

## The Effects of Question Clarification and Memory Aids on Young Children's Performance on Appearance-Reality Tasks

John H. Flavell  
Frances L. Green  
Kelly E. Wahl  
Eleanor R. Flavell  
Stanford University

In two studies 3-year-olds were given tasks in which they saw an object placed under a color filter and were asked to identify its real, no-filter color and its apparent, seen-through-the-filter one. Making the request for the real color more concrete, simple, and clear did not reduce the children's strong tendency to indicate the apparent color when asked for the real one. Providing visual reminders of the object's real color did reduce this tendency, although only in one study and for one set of tasks. These results support the conclusion from previous studies that young children have genuine, deep-seated, and hard to remedy conceptual difficulties with the appearance-reality distinction.

Three-year-olds are first given brief pretraining on the meaning and linguistic expression of real (nonillusory) versus apparent (illusory) color, using cardboard cutouts of familiar objects and color filters that change the objects' apparent colors. After that they see, for example, a light blue paper cutout of a rabbit slowly moved beneath a dark green filter that makes the rabbit look green. Then they are asked the sorts of color appearance and reality questions on the meaning of which they have just been pretrained: appearance—"When you *look* at that rabbit right now, does he *look* green or does he *look* blue?" reality—"What color is the rabbit *really and truly*? Is he *really and truly* blue or is he *really and truly* green?" There is now considerable evidence that many 3-year-olds perform poorly on tasks like this, and perform about equally poorly on similar tasks involving real versus apparent size, shape, object identity, object presence, actions, and emotions (Braine & Shanks, 1965a, 1965b; Flavell, 1986a, 1986b;

---

This research was supported by National Institute of Mental Health Grant No. MH 40687. We are grateful to the children, teachers, and parents whose cooperation made these studies possible.

Correspondence and requests for reprints should be addressed to John H. Flavell, Department of Psychology, Stanford University, Stanford, CA 94305.

Flavell, Flavell, & Green, 1983, 1986; Flavell, Green, & Flavell, 1986; Flavell, Zhang, Zou, Dong, & Qi, 1983; Harris, 1986; Harris, Donnelly, Guz, & Pitt-Watson, 1986; Taylor & Flavell, 1984). As would be expected, they usually err on color tasks by responding with the apparent color (here, green) to both questions.<sup>1</sup>

How this poor performance should be interpreted is of course the crucially important question. It might reflect either a genuine, deep-seated inability to understand and think about the appearance-reality distinction or it might reflect only task insensitivity. Thus far, the research evidence supports the genuine-inability conclusion (Flavell, 1986a). Beijing preschoolers perform just like northern Californian preschoolers on translated versions of these tasks, despite differences in language and childhood experiences. Efforts to make the tasks less demanding and more sensitive to rudimentary appearance-reality knowledge have not thus far resulted in improved task performance, nor have intensive efforts to teach the appearance-reality distinction.

These studies represent further attempts to reduce the performance problems and to increase the sensitivity of color appearance-reality tasks. We tried to liberate any nascent understanding 3-year-olds might have by clarifying the reality question (the one they usually get wrong) and by helping them keep in mind the object's original (real) color.

The principal research objective was to find a way to help children understand the experimenter's request for the real color of the object. The present method is, arguably, "too verbal" for 3-year-olds. Despite pretraining on the intended meaning of the expression, "really and truly," children of this age might misinterpret it as meaning "very" or "definitely." Such misinterpretation would certainly be understandable: "Really and truly" often does have those connotations and the apparent color is indeed very and definitely green. Also, the two questions are rather abstract, complex, and similar in form. Perhaps repeatedly hearing in alternation two such similar and verbose questions causes 3-year-olds to tune out everything but "blue?" and "green?" Thus, what has appeared to be a genuine conceptual limitation might actually ("really and truly") reflect only insufficient word knowledge or listening skills. This is a frequently encountered and seemingly very plausible objection to the existing research on this topic, and consequently one we wanted to evaluate empirically.

These possible assessment problems could be gotten around by using the following procedure. Put a stimulus, for example, a blue cutout of a rabbit, behind the filter, as usual. Press out a precut piece from the center of the rabbit and remove it from under the filter in one's closed hand, leaving the rest of the

<sup>1</sup> Unlike the case with its illusory apparent size or shape, an object's apparent color as seen through a filter is, properly speaking, its "real color." However, only adults who know something about color vision realize this. Our research experience indicates that other adults, and most children older than 4 or 5 years of age, routinely think of it the way they think of apparent size and shape—namely, as a momentary illusory appearance rather than as an enduring reality.

rabbit under the filter. Place on the table that blue piece together with another piece of the same size and shape, the green color of which approximates the green apparent color of the rabbit as seen through the filter. Then simply ask the child which is the piece that was just removed from the rabbit. Variations on this procedure were used in these studies. This question avoids the first problem by clearly referring to the real, filter-independent surface color of the rabbit without requiring an understanding of “really and truly.” It circumvents the second one by posing a simple, concrete, nonwordy question that sounds quite different from the appearance question. If misunderstanding of “really and truly” or other problems with the “standard” questions contribute to young children’s poor performance, their performance should improve if this more straight-forward “semantically transparent” questioning procedure is used.

