Mental image reversal and verbal recoding: When ducks become rabbits

MARIA A. BRANDIMONTE and WALTER GERBINO
University of Trieste, Trieste, Italy

A recent debate has concerned whether classical bistable configurations (e.g., duck/rabbit) can be reinterpreted using mental imagery. Research in this field indicates that image reversal is possible only when subjects change their specification of orientation. In a series of four experiments, we demonstrate that mental reversal of classical bistable configurations (CBCs) is impeded by verbally recoding the visual pattern at the time of input. When subjects were prevented from verbally recoding visual stimuli in short-term memory, they fared systematically better in mentally reversing the CBC, even when they received no instructions to change their reference frame or specification of orientation. On this basis, we suggest a model of image reversal that takes into account the interaction between memory codes and provides a new perspective on verbal recoding, verbal overshadowing, and mental discoveries.

Visual ambiguities indicate that the stimulus–percept relationship is interpretive in nature. Classical bistable configurations (CBCs), such as the Necker cube, are graphical patterns that support two incompatible percepts. Each percept corresponds to a particular description of the pattern (not necessarily a verbal one) involving the specification of part-whole relationships (Rock, 1983) and the assignment of top, bottom, and sides. The occurrence of analogous ambiguities in the domain of mental images has important theoretical implications. This article will explore some conditions under which image ambiguities can be revealed.

Image Reversal

Some experimental findings suggest an extreme position: Ambiguities of the kind discussed above do not exist in mental imagery. Chambers and Reisberg (1985) asked subjects to form a mental image of a CBC, the duck/rabbit by Jastrow (1900), when only one interpretation was experienced during perception. Despite training to reverse other bistable figures in perception, and despite prompting, no subject reversed the mental image of the CBC. Chambers and Reisberg explained the failure to reverse in mental imagery by hypothesizing that the mental image of a CBC is inherently unambiguous. Un-

like the stimulus, which requires an interpretation, the image embodies a specific interpretation. That is, an image is created to represent a particular object, and hence its "meaning" is in place at the outset. Furthermore, they argued that shape representations are accessed by means of semantic representations and that the two cannot be separated. Recent studies on auditory imagery support a converging line of evidence (Reisberg, Smith, Baxter, & Sonenshine, 1989).

However, most recently Reisberg and Chambers (1991) admitted that subjects can nonetheless learn from their images by changing some aspects of the interpretation. According to the authors, this conclusion would not contradict the assumption that mental images are "meaningful depictions" that, as such, are already interpreted. Indeed, because images preserve appearance, one can talk about what the image resembles and what it evokes. Many different associations become available. As a consequence, by demanding a change in the way the image is understood, the mode of representation will lead to a different set of associations. For example, if subjects change their specifications of orientation or their understanding of the figure-ground relationship, many are able to find target shapes in their images. Reisberg and Chambers (1991) conclude that what subjects can discover in a mental image is bounded by how the subjects understand the image. According to the authors, this understanding of the represented form seems to govern both what the form will be seen to resemble and what it is likely to call from memory. This recent position is much closer to that of other authors who have suggested that, given suitable conditions, subjects can discover new construals emerging from mentally transformed images (Brandimonte, Hitch, & Bishop, 1992a, 1992b, 1992c; Finke, 1989, 1990; Finke, Pinker, & Farah, 1989). However, these studies did not consider the reversals of bistable configurations, but rather other mental discoveries, like those resulting from the

combination or separation of forms representing familiar objects, geometrical shapes, or letters.

A distinction between different types of reversals was recently proposed by Peterson, Kihlstrom, Rose, and Glisky (1992). They distinguish between reversals that entail assigning a new interpretation to the image components only (*reconstruals*) and reversals that imply a change in the specification of orientation (top/bottom, back/front of the image) as understood by the subject (*reference-frame realignments*). Reconstruals were shown to occur more readily in imagery than reversals entailing reference-frame realignments (Peterson et al., 1992). Reversals of the figures used by Finke et al. (1989) and by Brandimonte et al. (1992a, 1992b, 1992c) can be regarded as predominantly reconstruals, given that either instructions made clear the kind of reference frame to employ for reinterpretation (Finke et al., 1989) or no change in the specification of the top/bottom or front/back was necessary in order to reinterpret the image (Brandimonte et al., 1992a, 1992b, 1992c).

In recent work, Hyman and Neisser (1991) showed that subjects can mentally reverse even bistable configurations, such as the duck/rabbit figure, if they are given the "right kind" of instructions. Hyman and Neisser gave their subjects both orientation and category information about the alternative interpretation of the figure. Namely, they told subjects to consider the back of the head they already saw as the front of the head of a different animal. Under this condition, 11 out of 20 subjects reversed the figure in mental imagery. Hyman and Neisser's findings emphasize the role of instructions in prompting CBC reversal. Similarly, Peterson et al. (1992) found that this kind of figure can be reversed in imagery when subjects are prompted by either explicit suggestions (e.g., hints to change their reference frame) or implicit suggestions (e.g., more appropriate demonstration figures). In addition, they showed that the rate of CBC reversal can be increased by manipulating the quality of the image components (good parts vs. poor parts of the image).

