in making logical inferences about questions.

Introduction

Adults take it for granted that appearances can be deceiving. A faraway stranger is mistaken for a friend; a clever reproduction fools an art collector; a straw seems to bend in water. However, young children might not understand such dissociations. A widely used test indicates that preschool children (i.e. three- and four-year-olds) confuse the appearances and identities of misleading objects. This raises questions about how children understand other misleading situations, such as fantasy play or social deception. Because of these wide-ranging implications, the appearance–reality (AR) test has become widely used as a test of children's 'theory of mind'. Yet evidence suggests that young children do in fact know that appearances can differ from reality. Also, evidence is lacking that AR understanding is related to theory of mind. This paper examines these claims and suggests a different conclusion: the AR test assesses how children respond to sequences of questions. As it happens, some preschoolers will repeat their first answer to every successive question about a topic. This odd tendency confounds any conclusions about what children really understand. Thus, although alternate methods have revealed children's facility in discriminating reality from appearances or fantasy, traditional methods have inadvertently revealed a puzzling gap in children's pragmatic skills.

Do children know that appearances can be deceiving?

In the traditional AR test, children are shown a deceptive object – for example, an eraser that looks like a crayon (Figure 1) – and asked if it 'looks like' an eraser or a crayon and if it 'really and truly is' an eraser or a crayon [1–3]. Many preschoolers give the same answer to both questions, suggesting some kind of cognitive inflexibility. (The protocol test uses preliminary questions to rule out basic perceptual or comprehension problems.) Also, children seldom focus on appearance words ('crayon') as some theories would predict [4]. Instead, children usually focus on the real identify (e.g. eraser), as if compelled to represent the 'best' identity of the object.

Perseverative responses to AR questions are related to age. Three-year-olds make many AR errors; five-year-olds make few and four-year-olds are intermediate [1–3,5]. This age trend is seen in various AR tests and other related tests (e.g. of fantasy versus reality). Several are summarized in Box 1. Because of this common trend, the AR test has become a benchmark of cognitive development [6–8]. For instance, in addition to AR errors, three-year-olds often fail to infer that someone could hold a belief that the child knows is false [9,10]. Also, three-year-olds sometimes seem to confuse pretense or fantasy, and reality [11]. Three-year-olds even seem to focus on shape and ignore functions when classifying objects [12]. Thus, AR errors seem to capture a broader problem of representational inflexibility.

However, closer examination raises serious doubts. Three-year-olds' understanding of false beliefs (FBs) is task dependent [13,14]. 'Leakages' of fantasy into reality (e.g. avoiding a box after pretending it contained a spider) might reflect incidental social and emotional processes, not confusion about what is real [15,16]. Also, three-year-olds can reason about non-obvious 'deep' properties and categories – for instance, classifying objects by functions as well as shape [17] – and reason about non-obvious biological categories in terms of unseen 'essences' [18] or unseen agents [19]. Thus, there is no three-year limitation on reasoning about misleading appearances. Still, preschoolers make striking errors in the AR test. By examining these in detail, we might better understand what preschoolers do (and do not) know about misleading objects and situations.

Although most sources ascribe children's AR errors to representational inflexibility or theory-of-mind deficits, recent evidence points to discourse-level misunderstandings. It is possible that these same discourse problems also explain preschoolers' fantasy–reality 'leakages' and some

Corresponding author: Deák, G.O. (deak@cogsci.ucsd.edu).
Available online 31 October 2006.

 1364-6613/$ – see front matter doi:10.1016/j.tics.2006.09.012

Figure 1. Sample deceptive objects used in AR and flexible-naming tests. Top row: sample deceptive objects used in AR tests. Left to right: crayon eraser, crayon candle, rubber rock, candy magnet. Bottom row: nondeceptive representational objects used in flexible-naming tests. Left to right: banana pen, crayon dinosaur, seashell soap. Deceptiveness (i.e. good fakes versus obvious toys) does not influence children's AR performance [5].

theory-of-mind errors. However, before detailing this argument, we must examine the popular assumption that children's AR errors, and related errors about the real and unreal, come from representational inflexibility.