The secondary research objective was to provide visual reminders of the object’s real color, its color before being put behind the filter. If young children have trouble keeping the object’s original color in mind when the questions are asked, such reminders might improve performance.

Finally, we probed briefly and minimally for 3 competencies that seemed particularly relevant to the new, semantically transparent method of requesting the object’s real color. These competencies are (a) recall of the object’s original prefilter color (e.g., blue); (b) knowledge that the object will look blue again when the filter is taken away; and (c) knowledge that its actual surface is still blue even while residing under the filter, and consequently that it would look blue rather than green if one viewed it other than through the filter. Competency *b* seems to require Competency *a*: The object would not look blue again if it had not originally been blue. The ability to know which piece came from the object, assessed by the semantically transparent question, probably requires Competency *b*, and therefore also Competency *a*: The piece was originally blue (Competency *a*); therefore it would be blue when it came out (Competency *b*); therefore this blue piece is the correct choice (semantically transparent question). However, the ability to pick the right piece would not necessarily require Competency *c*. That is, a child might vaguely believe that, even though the filter did not touch the object, by being put close to it the filter somehow made the object’s surface become temporarily blue, as if blue paint had been applied to its surface. On the other hand, possession of Competency *c* appears to require Competencies *a* and *b* and also should help make the semantically transparent question easier: The right piece is the one that matches in color the object under the filter, which is represented as still having a blue-colored surface while under there (Competency *c*). Competency *c* also seems to require a good understanding of the appearance-reality distinction (it looks green but is really blue underneath the filter) plus some perspective-taking ability (it looks green from here but would look blue if one viewed it from the other side of the filter).

We tested for Competency *b* (and therefore also *a*, presumably) by asking subjects which color the object would look if the filter were taken away. We

probed Competency *c* by asking which color it would look if they peeked under the filter (Study 1) or viewed it from the opposite side of the filter (Study 2). In a previous study (Flavell, Green, & Flavell, 1986, Study 1), 80% of a sample of 3-year-olds responded correctly on at least 2 of 3 Competency *b* probes; we therefore expected good performance on this probe. No previous assessment of Competency *c* has been made, but these considerations suggest that 3-year-olds should find the Competency *c* probe quite hard. These 2 probes were followed by a third, in the form of a standard “really and truly” reality question. We expected this probe to be harder than the Competency *b* probe and of uncertain relation to the Competency *c* probe.

### STUDY 1

In this study each child was first briefly pretrained and pretested, and then given a series of 8 tasks followed by 3 probes. In 6 of these tasks the children were asked only a semantically transparent reality question. For example, they were asked which of two pieces, if either, had just been removed from the rabbit, still visible under the filter with a hole where the missing piece had been. In 3 of these 6 (Pieces tasks), an object was under the filter and the 2 pieces were outside, as in the example just given. In the other 3 (Objects tasks), a piece was under the filter and 2 objects that the piece might belong to were outside. The 7th task (Semantically Transparent AR) was also of the Pieces type but the children were asked an appearance question in addition to a semantically transparent reality question. It could thus be directly compared with the 8th (Standard AR), a standard appearance-reality task that likewise included both an appearance and a (standard) reality question. Thus, there were 4 types of tasks, 2 of which were represented by 3 tasks each and the other 2 by only 1 task each.

The 3 probes were always given last. One probe tested for Competency *b* and one for Competency *c*; the third consisted of a standard reality question. In all tasks and probes subjects were provided with a visible reminder of the stimulus's real color.

### METHOD

#### Subjects

The subjects were 24 nursery schoolchildren (14 females, 10 males) drawn mostly from upper middle-class families. They ranged in age from 2;11-4 to 2 years with a mean of 3;8 years. All subjects were tested individually in a single session by the same female experimenter.

#### Materials

In all tasks and probes children viewed a cardboard cutout of a stimulus (an object or a square, 2.5 × 2.5 cm object piece) through a color filter that produced

a change in its apparent color. Each filter was mounted on a 20 × 25 cm plastic picture frame tilted 15 degrees from the horizontal such that the surface of the filter was approximately 4 cm above the stimulus (thus, clearly not touching it). Each stimulus to be put under the filter was taken from a stack of cutouts. The cutouts were of different shapes but were all the same color as the stimulus. The stack, clearly visible about 26 cm to the right of the filter throughout the task, thus provided a potential reminder of the object's or piece's original color. Only one stack was present on each task.

### Pretest and Pretraining

Each child received the same pretraining and pretesting in the following order. The child was shown a black triangle under normal viewing conditions. The experimenter said, "I'm going to take a piece out of this triangle." She hid the black square piece in her hand, placed a blue square piece and a white square piece in front of the child, and asked; "Where is that piece (points to the empty space in the triangle) that I took away? Is it here (points to the pieces) or is it someplace else?" After the child's response she said, "That's right/actually, I had it some place else. Here is that piece. It is just the same color as that triangle." The experimenter then said for each of the white and blue pieces, "This isn't the right piece because it's a different color," and held the piece to the triangle for comparison. "This (black piece) is the right piece because it is just the same color as the triangle."