To summarize, a growing body of experimental evidence suggests that the answer to the original question asked by Chambers and Reisberg (1985), "Can mental images be ambiguous?" is "It depends" (Hyman & Neisser, 1991). It seems, therefore, more productive to investigate the conditions under which CBCs *can* or *cannot* be reversed in mental imagery than to search for a yes/no answer.

Imagery and Memory

A prominent theory of the last few decades viewed imagery in relation to memory processes (Paivio, 1971, 1975). Much of Paivio's research addressed the question of *how imagery affects memory*. Imagery was, therefore, regarded as a mediator of memory performance.

More recently, the computational approach to imagery (Kosslyn, 1980), although mainly concerned with the nature of mental images, has nonetheless taken into account the role of memory processes in image generation. In particular, Kosslyn's research has been mainly concerned with long-term memory (LTM), in the sense that the tasks required subjects to generate an image by retrieving information from LTM. These studies greatly contributed to our understanding of *how memory supports imagery*.

A different way of looking at the relationship between imagery and memory has recently been proposed by Brandimonte et al. (1992a, 1992b, 1992c). They explored *how memory affects imagery* by studying people's ability to manipulate and transform visual mental images. In particular, they investigated whether and how the kind of short-term memory (STM) code (visual or verbal) used during initial learning of novel visual stimuli affects LTM encoding and, hence, subsequent image manipulation and processing. To clarify the role of STM codes on visual image processing, Brandimonte et al. (1992a) exploited the articulatory suppression (AS) technique (see, e.g., Baddeley, 1986; Murray, 1968; Slowiaczek & Clifton, 1980) and varied the nameability of the stimuli. In one study, subjects first learned a series of either easy-to-name or difficult-to-name pictures. Then, they were required to form an image of each picture and mentally subtract a specific part from it in order to discover a new construal. Half the subjects engaged in AS during initial learning, while half remained silent. Results showed that AS improved performance when the stimulus was easy to name, but had no effect when the stimulus was difficult to name. In the latter case, mental discoveries were as likely to occur with AS as without AS. In addition, performance in these conditions was as good as that in the condition using easy-to-name pictures with AS (Brandimonte et al., 1992a). These findings support the view that recoding visual stimuli into a phonological/verbal format in STM is detrimental when the task requires visual analysis. This hypothesis was confirmed by another study in which AS was not used, but verbal labels were supplied to both easy-to-name and difficult-to-name pictures. This manipulation led to the opposite pattern of results: Performance in the imagery task was severely disrupted by labeling for difficult-to-name pictures but not for easy-to-name pictures (Brandimonte et al., 1992b). Furthermore, performance in the condition using easy-to-name pictures without labels was as low as that in the condition using difficult-to-name pictures with labels.

The effect of AS strongly suggests that the encoding of visual stimuli, when they are easy to name, is constrained by the verbal component of the working memory system. The predominance of verbal encoding in STM affects LTM encoding and, therefore, the subsequent processing of visual mental images (e.g., Brandimonte et al., 1992a). In the next sections, we introduce the basic concepts and assumptions used throughout the article.

Verbal and Visual Codes

A great deal of research has provided evidence for two distinct modes of representation, visual and verbal, corresponding to different ways of encoding information in both STM (e.g., Baddeley, 1986) and LTM (Kosslyn,

1980; Paivio, 1971). In the present article, we will discuss the role of STM codes on image reversal. The term "verbal STM code" will be used as a synonym of the phonological code of the working memory system, which is active while maintaining and elaborating speech-based information (Baddeley, 1986; Baddeley & Hitch, 1974). Similarly, the term "visual STM code" will refer to the code that is active in the working memory system when loading and maintaining an image into the visual STM store. We also assume that verbal encoding in STM automatically prompts the verbal/semantic code of LTM, as when reading for meaning.

STM Verbal Recoding and
Verbal Overshadowing

In both STM and LTM, the visual and verbal subsystems have been regarded as interconnected, given that information can be transferred from one to the other very easily. The shift from the visual STM code to the verbal STM code will be termed "STM verbal recoding." Indeed, working memory research and visual imagery studies established that auditory inputs automatically access the phonological loop (Baddeley, 1986), whereas visual inputs automatically access the visual buffer (the visuospatial sketch pad, to use Baddeley's language; see also Kosslyn, 1980). The existence of this early stage does not preclude the visual recoding of verbal material (whenever this is possible) or the verbal recoding of visually encoded material (see also Posner, 1969; Posner, Boies, Eichelman, & Taylor, 1969). An important implication of this assumption is that, after recoding, part of the visual information may be lost (Bahrick & Boucher, 1968; Nelson & Brooks, 1973). A further implication is that verbal recoding of a visual stimulus may be the basis for the "overshadowing" of the visual code by the verbal code.