Box 1. Versions of AR and related tests

Although I have focused on the 'object identity' appearance–reality (AR) test here, other AR tests have been used – typically, color AR. Typical stimuli and questions from several AR tests and a pretense–reality test are shown. Bracketed text indicates wording that has varied across studies.

Object identity [1–3,5–8,31–33]
Stimulus = deceptive object (e.g. chocolate magnet)
Question 1 = 'What does this look like [to your eyes (right now)]? (Does it look like) a chocolate or (does it look like) a magnet?'
Question 2 = 'What is this really (and truly)? [Is it really (and truly)] a chocolate or [really (and truly)] a magnet?'

Color [1–3,31–33]
Stimulus = colored line drawing (e.g. pink rabbit) and (blue) cellophane envelope
Question 1 = '[When you look at this (right now)], does it look like it's pink or (does it look like) it's blue?'
Question 2 = 'What is this really (and truly)? [Is it really (and truly)] pink or [really (and truly)] blue?'

Emotion [43,44]
Stimulus = brief vignettes of a character trying to hide an emotion
Question 1 = 'How did Diana *really* feel, very happy or a bit happy, or very sad or a bit sad?'
Question 2 = 'How did Diana *look* when that happened to her? Did she look very happy or a bit happy, or very sad or a bit sad?'

Pretense [3]
Stimulus = everyday objects with plausible pretend identities (e.g. plastic cup, to be used as a hat for a plush bear)
Question 1 = 'What is it really and truly? Is it really and truly a cup or really and truly a hat?'
Question 2 = 'What am I pretending this is right now? Am I pretending this is a cup or pretending this is a hat?'

Ruling out representational inflexibility

Since the 1970s, AR errors have usually been attributed to representational inflexibility (e.g. keeping only one object construal in working memory) [1,2]. This fits claims, now considered over-simplistic [20], that preschoolers are unidimensional thinkers [4].

However, representational inflexibility cannot explain AR errors. Virtually all three-year-olds will readily and accurately assign a person or object to several categories [21–25]. For example, they accept unfamiliar label pairs (e.g. both crayon and eraser), even for deceptive objects [22,24,25]. Corroborating results from multiple studies and laboratories show that three-year-olds and even two-year-olds [21] can rapidly shift between representing and describing the appearance and function of an object. These results clearly disconfirm representational inflexibility accounts of AR errors. One reason (elaborated later) is that recent methods have used more natural conversational prompts [22,24]. Box 2 illustrates natural prompts, revealing a three-year-old's representational flexibility. Importantly, such findings fit other evidence that two- and three-year-olds can accurately and readily shuttle between representing pretense and reality, given adequate prompts [26,27]. Thus, three-year-olds, when asked the right questions, can easily and accurately describe real and fake or imagined aspects of an object or situation.

This conclusion might seem surprising because preschoolers appear to be challenged by nested category relationships [4], FBs [9] or multiple object labels [28], all of which might require representational flexibility. However, these claims are controversial [13,22,29]. Ultimately, the most relevant studies show that three-year-olds can, in natural conversations, flexibly describe misleading appearances and functions of objects. Thus, another account is needed. One alternative is that three-year-olds cannot always access or represent their own previous mistaken beliefs about a deceptive object – a theory-of-mind limitation.

Is it not a theory-of-mind task?