The purpose of this first part of the pretraining was to give the children experience in correctly stating, "someplace else" (20 of 24 did so) and to explain that pieces and objects should be the same color. The purpose of the second part was to teach them to match for real rather than apparent color, and to acquaint them with the words used on the appearance-reality tasks. The experimenter removed a pink square from a pile of pink memory aids saying, "Here is a piece of something. Watch." The piece was moved slowly behind a green filter, withdrawn, then placed behind it once more on a black ground. The appearance of the piece was green. Saying, "These ducks lost a piece," the experimenter introduced identical-size pink and green ducks. The child was asked: "Does that piece belong to one of these ducks (points) or does it belong to some other duck?" Regardless of the child's accuracy (only 12 of 24 correctly chose the pink one), the correct answer was demonstrated by lifting the filter away from the piece and by bringing the piece next to the pink duck. The experimenter said, "That's right/actually, it belongs to this duck." The filter was replaced and the experimenter continued: "Right now, that piece *looks* green. It *looks* green but it isn't *really* and *truly*. *Really* and *truly* it is *pink*." The experimenter then lifted the filter to demonstrate its actual color.

All of the children were then asked to provide the correct color labels for the colors used in this study. Somewhat surprisingly, no subject made a mistake on this pretest. Therefore, no subject was dropped from the study.

### Procedure

The 4 types of task are described here. The first 2 types, Pieces and Objects, have corresponding tasks. Pieces 1, 2, and 3 are similar to Objects 1, 2, and 3, respectively. In Pieces 1 and Objects 1, the correct choice is present but children will err if they are wont to match by apparent color, because the other available choice is of that apparent color. In Pieces 2 and Objects 2, the correct choice is also present but there is no potentially tempting apparent-color match, because the other available choice is not of the apparent color. We thought that some children who err when a perceptual match is possible might make the correct choice when it is not. To find out whether such correct choices would be due to chance or to latent knowledge about "real color," we added Pieces 3 and Objects 3, tasks in which there is neither a correct nor a color-match choice. Children who have this latent knowledge should respond, "someplace else," on these tasks; children who do not should randomly select one of the two pieces or objects present. We also thought that the Objects tasks should be easier than the Pieces tasks for two reasons. First, mentally removing the piece from the filter to see which object it belonged to seemed a likely, natural strategy. Second, as noted in the discussion of Competency *b*, Flavell, Green, and Flavell (Study 1, 1986) found that many children who fail the "really and truly" questions can nevertheless report accurately the color an object will look when the filter is removed.

**Pieces.** The experimenter said, "Here is a (e.g., ball)," and manually demonstrated the real and apparent colors of the object as described in the second part of the pretraining (moved it slowly behind the filter, withdrew it, etc.). The object was placed under the filter with one edge resting on a small block. Saying, "I'm taking this piece away," the experimenter removed a square piece from the object, taking care that the subject did not see its real color. She then placed 2 colored squares in front of the child and asked: "Where is that piece (points to the hole in the object) I took away? Is it here (points to pieces) or someplace else?"

**Pieces 1:** A red ball appears black when seen through a green filter; the pieces are red (correct) and black (incorrect, perceptual match).

**Pieces 2:** A green ball appears black when seen through a red filter; the pieces are green (correct) and blue (incorrect, no match).

**Pieces 3:** A white star appears green when seen through a green filter; the pieces are red and dark blue (both incorrect, no match).

**Objects.** The demonstration of the piece and the test question were exactly as in the second part of the pretraining, "These (e.g., boxes) are missing a piece. Does that piece belong to one of these (boxes) or does it belong to some other (box)?"

**Objects 1:** A blue piece looks black when viewed through a red filter; the objects are blue (correct) and black (incorrect, perceptual match) boxes with missing pieces.

**Objects 2:** A red piece looks black when viewed through a green filter; the objects are red (correct) and light blue (incorrect, no match) tents with missing pieces.

**Objects 3:** A white piece looks red when viewed through a red filter; the objects are green and blue (both incorrect, no match) kites with missing pieces.

**Semantically Transparent AR (Appearance-Reality) Task.** A white flower appears blue when seen through a blue filter. The pieces are white and blue. The flower's real and apparent colors were demonstrated and a piece taken away, just as in Pieces 1. However, both appearance and reality questions were asked about the 2 pieces outside the filter. The test questions were as follows: "Right now, which piece *looks* the same color the flower *looks right now*?" and "Which piece is *really* the piece of that flower, *really and truly*?"

**Standard AR Task.** A blue rabbit appears green when seen through a green filter. Saying, "Here is a rabbit," the experimenter demonstrated its real and apparent colors as in other tasks. The test questions were: "When you *look* at that rabbit right now, does he *look* (green) or does he *look* (blue)?" and "What color is the rabbit *really and truly*? Is he *really and truly* (blue) or is he *really and truly* (green)?"