The term "verbal overshadowing" has been recently used by Schooler and Engstler-Schooler (1990) to describe a kind of LTM recoding interference that does not eradicate the original visual memories, but makes them temporarily unaccessible. The overshadowing effect occurs when a weaker representation is made ineffective by a simultaneous stronger representation. In the present research, verbal overshadowing will indicate an effect of interference stemming from verbally recoding the visual input in STM.

Loading Versus Generating Visual Mental Images

It is commonly assumed that visual imagery involves operating on mental representations that are temporarily maintained in the visual buffer (Kosslyn, 1980). Information can be entered into the visual buffer in two ways: (1) by generating a mental image from information stored in LTM (Finke et al., 1989; Kosslyn, 1980) or (2) by loading new visual inputs from the eyes (Kosslyn, Brunn, Cave, & Wallach, 1984; Kosslyn, Holtzman, Farah, & Gazzaniga, 1985). While there may be a functional equivalence between generated and loaded images, in that

an image generated from LTM can be as rich as one just loaded into the visual buffer, the two processes seem to differ widely. For example, generated images or parts of them can be refreshed any time they start fading, while loaded images may become unretrievable (Brandimonte et al., 1992a). Verbal recoding in STM might be seen as an attempt to fixate at least the content of the picture, given that the verbal code is less consuming. As suggested by Brandimonte et al. (1992a, 1992b), verbally recoding an image just loaded in visual STM may disrupt the particular identity of the representation by prompting the generation of an image that is conceptually driven (based on a name) rather than visually driven (based on the visual characteristics of the original stimulus). According to Kosslyn's (1980) theory of imagery, images can be generated by transferring information from LTM into the visual buffer via propositional and/or literal encodings. We assume that images, as soon as they enter the visual buffer from the eyes, are typically recoded in a verbal format and that the consequence of this is the generation of another image from LTM on the basis of a name, unless people are prevented from verbally recoding the picture.

STM Verbal Recoding and Image Content

In studies using CBCs to investigate reinterpretation of mental images, experimenters attempted to bias subjects toward the visual code in several ways. Subjects were trained to reverse during inspection of different CBCs and systematically instructed to form a "mental picture" of the stimulus. In addition, to prevent subjects from altering their images, experimenters showed the glasses/barbells figure from Carmichael, Hogan, and Walter (1932) to illustrate possible distortions occurring in picture memory. Interestingly, despite instructions to capitalize on the visual code and prompting, few subjects were able to reverse the image. We suggest that verbal recoding spontaneously activated during initial inspection of the picture may have impaired people's ability to mentally reinterpret their image.

A number of arguments may be advanced in favor of this hypothesis. First of all, instructing subjects to form a "mental picture" of the stimulus may not be as stringent as informing them about the precise task that is to be performed (Brandimonte et al., 1992a, 1992b, in press). Similarly, warning subjects to avoid memory distortions when encoding their images does not ensure that they will actually succeed in doing this. Indeed, as reported in previous studies (Chambers & Reisberg, 1985; Peterson et al., 1992), results were unaffected by presence versus absence of the demonstration figure from Carmichael et al. (1932).

A number of studies have shown that concurrent verbal processing of visual stimuli can disrupt memory for their visual characteristics (Bahrick & Boucher, 1968; Carmichael et al., 1932; Nelson & Brooks, 1973) as well as performance in visual imagery tasks (Brandimonte et al., 1992a, 1992b). The most common interpretation of these results is that verbal processing reduces the

amount of encoded visual information, hence affecting the quality of the visual representation. This claim rests on the general assumption that a limited pool of resources can be used for initial encoding and that, in effect, instructions to verbalize impose a dual task upon subjects, who need to devote part of their resources to verbal encoding at the expense of visual encoding (Bahrick & Boucher, 1968). A similar interpretation, proposed by Nelson and Brooks (1973), suggested that the involvement of the verbal system may have subtracted time from the encoding of the visual aspects of the stimulus.

To be successful, image reversal requires visual analysis during initial encoding. If subjects spontaneously recode the picture in a verbal form when loading the image, it seems likely that the image content will differ accordingly with the kind of encoding processes. Therefore, any manipulation aimed at emphasizing visual encoding should facilitate image reversal of CBCs such as the duck/rabbit. One such manipulation is to prevent subjects from verbally recoding the picture at the time of presentation. Articulatory suppression is a means to gain this end. Accordingly, in the four experiments reported below, we investigated some conditions under which "ducks" can become "rabbits."

Overview

In this article, we address the question of whether people can reverse the duck/rabbit figure, provided AS prevents them from verbally recoding the picture during loading. Experiment 1 tests the hypothesis that adults tend to spontaneously rely on a verbal form of encoding when loading an image into the visual buffer. In an attempt to replicate and extend Hyman and Neisser's (1991) results, Experiment 2 examines the effects of stimulus exposure time and AS, comparing 2-sec versus 5-sec presentations. Experiment 3 analyzes the effect of AS when full (category and orientation), partial, or no information is given. Finally, Experiment 4 attempts to establish whether verbal recoding occurs during loading, during image manipulation, or during both stages.

