Title
Do children really confuse appearance and reality?

Permalink
https://escholarship.org/uc/item/9756839

Journal
Trends in cognitive sciences, 10(12)

ISSN
1364-6613

Author
Deák, Gedeon O

Publication Date
2006-12-01

Peer reviewed
Do children really confuse appearance and reality?

Gedeon O. Deák

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 identity (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.

Representational inflexibility

Three-year-olds make many AR errors; five-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 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.
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 3 considered over-simplistic [20], that preschoolers are unidimensional thinkers [4]. However, representational inflexibility cannot explain words, all with the same stimuli and scenarios. Thus, 6 AR errors. Virtually all three-year-olds will readily and 7 accurately assign a person or object to several categories 8 for example, they accept unfamiliar label pairs (e.g. both crayon and eraser), even for deceptive objects 10. {AR–FB associations below recommended thresholds [34]. Nonparametric analyses also show modest between- 11 objects such as a fur square with a bell attached and were reality or appearances. In one control test, children saw 12 causes some preschoolers to repeat their answer about a topic or percept, ending in the same two verbal options, and discourse format can be challenging [15,20]; however, the use of multiple object labels 13 associations [34]. Notably, the original study [30] states that two- and three-year-olds can accurately and readily 14, even for deceptive objects [22,24,25]. Corroborating results from multiple studies [24,25]. AR test associations [34]. Notably, the original study [30] states that two- and three-year-olds can accurately and readily 15 prompts, revealing a three-year-old's representational 18 flexibility. Importantly, such findings fit other evidence 19 that two- and three-year-olds can accurately and readily 20 shuffle between representing pretense and reality, given 21 adequate prompts [26,27]. Thus, three-year-olds, when 22 asked the right questions, can easily and accurately 23 describe real and fake or imagined aspects of an object or 25 situation. This conclusion might seem surprising because 27 preschoolers appear to be challenged by nested category 29 relationships [4], FBs [9] or multiple object labels [28], all 30 of which may require representational flexibility. However, these claims are controversial [13,22,29]. Ultimately, the most relevant studies show that three- 33 year-olds can, in natural conversations, flexibly describe 34 misleading appearances and functions of objects. Thus, 35 another account is needed. One alternative is that three- 36 year-olds cannot always access or represent their own previous mistaken beliefs about a deceptive object – a 37 theory-of-limitation.

39 Is it not a theory-of-limit task? 40 The AR test has recently been used to assess preschoolers' 100 or (labels)? and 'What is it really and truly, a (label a) or (labelb)?' Because the child initially provided both labels in 42 contents [6–8,31–33]. Such control tests share up to 10% variance with the AR test, after controlling for age 43 method variance have often reported nonsignificant results clearly disconfirm representational inflexibility. 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 shuffle 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 may 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-limitation.

40 The AR test has recently been used to assess preschoolers' 100 or (labels)? and 'What is it really and truly, a (label a) or (labelb)?' Because the child initially provided both labels in 42 contents [6–8,31–33]. Such control tests share up to 10% variance with the AR test, after controlling for age method variance have often reported nonsignificant results clearly disconfirm representational inflexibility. 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 shuffle 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 may 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-limitation.

40 The AR test has recently been used to assess preschoolers' 100 or (labels)? and 'What is it really and truly, a (label a) or (labelb)?' Because the child initially provided both labels in 42 contents [6–8,31–33]. Such control tests share up to 10% variance with the AR test, after controlling for age method variance have often reported nonsignificant results clearly disconfirm representational inflexibility. 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 shuffle 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 may 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-limitation.

40 The AR test has recently been used to assess preschoolers' 100 or (labels)? and 'What is it really and truly, a (label a) or (labelb)?' Because the child initially provided both labels in 42 contents [6–8,31–33]. Such control tests share up to 10% variance with the AR test, after controlling for age method variance have often reported nonsignificant results clearly disconfirm representational inflexibility. 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 shuffle 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 may 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-limitation.

40 The AR test has recently been used to assess preschoolers' 100 or (labels)? and 'What is it really and truly, a (label a) or (labelb)?' Because the child initially provided both labels in 42 contents [6–8,31–33]. Such control tests share up to 10% variance with the AR test, after controlling for age method variance have often reported nonsignificant results clearly disconfirm representational inflexibility. 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 shuffle 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 may 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-limitation.
1 and verbal abilities. In a more extreme control test, 61 factor. However, still other cognitive factors might affect 62 performance.

2 children answered easy question pairs (e.g. about a picture 63 Memory and inhibition
3 of a dog and bird: ‘which one…flies in the sky, a dog or a 64 Children might perseverate because they cannot maintain 65 two labels in working memory. Showing props for both 66 word choices (e.g. a rock and sponge, for a deceptive rock– 67 sponge) reduces errors [40]. However, working memory 68 span has not predicted AR errors in several studies 69 [5,25,31], and verbal memory cues do not reduce AR errors 70 [41]. Thus, the role of working-memory limitations in AR 71 errors is tenuous. Another possible factor is inhibition: 72 children must suppress their first answer to answer the 73 second AR question correctly, so poor inhibition might 74 cause errors. One study found a weak correlation between 75 verbal inhibition and AR tasks [32]. However, several 76 others did not [7,25,32,42], so it is unclear whether 77 inhibition has a significant role in children’s thinking 78 about real and misleading objects and situations.