The 4 types of tasks were administered in all possible orders. Within Pieces and Objects each possible order of the 3 tasks was represented 4 times. Half of the subjects received the appearance question prior to the reality question on both the Semantically Transparent and the Standard AR tasks and half the reverse. The order of choices within questions was randomly determined in the Standard AR Task and in the 3 probe questions described next. The probes were given following the last task.

**Probe 1. Color When Filter Removed (Competency b).** A green turtle, demonstrated as all stimuli were, was made to appear black when seen through a red filter. The experimenter asked: "If I pick this thing up (the experimenter raised the filter slightly) and take it away, will that turtle look (black like he does now) or will he look (green)?"

**Probe 2. Color of Object Under Filter (Competency c).** A white butterfly, demonstrated as before, appeared red when seen through a red filter. The question was: "If you could peek under this thing (the experimenter raised the filter slightly), do you think that the butterfly would look (red like it does now) or would it look (white)?"

Half of the subjects received Probe 1 first, half Probe 2 first. Probe 3 was the final question asked all subjects; it utilized either the turtle or butterfly stimulus, depending upon which Probe (1 or 2) had just been given.

**Probe 3. Standard Reality Question.** Immediately following either Probe 1 or 2, whichever was last, the subject was asked: "What color is that (turtle or butterfly) *really and truly*? Is it *really and truly* (blue, white) or is it *really and truly* (green, red)?" Probe 3 was included to check the aforementioned finding (Flavell, Green, & Flavell, Study 1, 1986) that children may correctly report the color an object will look when a color filter is removed (Probe 1) but still give a phenomenal answer to a standard "really and truly" question about the object's real color while under the filter.

## RESULTS AND DISCUSSION

Table 1 shows the numbers of children making each of the possible responses to the 8 tasks and 3 probes. The most important result of this study is its failure to show that the semantically transparent form of the reality question is any easier for young children to answer correctly than the standard "really and truly" one. The group performed very similarly on the Semantically Transparent AR and Standard AR tasks, showing in both the usual predominant error pattern of giving the apparent color in response to the reality question as well as to the appearance question (10 "Only Appearance Correct" in both tasks). Likewise, the mean

**Table 1. Responses to Study 1 Tasks and Probes ( $N = 24$ )**

| Tasks                       | Response     |                         |                      |                 |
|-----------------------------|--------------|-------------------------|----------------------|-----------------|
|                             | Real Color   | Apparent Color          | Other Color          | Some Place Else |
| Pieces 1                    | <i>17</i>    | 3                       | —                    | 4               |
| Pieces 2                    | <i>11</i>    | —                       | 2                    | 11              |
| Pieces 3                    | —            | —                       | 7                    | <i>17</i>       |
| Objects 1                   | <i>13</i>    | 10                      | —                    | 1               |
| Objects 2                   | <i>17</i>    | —                       | 5                    | 2               |
| Objects 3                   | —            | —                       | 9                    | <i>15</i>       |
|                             | Both Correct | Only Appearance Correct | Only Reality Correct | Neither Correct |
| Semantically Transparent AR | 10           | 10                      | 3                    | 1               |
| Standard AR                 | 11           | 10                      | 1                    | 2               |
| Probes                      | Correct      | Incorrect               |                      |                 |
| Probe 1                     | 20           | 4                       |                      |                 |
| Probe 2                     | 13           | 11                      |                      |                 |
| Probe 3                     | 13           | 11                      |                      |                 |

*Note:* Correct responses to the Pieces and Objects tasks are italicized.



number of correct responses to semantically transparent reality questions on the 6 Pieces and Objects tasks was 15 (range 11–17), not much higher than the 12 correct responses to the standard reality question on the Standard AR task (11 “Both Correct” plus 1 “Only Reality Correct”).

There was also no evidence that children’s performance benefited from the continued presence throughout each task of a stack of same-color objects that could serve as a visible reminder of the stimulus’ original color. The experimenter had the strong impression that the children completely ignored the stacks once the target objects had been removed from them. Consistent with this impression, their performance on the Standard AR task ( $1\frac{1}{24}$  or 46% correct pairs of answers) was similar to that of previous groups of 3-year-olds on standard color appearance-reality tasks that provided no such memory aids: Flavell, Flavell, and Green (1983) Study 2—32%, Study 3—38%; Flavell, Green and Flavell (1986) Study 1—48%, Study 2—54%, Study 4—43%.

Also similar to previous findings (Flavell, Green, & Flavell, 1986, Study 1), on Probe 1, 20 of 24 or 83% of the children correctly said that the turtle would look green rather than black if the filter were removed (Competency *b*), and therefore undoubtedly remembered that it had been green originally (Competency *a*). As expected, however, they performed near-significantly ( $p < .10$ , Sign tests) more poorly on Probes 2 and 3 than on Probe 1 (54% each) suggesting that they were less clearly aware that the object retained its original surface color while under the filter (Competency *c*) and that the object was “really and truly” that original color.