EXPERIMENT 1

In this experiment, adults' performance was compared with that of children aged 6 and 10, under the assumption that when processing visual materials in STM, older children and adults tend to rely on a verbal form of encoding, whereas younger children tend to use the visual code. This assumption is based on results from research on development of working memory. It has been shown that in immediate memory tasks young children (5-year-olds) use some form of visual storage to retain visually presented materials, whereas older children (10-year-olds) and adults tend to rehearse the materials using a speech-based articulatory loop (Hitch, Halliday, Schafstaal, & Schraagen, 1988).

Articulatory suppression is known to inhibit the use of a phonological code in STM (Baddeley, 1986), and there is evidence that it forces a reliance on visual STM (Hitch, Woodin, & Baker, 1989).

Within each age group, we manipulated AS; half of the subjects were required to remain silent while looking at the picture (no-AS condition), and half were asked to repeat irrelevant sounds ("la-la") (AS condition). If older subjects tend to verbally recode the visual stimulus during loading, then improved performance should be observed in the AS condition as compared with the no-AS condition. On the other hand, no effect of AS should be expected on young children's performance.

Method

Subjects. One hundred twenty subjects participated in the experiment (40 per age group). Subjects from each group were randomly assigned to the two experimental conditions (AS/no AS). Children were from two primary schools of Trieste; adults were undergraduates voluntarily participating in the experiment. No psychology student was included in the sample.

Materials. Three reversible figures were used (Figure 1). They were the rat/man by Bugelski and Alampay (1961), the young girl/old woman by Boring (1930), and the duck/rabbit by Jastrow (1900). The rat/man and the young girl/old woman served as training figures. The duck/rabbit was the test figure, and it was prepared
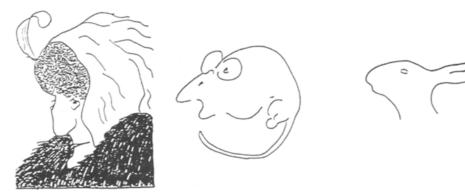

Figure 1.

in both right and left orientations. The experimenter used a stopwatch to control stimulus exposure time.

Procedure. The subjects were tested individually, the children in a quiet room of the school and the adults in a laboratory of the Department of Psychology at Trieste University. The procedure was modeled after the Chambers and Reisberg (1985) and Hyman and Neisser (1991) procedures.

In the training phase, the subjects were shown the rat/man and the young girl/old woman figures and were asked to find and identify both interpretations of each CBC. The order of presentation of the two training figures was counterbalanced across subjects. Then, they were required to describe both construals and to specify which features of the figure made, in their opinion, the reversal possible (Peterson & Hochberg, 1983). This procedure allowed the subjects to focus their attention, during perceptual observation, on the relevant characteristics of the two training CBCs (Tsal & Kolbet, 1985). Indeed, the strength of local cues and the viewer's intentions have been shown to play an important role in form perception and to influence the durations of alternative organizations (Peterson & Hochberg, 1983). If a subject had difficulty in seeing the second interpretation, the experimenter progressively indicated the parts of the depicted object that could serve as cues, proceeding from the least relevant to the most relevant parts. In doing so, the experimenter encouraged the subject to discover the new interpretation, without providing the correct response. If the subject claimed not to be able to see it, the experimenter named the correct response and described the figure while finger tracking its contours.

In the inspection phase, before showing the test figure, the experimenter informed the subjects that the next part of the experiment was about visual memory. More precisely, the verbal instruction was: "Now, look very carefully at the next figure and try to remember it as well as you can." No information about the imagery task was provided during this phase of the experiment. A pilot study with 25 naive adults showed that 3 sec was sufficient for 6 subjects to find the second interpretation. Different results are reported by Girgus, Rock, and Egatz (1977) and Chambers and Reisberg (1985). In these studies, 5 sec is indicated as the minimum time necessary to reverse the configuration during perceptual observation. Using a 5-sec presentation time with adults, Hyman and Neisser (1991) were forced to eliminate 6 out of 26 subjects after the inspection phase. We decided to reduce the presentation time to avoid the elimination of a great number of subjects. The duck/rabbit figure was then presented for 2 sec, which was still sufficient for verbal recoding to occur (Bahrick & Boucher, 1968). Half of the subjects were required to remain silent during presentation, whereas half were asked to articulate the sound "la-la" at a rate of 3-4 "la"s per second, which is known to be the ad hoc rate sufficient to produce effects of suppression, but not to be demanding in terms of attentional cost (see Baddeley, 1986). The orientation of the figure was counterbalanced across subjects; half saw the figure with the bill/ears oriented to the left and half to the right. After the figure was taken away, the experimenter asked the subject to say what he/she had seen and was prepared to eliminate him/her if both interpretations were reported.