The AR test has recently been used to assess preschoolers' theory of mind for inferences about mental states and contents [6–8]. This capacity expands qualitatively from

Box 2. Excerpt from transcript of 41-month-old female in flexible naming test

Experimenter (E): 'All right, now I'm going to show you some things and I want you to tell me what they are called.' (Shows Dalmatian puppet) 'What is this called?'
Child (C): '(It's a) Dalmatian.'
E: 'What kind of thing is a Dalmatian?'
C: 'A dog. . .(It's a) fire-engine dog.'
. . .
E: 'Now watch this.' (E puts puppet on hand; pretends to talk with it) 'What do you call something that does this?'
C: 'Puppet.'
. . .
E: 'Is it a dog and a cat?'
C: (Shakes head)
E: 'Is it a cat and a puppet?'
C: (Shakes head)
E: 'Is it a dog and a puppet?'
C: 'Yes.'
(Taken from a study described in Ref. [25].)

two to five years of age. With regard to the AR test, the logic is that three-year-olds first identify the deceptive object by appearance (e.g. crayon) but revise this identity belief after seeing its 'true' function (i.e. eraser). However, they cannot then reflect on their prior mistaken belief. This seems akin to reasoning about FBs, a cornerstone theory-of-mind skill [9,10]. As a consequence, at least 18 peer-reviewed studies between 2000 and 2005 used AR as a theory-of-mind test.

The justification for considering the AR as a theory-of-mind test began with a report [30] of moderate correlations between the AR and two theory-of-mind tests: FB and representational change (RC). Later studies have usually found AR–FB associations below $r = 0.30$ [6–8,31–33]. Nonparametric analyses also show modest between-test associations [34].

Yet there is little support for treating AR as a theory-of-mind test. Studies with the highest AR–FB correlations (highest biserial $r = 0.65$) had overlapping task content, including very similar wording of questions (i.e. using 'really and truly' in both tests) and discourse format (discussed later) [6,30]. Notably, the original study [30] used AR, FB and RC questions that differed by only a few words, all with the same stimuli and scenarios. Thus, shared method variance has greatly confounded the highest reported correlations. Studies that controlled method variance have often reported nonsignificant correlations (controlling for age) no greater than $r \approx 0.25$ [31,33]. Thus, AR and FB tests seem to measure mostly different abilities.

Other results suggest that the AR test does not require inferences about mental states. First, if three-year-olds cannot reason about changing representations, they should not accept both appearance and function labels for an object (e.g. 'dog' and 'puppet'), yet they consistently do [22,23]. Second, three-year-olds succeed in a non-verbal AR test [35], yet it is mystifying how nonverbal responses could reduce the belief–representation demands. Third, control tests (described later) associated with the AR test share its discourse format but require no belief representations [5,25]. Thus, although it is difficult to explain children's responses to complex tasks such as AR and FB, there is no compelling evidence that the AR test assesses belief–state inferences. This means we should be cautious in interpreting children's AR responses as theory-of-mind indices. It also demands an alternative account of AR errors. One alternative focuses on the discourse structure of the test and what it reveals about children's understanding of questions.

Role of discourse understanding

Siegal [36] argued that the AR test protocol is pragmatically odd and this contributes to children's errors. The oddness is hard to define but is inarguable: after seeing the object and labeling it by appearance, and seeing its function and agreeing with the function label, children are asked two successive forced-choice questions ending with the same choices – 'What does this look like, a ($label_a$) or ($label_b$)?' and 'What is it really and truly, a ($label_a$) or ($label_b$)?' Because the child initially provided both labels in the appropriate context, the explicit questions violate a Gricean maxim of efficient communication, and thus seem 'jarring'. Consequently, the adult's meaning might be unclear. This hypothesis is fleshed out by recent findings.

Several experiments [5,25] have shown that the crucial discourse format of two successive forced-choice questions about a topic or percept, ending in the same two verbal options, causes some preschoolers to repeat their answer inappropriately. This happens in both the AR test and in control tests, with no deceptive objects or questions about reality or appearances. In one control test, children saw objects such as a fur square with a bell attached and were asked: 'What is this made of, fur or a bell?' and 'What does it have, fur or a bell?'. Children who make AR errors also tend to give the same response to both questions [5], as if they believe that the object 'has a bell' and is 'made of bell(s)'. (Of course, they do not actually hold this bizarre perception; see Ref. [37].) Such control tests share up to 40% variance with the AR test, after controlling for age and verbal abilities. In a more extreme control test, children answered easy question pairs (e.g. about a picture of a dog and bird: 'which one…flies in the sky, a dog or a bird?' and 'which one…chews on bones, a dog or a bird?'. Children who perseverated in answering these questions also made many AR errors. It seems that AR errors indicate that children have a tendency to process series of questions partly independently of the content or topic of each question.