The other, minor predictions were not confirmed by the data. The Objects tasks did not prove to be easier than the Pieces tasks. Likewise, the children did not perform substantially worse when an incorrect stimulus choice of the same color as the target’s apparent color accompanied the correct, real-color choice (Pieces 1 and Objects 1) than when an incorrect choice of some other third color accompanied the correct choice (Pieces 2 and Objects 2). They also did not always select the apparent-color choice when failing to select the correct choice: As Table 1 shows, they did so 10 of 11 times on Objects 1 but only 3 of 7 times on Pieces 1.

The results of this study should be taken as suggestive rather than definitive. First, although the children did not appear to use the memory aids provided them in this study, they might have used more salient, attention-catching reminders, and such use might have improved their task performance. Second, the failure to find better performance on semantically transparent reality questions than on standard ones could conceivably have been an artifact of inadequate assessment. For reasons that are not clear, subjects may have been less consistently on-task in this study than in previous ones. The failure of groups of 7 and 9 children to respond “someplace else” on seemingly easy Pieces 3 and Objects 3, respectively, suggests this possibility. So also does the surprising finding, atypical in

appearance-reality research, that children's performance on one task or probe did not predict their performance on another. As one example among many, although about as many children were fully correct on the Semantically Transparent AR task (10 children) as on the Standard AR task (11 children), only 6 were fully correct on both—thus, a zero-order correlation between the two tasks. Study 2 was designed to provide additional, more definitive results on these same questions.

## STUDY 2

The main objective of Study 2 was to provide a methodologically more adequate test of the possibility that 3-year-olds would perform better on appearance-reality tasks that utilize semantically transparent reality questions than on those that utilize standard, "really and truly" reality questions. Its secondary objective was to provide a better test of the possibility that helping 3-year-olds remember and keep in mind the object's original (real) color would also benefit performance.

We tried to achieve the first objective by using a somewhat larger subject sample, providing direct training for each of the 2 types of tasks, demonstrating that the actual surface of the object was not altered while behind the filter by having the child move around the stimulus display, reducing the number of different color transformations, including 4 tasks of each type, and administering each set of 4 tasks as a block. In addition, the same real and apparent colors were used in corresponding tasks of the 2 types in an attempt to make them as similar as possible.

We tried to achieve the second objective by making the object's real color more available to memory and attention than was done in Study 1. Each cardboard cutout object had extending horizontally from each side, like a wing, a thin strip of cardboard that widened at the end to form a trapezoid-shaped "handle" about half the size of the object. Handles, strips, and object were all of the same color and constituted a single, complex form 40 cm long. If this form were white and the filter green, for example, the object part of the form and the parts of the strips nearest to the object would look green when the form was placed under the filter, whereas the outer portions of the strips and the handles would look white because they extended beyond the edges of the filter. About half the surface area of the form looked white and about half green when it was placed behind the filter. For adults, at least, these white parts constitute an extremely powerful, hard to ignore reminder of the object's original, real color, especially having just seen the object put under the filter.

Finally, the same 3 probes used in Study 1 were administered at the end of the testing session. As in Study 1, Probes 1 and 2 were given in counterbalanced order and Probe 3 was always given last.

## METHOD

### Subjects

The subjects were 32 nursery schoolchildren (14 females, 18 males) drawn mostly from upper middle-class families. They ranged in age from 3-1 to 4-1 years, with a mean of 3-6 years. All subjects were tested individually in a single session by the same male experimenter.

### Materials

The filter setup and other materials were similar to those used in Study 1, except that the objects were outfitted with lateral strips and handles as just described.

### Pretest and Pretraining

Children were pretested for their ability to name all the colors used in this study. Of the 35 children who took the pretest, 32 performed perfectly and were retained as subjects.

Children were pretrained on each task type immediately prior to being given the 4 tasks of that type. In the pretraining for the Semantically Transparent AR tasks, the experimenter said, "Here is a boat. It has some handles so I can hold it. Watch." Then he slowly moved the boat (with handles) behind the filter, out again, then behind again so the child could see the changes in apparent color. The child then was asked to come around to see how the boat looked from the opposite, nonfilter side; the experimenter raised the filter slightly so that no reflective color shadows would be cast on the boat. This experience was included to show children that the filter did not somehow "coat" or "paint" the actual surface of the object. When the child returned to his or her seat the experimenter removed a precut piece from the boat in his closed hand and put on the table, in front of the filter, that piece plus another that matched the boat's apparent color. The child was asked which piece looked the same color as the boat looked right now, and then asked which piece the experimenter had just taken out of the boat. Corrective feedback and careful explanation were given after each question, much as in Study 1. The Standard AR pretraining was the same except that no piece was removed from the object (an egg) and the appearance and reality questions were the standard ones.