In the imagery phase, which started after picture naming, the subject was required to hold the image in his/her mind. The experimenter then stated that, like the previous training figures, the imaged figure also had two interpretations. Following Hyman and Neisser's procedure (1991), the subject was informed about the orientation and the conceptual category of the second interpretation. We chose to use full information in order to facilitate younger children's understanding of the task.

Instructions given to adults were: "Now, pay attention. The second interpretation of this figure is another animal, but you can see it only if you look at the image from the opposite side. Consider the back of the head of the animal you have seen as the front of

the head of a different animal." Then, the subjects were required to name the alternative interpretation as soon as they had discovered it. These instructions characterize the full information condition (category plus orientation), also used in Experiments 2-4, in which only adult subjects were used.

For the two groups of children, the verbal instructions were based on the description of a practical example. Namely, they were required to show, by using their hands, through which part the animal they had seen could be fed. Then, they were told to imagine feeding the animal through the opposite part. It was clarified that in this way the back of the head of the first animal appeared as the front of the head of a different animal. At this point, the children were required to look for a second animal and to name it when they had seen it.

Results and Discussion

No subject was eliminated from the imagery phase. The results are summarized in Table 1. Answers were considered correct when the subjects used the expected name ("duck" or "rabbit") or an obvious synonym. The answer "bird" was considered correct because some subjects named the first interpretation as a "bird," rather than as a "duck," during perceptual observation. No effect of figure orientation was found.

As predicted, older children's and adults' performance significantly improved in the AS condition. Both in the older children's group and in the adults' group, 13 subjects out of 20 reversed the imaged duck/rabbit after AS, whereas only 5 subjects out of 20 reversed the image in the no-AS condition. This difference (5/20 vs. 13/20) was significant [$\chi^2(1) = 4.94, p = .026$]. The difference between the youngest and any of the older groups in the AS condition was close to significance. No statistical difference was found between age levels in the no-AS condition.

In accordance with previous results in the literature (Bahrick & Boucher, 1968; Brandimonte et al., 1992a, 1992b; Carmichael et al., 1932; Nelson & Brooks, 1973), these findings support the view that verbal recoding disrupts subjects' use of the visual code, hence impairing mental image processing. By blocking the adults' and older children's tendency to verbally recode visual stimuli, AS enhanced use of the visual STM code and improved performance in the imagery task. Young children's performance supports this view in that AS should not have any effect when a visual code is used. Moreover, it is in accordance with previous results in the literature showing that young children tend to use visual STM to retain visually presented materials (Hitch et al., 1988).

Further support for the view that older subjects use the visual code in the AS condition, but not in the no-AS condition, comes from the finding that in the no-AS condi-

Table 1
Number of Subjects ($N=20$) who Reversed the Image as a Function of Age and Presence of AS (Experiment 1)

Age	AS	No AS
6 years	8	7
10 years	13	5
Adult	13	5

tion, older children and adults performed worse than the younger children, whereas in the AS condition they performed better (age differences in this condition were close to significance). If control processes are taken into account, this result is the most reasonable: ceteris paribus, adults should always be expected to perform any cognitive task better than young children.

It should be noted that, contrary to the findings of Hyman and Neisser (1991), only 5 adults out of 20 reversed the image in the no-AS condition, despite full information. However, this finding cannot be taken as indicating a failure to replicate Hyman and Neisser's results. Indeed, the difference between their results and ours might be due to the different exposure time. We used a 2-sec presentation rate; Hyman and Neisser used 5 sec. Having seen the figure for only 2 sec, our subjects might have had less stored information to work with (U. Neisser, personal communication, June 26, 1991). By enhancing use of the visual code, AS may have increased the amount of encoded visual information as a consequence of the elimination of concurrent verbal processing.

The purpose of Experiment 2 was to explore the effect of AS with short versus long exposure time. Full information was provided in order to make Hyman and Neisser's (1991) results comparable with ours.

EXPERIMENT 2

In this experiment, the subjects viewed the duck/rabbit figure either for 2 sec or for 5 sec while articulating or not. Then, they were informed about the category and the orientation of the second interpretation of the image. From Experiment 1, we know that AS is sufficient to improve image reversal. It is reasonable to suppose that inspection time has a similar positive effect, as suggested by U. Neisser (personal communication, June 26, 1991). However, the conjoint effect of AS and inspection time remains to be demonstrated. In view of the results of Experiment 1, and since AS had no effect on young children's performance but did have the same beneficial effect for older children and adults, only adults were tested in this experiment.

Method

Subjects. Eighty subjects took part in this experiment. They were selected from a pool of 102 volunteers randomly assigned to one of four conditions (2 sec/AS, 2 sec/no AS, 5 sec/AS, 5 sec/no AS). Seventeen subjects out of 57 in the 5-sec groups and 5 subjects out of 45 in the 2-sec groups were eliminated for reporting both interpretations immediately after viewing the picture. The subjects were high school students (mean age: 17.4 years).