Other findings back up this claim. Three-year-olds in a language-modified AR test correctly choose between objects instead of labels [35]. In more natural conversations, three-year-olds will label the appearance and function of an object [22,23]. After adults describe the appearance and function of an object using formulaic phrasing (e.g. '…this looks like a rock but it's really a sponge'), three-year-olds produce analogous formulas to describe deceptive objects. Thus, when answering successive forced-choice questions (standard AR or control tests), preschoolers erroneously repeat their answer. However, when describing deceptive objects in altered-discourse tests, three-year-olds are fairly accurate (Figure 2).

A few intriguing studies suggest that this discourse-dependent pattern extends beyond the AR test. Children's FB performance improves in altered discourse conditions [14]. Also, fantasy–reality intrusions are reduced when termination of fantasy play is pragmatically highlighted [15]. Thus, preschoolers seem to describe misleading or deceptive situations more accurately when paralanguage and pragmatics clarify the test questions [38,39]. Even in the AR test, when questions are couched in a pretend-play interaction or with the premise of deceiving someone, preschoolers can more easily follow the meanings of successive questions and make fewer errors [3,40].

Yet it remains unclear exactly why preschoolers repeat answers to successive forced-choice questions. One hint is that children's ability to determine whether a question is indeterminate (i.e. cannot be resolved without further information) predicts their AR accuracy [5,25]. Preschoolers tend to be overconfident when interpreting ambiguous messages, and awareness of indeterminacy develops through childhood [38,39]. In the AR test, children who are overconfident (i.e. unaware of indeterminacy) might ignore possible alternative meanings of successive