### Procedure

**Tasks.** The procedure was essentially the same as the pretraining, except that different objects were used, no feedback or explanation was given, and the children were not given the opportunity to walk around the display. The experimenter mentioned the handles on every trial, as in pretraining. The 4 tasks in each set were always given in the same order. The questions in the Semantically Transparent AR tasks were: "See the color the \_\_\_\_\_ looks right now (points to

object)? Which piece *looks* the same color the \_\_\_\_\_ *looks* right now?" and "Which is the piece I just took out of the \_\_\_\_\_ (points to object)?" The questions in the Standard AR tasks were "When you *look* at the \_\_\_\_\_ right now (points), does the \_\_\_\_\_ *look* X or does it *look* Y?" and "What color is the \_\_\_\_\_ (points) *really* and *truly*? Is it *really* and *truly* X or *really* and *truly* Y?" The displays used in the Semantically Transparent AR tasks were: a white kite that looked green through a green filter (Task 1), red ball—looked black—green filter (Task 2), white flower—looked red—red filter (Task 3), and green box—looked black—red filter (Task 4). Four different objects were used in the Standard AR tasks but their real and apparent colors and the filters used to produce their apparent colors were the same as in the Semantically Transparent AR tasks.

Half of the sample received the 4 Semantically Transparent AR tasks first and half the 4 Standard AR tasks first. Within each half, 8 subjects were asked the appearance question before the reality question on the first 2 tasks and the reality question before the appearance question on the last 2 tasks. The other 8 subjects received the reverse order of questions. Within a subject the order in which questions were asked and the ordering of the question choices within questions were identical for the 2 types of tasks. Orders of choices within questions were randomly assigned for each subject. The right-left location of the correct piece was counterbalanced.

**Probes.** As already noted, the 3 probes were essentially the same as in Study 1, except for a minor change in the wording of Probe 2: "If you could walk around here and look at it from where I am. . . ." rather than "If you could peek under this thing. . . ."

## RESULTS AND DISCUSSION

A 2 (type of task)  $\times$  2 (order of administration of the 2 types of tasks) ANOVA was performed on the numbers of tasks of each type in which both the appearance and the reality questions were answered correctly. The only significant finding was a main effect for type of task,  $F(1,30) = 6.00$ ,  $p < .02$ . The direction of that effect was very surprising. Not only did the children not perform better on the Semantically Transparent AR tasks than on the Standard AR tasks, they actually performed significantly worse:  $M = 2.16$  of 4 tasks wholly correct for Semantically Transparent tasks,  $M = 2.91$  for Standard tasks. Near-significant differences favoring Standard over Semantically Transparent tasks were found for both the appearance questions ( $M$ s of 3.91 vs. 3.59) and the reality questions ( $M$ s of 2.97 vs. 2.53). Table 2 shows children's patterns of answers to each of the four tasks of each type. The difference between the 2 types of 2 tasks is consistent from task to task: Children answered both questions correctly more often on each Standard task than on its Semantic Transparent counterpart. The

**Table 2. Responses to Study 2 Tasks and Probes ( $N = 32$ )**

| Tasks                       | Response       |                         |                      |                 |
|-----------------------------|----------------|-------------------------|----------------------|-----------------|
|                             | Both Correct   | Only Appearance Correct | Only Reality Correct | Neither Correct |
| Semantically Transparent AR |                |                         |                      |                 |
| Task 1                      | 15             | 16                      | 0                    | 1               |
| Task 2                      | 16             | 11                      | 5                    | 0               |
| Task 3                      | 18             | 12                      | 2                    | 0               |
| Task 4                      | 20             | 7                       | 5                    | 0               |
| Standard AR                 |                |                         |                      |                 |
| Task 1                      | 25             | 7                       | 0                    | 0               |
| Task 2                      | 23             | 8                       | 0                    | 1               |
| Task 3                      | 22             | 10                      | 0                    | 0               |
| Task 4                      | 23             | 7                       | 2                    | 0               |
| <b>Probes</b>               |                |                         |                      |                 |
|                             | <b>Correct</b> | <b>Incorrect</b>        |                      |                 |
| Probe 1                     | 27             | 5                       |                      |                 |
| Probe 2                     | 15             | 17                      |                      |                 |
| Probe 3                     | 24             | 8                       |                      |                 |

data are also orderly in other ways. The predominant error pattern in both sets of tasks is to get the appearance question right and the reality question wrong, the usual pattern for appearance-reality tasks involving illusory color. Double errors (neither correct), suggestive of random responding, are almost totally absent. Absence of random responding is also suggested by a Pearson product-moment correlation of .43 ( $p < .01$ ) between performance on the 2 types of tasks (cf. Study 1). These results, together with those of Study 1, clearly indicate that 3-year-olds do not perform better on color appearance-reality tasks that include a simple, concrete, seemingly hard-to-misinterpret reality question than on tasks that employ the standard, seemingly more ambiguous "really and truly" question.