Materials and Procedure. In the training phase, the rat/man and the young girl/old woman pictures were again used as familiarization figures. The procedure was the same as that used in Experiment 1, with the exception that the subjects assigned to the experimental conditions viewed the test figure for 5 sec. The order of presentation of the training figures was counterbalanced across subjects. Prior to viewing the duck/rabbit figure, the subjects were warned that they would be asked to perform an imagery task and were encouraged to remember the duck/rabbit figure exactly as it

Table 2
Number of Subjects (N=20) who Reversed the Image as a Function of AS and Time of Inspection (Experiment 2)

	2 Sec	5 Sec
AS	13	13
No AS	5	12

appeared. We chose not to use the demonstration figure by Carmichael et al. (1932), as previous results showed that its presence is irrelevant (Peterson et al., 1992).

Results and Discussion

Results are shown in Table 2. The 2-sec conditions replicated results from Experiment 1: AS increased the number of subjects who reversed their images from 5 to 13 (out of 20). This difference was significant [$x^2(1) = 4.94, p = .026$]. In the 5-sec conditions, the number of subjects who reversed their images was equally high both without AS (12/20) and with AS (13/20).

These results showed that AS is effective only when combined with a short inspection time, whereas inspection time is effective only in the absence of AS [$x^2(1) = 3.68, p = .055$]. The presence of AS alone or the lengthening of inspection time alone are sufficient to produce a proportion of reversals equivalent to that obtained when the two facilitatory treatments are combined.

In this experiment, we also replicated part of Hyman and Neisser's (1991) results, and showed that, when full information is given after a 5-sec presentation even without AS, a larger proportion of the subjects reversed their image. It seems that when subjects are given more time to load the image, they store more information to work with. On the other hand, time showed no beneficial effect when the subjects engaged in AS. One interpretation of this finding is that because AS prevents verbal recoding and forces reliance on the visual code, additional visual analysis may be ineffective. However, an alternative explanation for the lack of further improvement in the 5-sec/AS condition is that performance was at ceiling. On a more general ground, this might reflect a difference between imagery and perception. Indeed, fewer surface details are present in an image than in a picture (Kosslyn, 1980).

Experiment 2 replicated Experiment 1 in showing that after a 2-sec presentation, reversal of the duck/rabbit in mental imagery is facilitated when subjects engage in AS during initial observation and are given full information. However, one might ask whether a significant improvement in image reversal could be obtained without any of Hyman and Neisser's (1991) hints. To ascertain whether the AS effect generalizes to situations in which only a partial hint or no hint is provided, we conducted a third experiment.

EXPERIMENT 3

In Experiment 3, we manipulated instructions by combining presence versus absence of two kinds of informa-

tion: category and orientation. We obtained four conditions (category, orientation, category plus orientation, and no information) and tested them with and without AS. If our hypothesis is correct, and verbal recoding occurs spontaneously when initially looking at the stimulus, AS should have a beneficial effect regardless of the amount of information.

Method

Subjects. We examined a total of 178 students from the University of Trieste, who voluntarily participated in the experiment. Eighteen were eliminated, 15 because they reported both interpretations prior to the imagery phase and 3 because they were unable to reverse even during re-presentation of the figure, at the end of the task. Therefore, 160 subjects were randomly assigned to one of eight conditions resulting from the combination of 4 levels of instruction × 2 levels of AS. To avoid experimenter bias, each condition was run by a different experimenter. Eight first-year psychology students acted as experimenters, in the fulfillment of course requirements. They were completely blind as regards the hypothesis and were carefully trained so as to make experimenter effects unlikely.

Materials and Procedure. Materials and procedure were the same as in Experiments 1 and 2. After training, the subjects were shown the test figure for 2 sec. Different instructions were given according to the experimental conditions. The group assigned to the full-information condition (category plus orientation) received the same instructions as in previous experiments. For the subjects assigned to the no-information condition, instructions were: "This figure also has a second interpretation. Try to discover what this new interpretation could be." In the category condition, instructions were: "This figure also has a second interpretation. It is another animal. Try to discover which animal it is." For the orientation group, instructions were as follows: "This figure also has a second interpretation. Consider the back of the figure you have seen as the front of a different figure. Shift from right to left [or vice versa]." Those subjects who failed to mentally reverse their images were presented the original stimulus again and asked to look for the second interpretation while viewing the stimulus. If a subject was still unable to reverse, he/she was eliminated from the sample and replaced with a new volunteer.

Results and Discussion

Table 3 shows the duck/rabbit reversals obtained with and without AS in each of the four instruction conditions. AS significantly improved performance in the imagery task regardless of the amount of information provided to the subject. For each level of instruction, a chi-square analysis confirmed that the difference between performance with and without AS was significant [no information, $\chi^2(1) = 5.16, p = .023$; orientation, $\chi^2(1) = 3.9$, $p = .048$; category, $\chi^2(1) = 6.54, p = .01$; full information, $\chi^2(1) = 4.1, p = .042$].