3 Flavell, J.H. (1956). The Early Growth of Logic in 80 Concluding remarks
4 and anonymous referees for helpful comments.

5 Deák, G.O. (1983) Development of the appearance–reality 83 Thanks to Anna Holt, Anna Krasno, Leah Welch, Susan Welch, Natalie Wong
6 and verbal inhibition and AR tasks [32]. However, several 84 Acknowledgements
7 Andrews, G. et al. (2003) Theory of mind and relational 85 This work was supported by the National Science Foundation (BSC-0092027).
8 inhibition has a significant role in children’s thinking 86
9 one that incorporates linguistic factors – is needed. 87
10 Concluding remarks
11 language-modified AR test correctly choose between
12 objects instead of labels [35]. In more natural
13 conversations, three-year-olds will label the appearance
14 and function of an object [22,23]. After adults describe the
15 appearance and function of an object using formulaic
16 phrasing (e.g. ‘...this looks like a rock but it’s really a
17 sponge’), three-year-olds produce analogous formulas to
18 describe deceptive objects. Thus, when answering
19 successive forced-choice questions (standard AR or control
20 tests), preschoolers erroneously repeat their answer
21 However, when describing deceptive objects in altered-
22 discourse tests, three-year-olds are fairly accurate
23 (Figure 2).
24 A few intriguing studies suggest that this discourse-
25 dependent pattern extends beyond the AR test. Children’s
26 FB performance improves in altered discourse conditions
27 [14]. Also, fantasy–reality intrusions are reduced when
28 termination of fantasy play is pragmatically highlighted
29 [15]. Thus, preschoolers seem to describe misleading or
30 deceptive situations more accurately when paralanguage
31 and pragmatics clarify the test questions [38,39]. Even in
32 the AR test, when questions are couched in a pretend-play
33 interaction or with the premise of deceiving someone,
34 preschoolers can more easily follow the meanings of
35 successive questions and make fewer errors [3,40].
36 Yet it remains unclear exactly why preschoolers repeat
37 answers to successive forced-choice questions. One hint is
38 that children’s ability to determine whether a question is
39 indeterminate (i.e. cannot be resolved without further
40 information) predicts their AR accuracy [5,25,31].
41 Preschoolers tend to be overconfident when interpreting
42 ambiguous messages, and awareness of indeterminacy
43 develops through childhood [38,39]. In the AR test,
44 children who are overconfident (i.e. unaware of
45 indeterminacy) might ignore possible alternative
46 meanings of successive questions, and treat their first
47 response as an all-purpose correct answer. That is,
48 whereas adults assume that different question have
49 different answers, preschoolers make the opposite
50 assumption, if confident in their first answer. This
51 hypothesis currently has only correlational support, and
52 further research is needed. However, it is not the only
53 alternate proposed cause of AR errors.
54 Semantics
55 Semantics also have a role in AR errors. The ‘looks like’
56 question is ambiguous [24,37] and this seems to increase
57 AR errors [24] compared with a semantically simple
58 control test [25] (Figure 2). However, three-year-olds still
59 make errors when AR questions are worded less
60 ambiguously [5], so discourse format remains a significant

References
1 Flavell, J.H. et al. (1983) Development of the appearance–reality
2 distinction. Cognit. Psychol. 15, 95–129
3 Flavell, J.H. et al. (1986) Development of knowledge about the
4 appearance–reality distinction. Monogr. Soc. Res. Child Dev. 51,
5 1–87
6 Flavell, J.H. et al. (1987) Young children’s knowledge about the
7 apparent–real and pretend–real distinctions. Dev. Psychol. 23,
8 816–822
9 Inhelder, B. and Piaget, J. (1956). The Early Growth of Logic in
10 the Child, Harper & Row
12 errors are related to emerging language skills. Child Dev.
13 74, 944–964
15 complexity. Child Dev. 74, 1476–1499
16 Carlson, S.M. and Moses, L.J. (2001) Individual differences in
17 inhibitory control and children’s theory of mind. Child Dev.
18 72, 1032–1053
20 Cogn. Dev. 10, 483–527


17 Kemler Nelson, D.G. et al. (2000) Two-year-olds will name artifacts by their functions. Child Dev. 71, 1271–1288


26 Amsel, E. et al. (1996) Young children’s memory for the true and pretend identities of objects. Dev. Psychol. 32, 479–491


33 Miller, S.A. et al. (1997) Children’s understanding of false beliefs that result from developmental misconceptions. Cogn. Dev. 12, 21–51


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


38 Markman, R.M. (1979) Realizing that you don’t understand: elementary school children’s awareness of inconsistencies. Child Dev. 50, 643–655


43 Harris, P.L. et al. (1986) Children’s understanding of the distinction between real and apparent emotion. Child Dev. 57, 895–909

1 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.

5 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?’

9 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?’

13 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?’

17 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?’

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.’

‘What kind of thing is a Dalmatian?’

C: ‘A dog…(It’s a) fire-engine dog.’

…

‘Now watch this.’ (E puts puppet on hand; pretends to talk with it) ‘What do you call something that does this?’

C: ‘Puppet.’

…

‘Is it a dog and a cat?’

C: (Shakes head)

‘Is it a cat and a puppet?’

C: (Shakes head)

‘Is it a dog and a puppet?’

C: ‘Yes.’

(Taken from a study described in Ref. [25].)

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?
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].

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].