The presence of the object handles as memory and attention aids seems to have benefited children's performance on the Standard tasks. Percentages of pairs of questions correct (i.e., tasks completely correct) are 73% for Standard tasks in this study versus 32–54% for standard color tasks in Study 1 and in previous studies. Consistent with this result, in a previous study (Flavell, Green, & Flavell, 1986, Study 1) 75% of a sample of 3-year-olds answered both questions correctly on a color test in which part of the target object itself (rather than its handles, as here) was visible outside of the filter. On the present Standard tasks, 69% of the children were fully correct on 3 of the 4 tasks and 53% were fully correct on all 4 tasks. This is clearly better performance than 3-year-olds have previously shown on standard color appearance-reality tasks. Notice that

the presence of the handles actually seems to have helped the children make the appearance-reality distinction rather than only increase the relative salience of the real color. If it had only increased salience, the children would simply have mindlessly reported the real color more often—to the appearance questions as well as to the reality questions. The virtual absence in Table 2 of entries under Only Reality Correct for the Standard AR condition shows that this did not happen. In accord with these results, and unlike what was observed in Study 1, the experimenter noticed some of the children glancing at or pointing to the handles when answering the reality question, especially in the Standard condition. On the other hand, it is not clear that the presence of the handles helped performance on the Semantically Transparent tasks. Percentages of pairs of questions correct averaged 54% for the four tasks of Study 2, as compared with 42% for the single Semantically Transparent task in Study 1. Considering only correctly answered reality questions, the averages are 63% in Study 2 and 62% in Study 1 (averaging together the data from the Semantically Transparent task and the similar Pieces 1 task).

Finally, performance on Probes 1 and 2 was similar to that observed in Study 1: Probe 1—83% and 84% correct in Studies 1 and 2, respectively; Probe 2—54% and 47% respectively. Having twice seen in pretraining that objects retain their original surface color while under the filter obviously did not help the Study 2 subjects answer Probe 2 correctly. Children performed better on Probe 3 in Study 2 than in Study 1 (75% vs. 54% correct), consistent with their better performance on the “really and truly” questions of the Study 2 Standard tasks (74% correct). In Study 2 children performed significantly ( $p < .02$ , Sign tests) better on both Probes 1 and 3 than on Probe 2.

Interestingly, there was little correlation between children’s answers to Probe 2 and their answers to either the Probe 3, the Standard, or the Semantically Transparent reality questions. This suggests that a number of children may have correctly identified an object’s “really and truly” color (Probe 3 and Standard reality question) or the color a piece of it would be when taken out from under the filter (Semantically Transparent reality question) as its original color, without also assuming that its surface retained that original color while under the filter (Probe 2). For instance, for half the sample Probe 3 immediately followed Probe 2 and addressed the same stimulus display. Of the 12 children in this subsample who correctly identified the object’s original color in response to Probe 3, 5 had just incorrectly said, in response to Probe 2, that the object would look the apparent color if viewed from the other side. Such children would presumably not be conceiving of the object as *simultaneously* having both an apparent color and a (different) real color in this situation, as most adults probably would. Perhaps children begin to understand the appearance-reality distinction by separating the appearance and the reality in time: It looks X now but it really is Y, except now. Only later can they think of them as holding simultaneously: It looks X now but it really is Y now. Watson (1986) has observed such a successive-to-

simultaneous transition in young children's understanding of social role relations. They were able to understand a person changing from, say, a father to a doctor before they were able to understand that the person could be in both roles simultaneously.

## GENERAL DISCUSSION

This investigation yielded two major findings. The first was that asking 3-year-olds a simple, concrete, and clear (Semantically Transparent) reality question instead of the possibly ambiguous, "really and truly" (Standard) one did not bring to light any hitherto unrevealed understanding of the appearance-reality distinction. Children either performed at the same level, on the average, when asked the two different kinds of reality questions (in Study 1) or, very surprisingly, actually performed worse when asked the Semantically Transparent one (in Study 2). The second finding was that provision of a highly salient, hard-to-ignore reminder of the object's original, real color benefited performance on Standard tasks but not, apparently, on Semantically Transparent ones (Study 2). Provision of less salient reminders did not benefit performance (Study 1).

The first finding bolsters the impression from previous studies that young children have genuine, deep-lying difficulties in thinking about the appearance-reality distinction. As mentioned in the introduction, previous studies have shown that giving them easy-looking tasks and providing them with systematic training does not reduce these difficulties (Flavell, 1986a; Flavell, Green, & Flavell, 1986). The present studies show that clarifying the reality question does not reduce them either. If subjects had clearly understood the appearance-reality distinction and had merely failed to comprehend the Standard reality question correctly, one would have expected them to do well on the Semantically Transparent forms of the appearance-reality task. The fact that they did not do well on them suggests that they did not clearly understand the distinction. Thus, the first finding is important because it suggests that existing appearance-reality tasks assess conceptual competencies rather than just semantic or communicative ones.

There remains the question of why the children so often chose the wrong piece when asked the Semantically Transparent question. Probably because, not clearly understanding the distinction between real and apparent color, choosing a piece that matched the object's visible apparent color was an easy, "default" response for 3-year-olds to make. The object under the filter had a conspicuous hole in it and the incorrect piece was the only one that looked like it belonged in that hole. Moreover, their poor performance on Probe 2 suggests that they implicitly assumed that the object's surface while underneath the filter bore the incorrect piece's color (i.e., bore the apparent rather than the real color). It is possible that the handles in Study 2 were of little help when the Semantically Transparent reality question was asked because that question ("Which is the

piece I just took out of the \_\_\_\_\_ (points)?') may have directed children's visual attention toward the 2 pieces and the object and away from the peripheral handles. In the Standard condition, on the other hand, there were no pieces present to direct attention away from the handles.