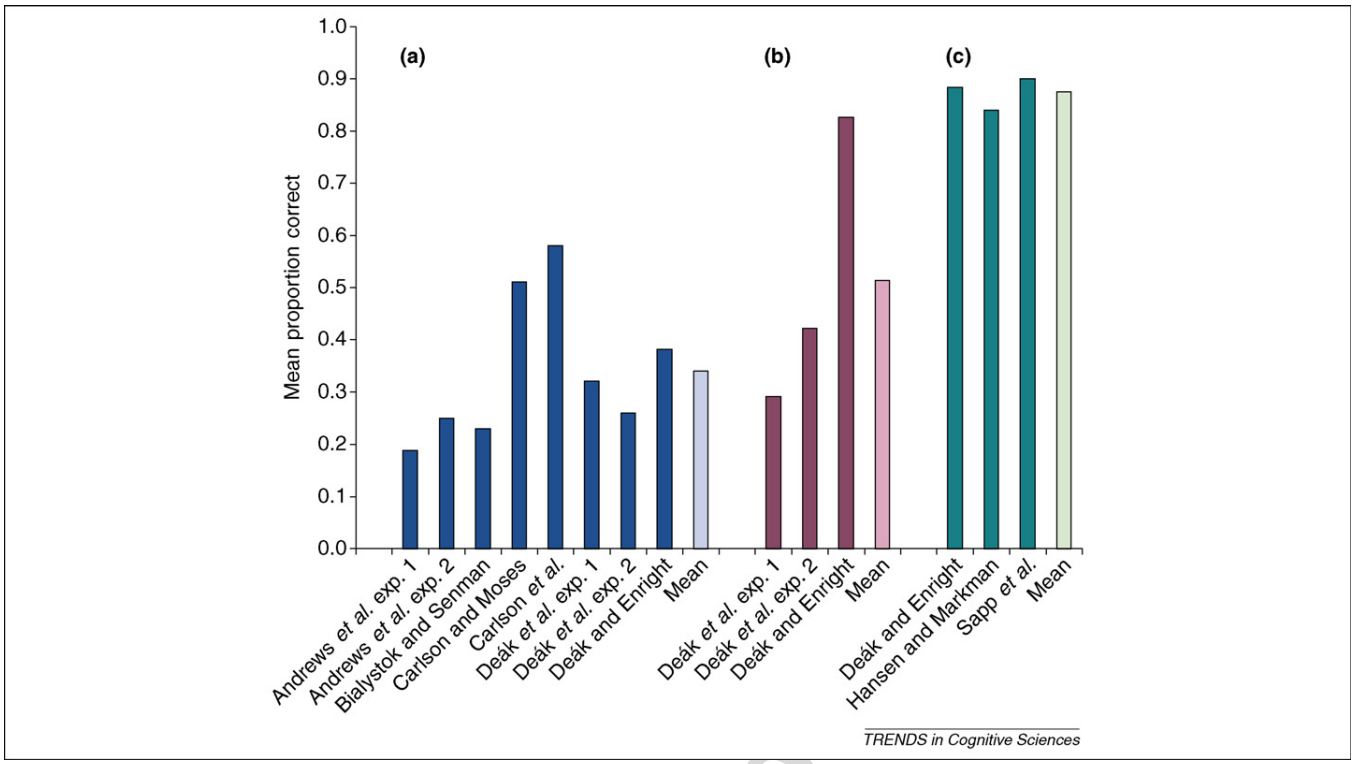

Figure 2. Three-year-old children's performance in studies that use the standard AR test (blue bars), control tests (red bars) and discourse-altered AR tests (green bars). Correct responses (to both questions about an object) in **(a)** studies using AR tests with standard discourse format; **(b)** control tests with the same discourse format; and **(c)** alternate AR test versions with altered discourse formats (e.g. non-verbal responses or object-description formulas). The lighter bars are averages of all experiments in the relevant category. In general, the control-test results are closer to standard AR results than are altered-discourse AR results. The different control study (Deák and Enright [25]) used easy questions about familiar stimulus pictures. Using data from Refs [5–7,24,25,35,42].

questions, and treat their first response as an all-purpose correct answer. That is, whereas adults assume that different question have different answers, preschoolers make the opposite assumption, if confident in their first answer. This hypothesis currently has only correlational support, and further research is needed. However, it is not the only alternate proposed cause of AR errors.

Semantics

Semantics also have a role in AR errors. The 'looks like' question is ambiguous [24,37] and this seems to increase AR errors [24] compared with a semantically simple control test [25] (Figure 2). However, three-year-olds still make errors when AR questions are worded less ambiguously [5], so discourse format remains a significant factor. However, still other cognitive factors might affect performance.

Memory and inhibition

Children might perseverate because they cannot maintain two labels in working memory. Showing props for both word choices (e.g. a rock and sponge, for a deceptive rock–sponge) reduces errors [40]. However, working memory span has not predicted AR errors in several studies [5,25,31], and verbal memory cues do not reduce AR errors [41]. Thus, the role of working-memory limitations in AR errors is tenuous. Another possible factor is inhibition: children must suppress their first answer to answer the second AR question correctly, so poor inhibition might cause errors. One study found a weak correlation between verbal inhibition and AR tasks [32]. However, several others did not [7,25,32,42], so it is unclear whether inhibition has a significant role in children's thinking about real and misleading objects and situations.

Concluding remarks

Children's AR errors are misleading. They are not tied to the 'perceptual seduction' of deceptive objects or to

Box 3. Questions for future research

- Which elements of the crucial discourse format contribute to children's errors? Errors might be related to one or more elements: (i) successive questions about one referent; (ii) repetition of the verbal choices; or (iii) verbal or lexical choices.
- Which tests of children's thinking pose one question, then ask a second question which children might mistakenly interpret as a repetition of the first? For example, the FB test [10] sometimes uses successive forced-choice questions [6,8], and other times uses open-ended questions or a combination [31,32]. This procedural variable might contribute to differences between studies.
- Is it somewhat harder (albeit not very hard) for three-year-olds to describe deceptive objects than nondeceptive objects? Which content factors contribute to different results across versions of the AR test (Box 1 and Figure 2)?
- Lexical knowledge (e.g. vocabulary) correlates with children's AR performance. Is this because vocabulary and discourse knowledge are associated? Or does vocabulary independently contribute to correctly answering AR questions?
- How does ability to recognize whether or not a question is indeterminate [40] develop?