Table 3
Number of Subjects (N=20) who Reversed the Image as a Function of Instructions and Presence of AS (Experiment 3)

Instructions	AS	No AS
No information	8	1
Orientation	7	1
Category	13	4
Category plus orientation	10	3

Without AS, only one subject reported a reversal in the no-information condition. This result is consistent with those reported by Chambers and Reisberg (1985), Hyman and Neisser (1991), and Peterson et al. (1992). The orientation condition replicated the results in the no-information condition in that only 1 subject was able to reverse the duck/rabbit, whereas 4 subjects reversed in the category condition. The full-information condition replicated results from Experiments 1 and 2 in showing that only a few subjects (3 out of 20) reversed their images. Differences among the four levels of instruction in the no-AS condition were not significant. Analogously, differences among the four levels of instruction in the AS condition were not significant.

However, inspection of Table 3 suggests an interesting difference between the two conditions without category information versus the conditions with category information. In contrast, the presence of orientation hints seems to be irrelevant. To test the relative contribution of category versus orientation information, two chi-square analyses were calculated by pooling the four conditions without category information against the four conditions with category information (regardless of AS and orientation). Seventeen subjects out of 80 reversed the image in absence of category information, whereas 30 subjects out of 80 reversed in presence of category information [$\chi^2(1) = 4.33, p = .037$]. On the other hand, no overall effect of orientation was found when results were pooled across category information.

This experiment demonstrated that CBC reversals can occur in imagery even when no information is given, provided subjects are prevented from verbally recoding the stimulus at the time of perceptual inspection. While the results from the conditions without AS support previous findings in the literature, given that few reversals occurred without explicit hints, the results from the AS condition argue against the general conclusion that the kind of information is the only critical factor in prompting CBC reversals. Nonetheless, our data also support the conclusion that instructions can facilitate image reversal. When performance was analyzed regardless of AS, category information was found to be more helpful than orientation information.

In examining the effect of AS, we demonstrated that it has a beneficial effect on image reversal. However, so far our results are not informative about whether some sort of verbal overshadowing may take place during the imagery phase as well. That is, even if subjects are prevented from verbally recoding the picture in STM, there might be a possibility of verbally recoding some aspects of the representation during image manipulation. In turn, this kind of later verbal recoding might have an overshadowing effect on the original visual memories. If verbal recoding occurs during image manipulation, one would expect further improvement when subjects engage in AS during both initial loading and image manipulation. A good performance should also be observed when AS is used in the imagery phase only, as compared with the con-

dition in which AS is not used. Alternatively, if no additional verbal recoding takes place during imagery, AS should be expected to help performance whenever it is present during loading, but not during image manipulation. Yet, AS during both phases should not further increase the proportion of reversals.

EXPERIMENT 4

The purpose of Experiment 4 was to explore whether verbal recoding occurs not only during initial loading, but also during image manipulation. Although studies on visual imagery (Kosslyn, 1980) have repeatedly shown that during image manipulation subjects tend to rely predominantly on the visual code, there remains the possibility that people partially redefine propositionally the image while elaborating and transforming it. If so, the subject may not have exploited all his/her resources in order to successfully perform the task. Therefore, in this experiment, the subjects were asked to articulate irrelevant sounds during loading, during manipulation, or during both phases. In a further control condition, the subjects were requested to remain silent. Besides AS, we manipulated instructions by providing either no information or full information to maximize their possible difference.

Method
Subjects. We examined a total of 174 high school students, who voluntarily participated in the experiment. Fourteen were eliminated because they reported both interpretations of the duck/rabbit figure immediately after viewing the stimulus. Therefore, 160 subjects were randomly assigned to one of eight conditions: AS during loading, AS during image manipulation, AS during both, and no AS, each with or without information.

Materials and Procedure. Materials and procedure were the same as in previous experiments, except that two new conditions were added in which subjects were asked to start articulating the sound "la-la" either immediately after the figure had been shown and instructions had been given or during both presentation of the figure and manipulation of the image. Again, to avoid experimenter bias, two experimenters tested the subjects. However, in this experiment, each experimenter ran half of each condition.

Results and Discussion
Table 4 shows the number of subjects who reversed the image in each of the eight conditions. Given that the same pattern of results was observed in both the full-information and no-information conditions, the data were collapsed across information. A chi-square analysis showed a highly significant difference between the AS-during-loading and

Table 4
Number of Subjects ($N=20$), as a Function of Presence of AS at Different Stages, who Reversed the Image (Experiment 4)

	No Information	Full Information
AS during loading	9	12
AS during image manipulation	2	3
AS during both stages	9	11
No AS	1	4

the AS-during-manipulation conditions [$\chi^2(1) = 12.82$, $p = .0003$]. No difference was found between the AS-during-loading and the AS-during-both conditions, whereas the difference between the AS-during-both and the AS-during-manipulation conditions was highly significant [$\chi^2(1) = 11.40, p = .0007$]. Finally, performance in the AS-during-manipulation condition was as low as that in the no-AS condition.