It is not clear to us precisely what role the handles played in improving performance on the Standard tasks. We can surely rule out one possibility, however: On hearing the reality question, the children spontaneously thought of retrieving the object's original color but simply could not without seeing it on the handles. Their very good performance on Probe 1 shows that they usually could remember it if induced to try, whether the handles were present (Study 2, 84% success) or not (Study 1, 83% success). Appearance-reality tasks intended for use with young children are of course designed to make deliberate retrieval of the object's real property or identity very easy for children who know to try. For example, there are always only two options, the real and the apparent; both real and apparent are always named in each question, making the retrieval task one of recognition memory only; and the real has just recently been visible. For young children who understand the task and realize that the answer wanted is the original, real color, it is hard to see how recognizing its name could be a significant problem.

However, the handles might be helpful for young children whose understanding is minimal and who therefore do not spontaneously try to recall the original color. They actually see the original color on the handles and outer ends of the strips, and also see that handles, strips, and object form a single connected whole. This special setup, unlike the appearance-reality situations life usually presents, may impel them to think of the object as having two color descriptions: Apparent Color *X* when looked at as a separate object; Real Color *Y* when considered as part of the whole, the other parts of which still display this color. Whatever the exact process, we tentatively conclude that this setup somehow helped children assemble a genuine but temporary appreciation of the distinction between apparent and real color. The appreciation was probably genuine, given the high frequency of Both Correct responses and the low frequency of Only Reality Correct responses (Table 2); as indicated previously, merely becoming more attentive to the real color would not produce that pattern. It was also probably only temporary, however. The data showed that the 16 children who had just experienced the Standard tasks actually performed slightly worse ( $M$  of 2.16 vs. 2.25) on the subsequent Semantic Transparency tasks than the 16 who were given these latter tasks first. Had the Standard tasks induced or released any durable understanding in the former subgroup, it probably would have shown up in better performance on the immediately following Semantically Transparent tasks. Recall again that previous attempts to engender durable understanding through direct training have likewise not been successful (Flavell, 1986a).

These findings taken together with previous ones suggest that heroic measures are needed to bring about even a temporary understanding of the appearance-



reality distinction in many 3-year-olds, and that even heroic measures may not suffice (e.g., previous training studies). Why is it so hard for them to grasp and hold onto the distinction? We have argued that the root cause is an immature conception of the mind and how it relates to the world (for details, see Flavell, 1986a, 1986b). According to this argument, young children do not clearly understand that a single external stimulus (object, event, property, etc.) may be simultaneously representable in two or more mutually contradictory ways, not just in pretend play and fantasy but also in serious cognition. But this understanding is precisely what is needed when thinking about appearances versus realities: One has to seriously, nonplayfully construe the very same stimulus as simultaneously "being," in a sense, two incompatible things, for example, both all blue and all green. Because young children do not easily imagine the possibility of multiple different, simultaneously applicable representations of single stimuli, they are likely to cope with appearance-reality tasks by simply focusing on the single most salient identity or property the stimulus presents and giving that as their response to both questions; at best, they may distinguish appearance and reality but separate them in time, as the Probe 2 data suggested may sometimes happen. As children come to understand that the same thing can be simultaneously represented in contrasting ways, within the same person as well as by different people, the way is paved for the acquisition of appearance-reality, perspective-taking, and other important metacognitive and social-cognitive competencies (Flavell, 1986b).

## REFERENCES

- Braine, M. D. S., & Shanks, B. L. (1965a). The development of conservation of size. *Journal of Verbal Learning and Verbal Behavior*, *4*, 227-242.
- Braine, M. D. S., & Shanks, B. L. (1965b). The conservation of a shape property and a proposal about the origin of the conservations. *Canadian Journal of Psychology*, *19*, 197-207.
- Flavell, J. H. (1986a). The development of children's knowledge about the appearance-reality distinction. *American Psychologist*, *41*, 418-425.
- Flavell, J. H. (1986b, May). *From Level 1 to Level 2 knowledge about the mind*. Paper presented at the International Conference on Developing Theories of Mind, University of Toronto, Canada.
- 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. (1986). *Young children's knowledge about the apparent-real and pretend-real distinctions*. Unpublished manuscript, Stanford University, Stanford, CA.
- 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, Serial No. 212).
- 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.
- Harris, P. L. (1986, June). *Children's understanding of real and apparent emotion*. Paper presented at the workshop on children's early concept of mind, Oxford University, England.

- Harris, P. L., Donnelly, K., Guz, G. R., & Pitt-Watson, R. (1986). Children's understanding of the distinction between real and apparent emotion. *Child Development, 57*, 895-909.
- Taylor, M., & Flavell, J. H. (1984). Seeing and believing: Children's understanding of the distinction between appearance and reality. *Child Development, 55*, 1710-1720.
- Watson, M. W. (1986). The breadth of the appearance-reality distinction. Commentary on Flavell, Green, and Flavell's development of knowledge about the appearance-reality distinction. *Monographs of the Society for Research in Child Development, 51*(1, Serial No. 212).