difficulties of reflecting on changing beliefs. More generally, children in natural conversation seem rarely to confuse misleading or pretended identities with real identities. However, AR errors might manifest a generalized confusion about successive forced-choice questions about a referent. Younger children might not, after answering one question correctly, realize that subsequent questions pertain to different aspects of the topic. This failure of 'rational uncertainty', along with other possible causes of discourse-dependent AR errors, requires further investigation (Box 3). Moreover, a wide variety of paradigms have been used to test children's discrimination of the real and unreal; discourse and semantic factors alone almost certainly cannot explain all findings. Ultimately, then, a more synthetic, multivariate model of children's AR, FB and fantasy–reality errors – one that incorporates linguistic factors – is needed.

Acknowledgements

This work was supported by the National Science Foundation (BSC-0092027). Thanks to Anna Holt, Anna Krasno, Leah Welch, Susan Welch, Natalie Wong and anonymous referees for helpful comments.

References

1 Flavell, J.H. *et al.* (1983) Development of the appearance–reality distinction. *Cognit. Psychol.* 15, 95–120
2 Flavell, J.H. *et al.* (1986) Development of knowledge about the appearance–reality distinction. *Monogr. Soc. Res. Child Dev.* 51, 1–87
3 Flavell, J.H. *et al.* (1987) Young children's knowledge about the apparent–real and pretend–real distinctions. *Dev. Psychol.* 23, 816–822
4 Inhelder, B. and Piaget, J. (1956) *The Early Growth of Logic in the Child.* Harper & Row
5 Deák, G.O. *et al.* (2003) Children's perseverative appearance–reality errors are related to emerging language skills. *Child Dev.* 74, 944–964
6 Andrews, G. *et al.* (2003) Theory of mind and relational complexity. *Child Dev.* 74, 1476–1499
7 Carlson, S.M. and Moses, L.J. (2001) Individual differences in inhibitory control and children's theory of mind. *Child Dev.* 72, 1032–1053
8 Frye, D. *et al.* (1996) Theory of mind and rule-based reasoning. *Cogn. Dev.* 10, 483–527
9 Perner, J. (1991) *Understanding the Representational Mind.* MIT Press
10 Russell, J. (2005) Justifying all the fuss about false belief. *Trends Cogn. Sci.* 9, 307–308
11 Harris, P.L. *et al.* (1991) Monsters, ghosts, and witches: testing the limits of the fantasy–reality distinction in young children. *Br. J. Dev. Psychol.* 9, 105–123
12 Smith, L.B. *et al.* (1996) Naming in young children: a dumb attentional mechanism? *Cognition* 60, 143–171
13 Bloom, P. and German, T.P. (2000) Two reasons to abandon the false belief task as a test of theory of mind. *Cognition* 77, B25–B31
14 Carpenter, M. *et al.* (2002) A new false belief test for 36-month-olds. *Br. J. Dev. Psychol.* 20, 393–420
15 Golomb, C. and Galasso, L. (1995) Make believe and reality: explorations of the imaginary realm. *Dev. Psychol.* 31, 800–810
16 Gergely, G. (2002) Some confusions about pretense–reality confusions. *Dev. Sci.* 5, 417–419
17 Kemler Nelson, D.G. *et al.* (2000) Two-year-olds will name artifacts by their functions. *Child Dev.* 71, 1271–1288
18 Gelman, S.A. (2004) Psychological essentialism in children. *Trends Cogn. Sci.* 8, 404–408
19 Siegal, M. and Share, D.L. (1990) Contamination sensitivity in young children. *Dev. Psychol.* 26, 455–458
20 Siegler, R.S. (1996) Unidimensional thinking, multidimensional thinking, and characteristic tendencies of thought. In *The Five to Seven Shift: The Age of Reason and Responsibility* (Sameroff, A. and Haith, M., eds), pp. 63–84, University of Chicago
21 Clark, E.V. and Svaib, T.A. (1997) Speaker perspective and reference in young children. *First Lang.* 17, 57–74
22 Deák, G. and Maratsos, M. (1998) On having complex representations of things: preschoolers use multiple words for objects and people. *Dev. Psychol.* 34, 224–240
23 Deák, G.O. *et al.* (2001) By any other name: when will preschoolers produce multiple labels for a referent? *J. Child Lang.* 28, 787–804
24 Hansen, M.B. and Markman, E.M. (2005) Appearance questions can be misleading: a discourse-based account of the appearance–reality problem. *Cognit. Psychol.* 50, 233–263
25 Deák, G.O. and Enright, B. (2006) Choose and choose again: appearance–reality errors and the logic of questioning. *Dev. Sci.* 9, 323–333
26 Amsel, E. *et al.* (1996) Young children's memory for the true and pretend identities of objects. *Dev. Psychol.* 32, 479–491
27 Walker-Andrews, A.S. and Kahana-Kalman, R. (1999) The understanding of pretence across the second year of life. *Br. J. Dev. Psychol.* 17, 523–536
28 Markman, E.M. and Wachtel, G.F. (1988) Children's use of mutual exclusivity to constrain the meanings of words. *Cognit. Psychol.* 20, 121–157
29 Trabasso, T. *et al.* (1978) How do children solve class-inclusion problems? In *Children's Thinking: What Develops?* (Siegler, R.S., ed.), pp. 151–180, Lawrence Erlbaum
30 Gopnik, A. and 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 Dev.* 59, 26–37
31 Carlson, S.M. *et al.* (2002) Speaker perspective and reference in young children. *Infant Child Dev.* 11, 73–92
32 Carlson, S.M. *et al.* (2004) Executive function and theory of mind: stability and prediction from ages 2 to 3. *Dev. Psychol.* 40, 1105–1122
33 Miller, S.A. *et al.* (1997) Children's understanding of false beliefs that result from developmental misconceptions. *Cogn. Dev.* 12, 21–51
34 Melot, A-M. and Angeard, N. (2003) Theory of mind: is training contagious? *Dev. Sci.* 6, 178–184
35 Sapp, F. *et al.* (2000) Three-year-olds' difficulty with the appearance–reality distinction: is it real or is it apparent? *Dev. Psychol.* 36, 547–560
36 Siegal, M. (1997) *Knowing Children: Experiments in Conversation and Cognition* (2nd edn), Psychology Press
37 Deák, G.O. (2000) The growth of flexible problem solving: preschool children use changing verbal cues to infer multiple word meanings. *J. Cognit. Dev.* 1, 157–192
38 Markman, E.M. (1979) Realizing that you don't understand: elementary school children's awareness of inconsistencies. *Child Dev.* 50, 643–655
39 Fay, A.L. and Klahr, D. (1996) Knowing about guessing and guessing about knowing: preschoolers' understanding of indeterminacy. *Child Dev.* 67, 689–716
40 Rice, C. *et al.* (1997) When 3-year-olds pass the appearance–reality test. *Dev. Psychol.* 33, 54–61
41 Flavell, J.H. *et al.* (1987) The effects of question clarification and memory aids on young children's performance on appearance–reality tasks. *Cogn. Dev.* 2, 127–144
42 Bailystok, E. and Senman, L. (2004) Executive processes in appearance–reality tasks: the role of inhibition of attention and symbolic representation. *Child Dev.* 75, 562–579
43 Harris, P.L. *et al.* (1986) Children's understanding of the distinction between real and apparent emotion. *Child Dev.* 57, 895–909
44 Friend, M. and Davis, T.L. (1993) Appearance–reality distinction: children's understanding of the physical and affective domains. *Dev. Psychol.* 29, 907–914