As is clear, AS had a beneficial effect only if it was used during initial loading. The rate of reversal in the AS-during-both condition replicated that obtained in the AS-during-loading condition. Adding AS during manipulation of the image did not help image reversal. Yet, this condition equated to that without AS. Furthermore, there was no overall difference between the full- and no-information conditions.

Taken together, these results confirm that it is critical that the visual STM code be active at the time of loading the picture into the visual buffer. We can, therefore, conclude that when AS is used, later verbal recoding that would overshadow visual information does not take place.

GENERAL DISCUSSION

This research has provided new findings on the reversal of bistable configurations in imagery and critical constraints for a model of the interaction between imagery and memory codes.

The major finding is the demonstration that CBC reversals can occur in mental imagery, even when no information is provided, contrary to the prevailing trend in the literature. We found that image reversal occurs even when subjects are not prompted to change their reference frame or specification of orientation. Under the particular conditions examined in this research, the likelihood of reversal seems to depend on the kind of memory code used during the initial loading of the visual input into the visual buffer. Whenever people are forced to use the visual code, the proportion of reversals increases.

Another important finding regards the experimental paradigm used to explore image reversals. A basic methodological requirement of imagery studies using CBCs is that the reversal should not occur in perception (during initial encoding), but only in imagery. Accordingly, precise instructions about the task (i.e., asking people to look for the alternative interpretation) have to be given only after picture presentation (Chambers & Reisberg, 1985; Hyman & Neisser, 1991; Peterson et al., 1992; Reisberg & Chambers, 1991). We suggest that such a requirement increases the likelihood that the picture will be spontaneously recoded into a verbal form. Indeed, our results support the hypothesis that people's ability to reverse mental images is affected by the way in which the image is initially encoded.

Once they are compared with other results in the literature, our findings give rise to a number of questions concerning the underlying processes that take place when visual and verbal codes interact. For example, one might

ask how verbal codes interfere with the original visual memories or in what way verbal recoding differs from verbal overshadowing. Although further research is necessary to answer the above questions, it seems appropriate to speculate on how verbal recoding and verbal overshadowing might explain the results so far obtained in imagery studies that employed the reversal paradigm and to propose a tentative model of image reversal.

The main assumption of the model holds that verbal recoding occurs spontaneously, almost immediately at the time of input (see also Bahrick & Boucher, 1968). Another assumption is that the duration of the inspection phase influences the quality of the visual representation.

If subjects are shown the picture for a short time (e.g., 2 sec), part of the visual information is lost as a consequence of verbal recoding. Another consequence is that, after verbal recoding, the remaining part of the visual information is overshadowed by the verbal code that is now dominant.

If subjects are shown the picture for a longer time (e.g., 5 sec), they learn more visual details by virtue of a double shifting in the use of STM codes (visual-verbal-visual). That is, a 5-sec presentation may prompt a process in which the visual input automatically loaded into the visual buffer is readily (e.g., within the first 2 sec) recoded in a verbal form. However, because the stimulus is easily recognizable and nameable, one could use the time left available after recoding to continue its visual analysis. As stated earlier, when both codes are present, the verbal information tends to overshadow the visual information. In the case of a longer inspection time, though, the visual representation overshadowed by the verbal code is more detailed.

Our tentative model also takes into account the possible removal of verbal overshadowing by manipulating instructions. Providing appropriate suggestions about the alternative interpretation may facilitate removal of verbal overshadowing by inducing the most effective strategy to use for exploiting the encoded visual information. Therefore, performance in image-reversal tasks should reflect both the amount and the accessibility of the visual information originally encoded during loading.

However, it should be noted that more or less effective strategies may be prompted by different kinds of suggestions. Implicit suggestions (such as the use of demonstration figures whose reversals entail a change in the reference frame that is similar to the change necessary to reverse the test figure) were found to be more helpful in prompting reversals than were explicit hints (Peterson et al., 1992). Indeed, explicit hints such as information about either the orientation or the category of the second interpretation of the image showed no beneficial effect on reversals, unless they were combined (Hyman & Neisser, 1991; Peterson et al., 1992). One way of explaining these findings is that implicit hints intervene prior to picture presentation, hence affecting the way in which the picture is initially encoded in STM. That is, implicit hints may act similarly to AS: they simply prevent people from

verbally recoding the picture in STM. On the other hand, explicit hints are active after encoding. Hence, their effect may be due to the removal of verbal overshadowing.

In summary, we propose a model based on the following assumptions: interaction between visual and verbal codes, automatic verbal recoding of the visual input in STM, possible prevention of verbal recoding by AS, influence of presentation time on the quality of the visual representation, existence of verbal overshadowing, and possible removal of verbal overshadowing by post-presentation information. If we consider conditions in which presentation time, presence of AS, and information are manipulated, the model generates a set of predictions that fit most data so far reported in the literature on image reversal.

1. *2-sec presentation/no AS/no information:* Low performance because of verbal recoding in STM (confirmed by our data).

2. *2-sec presentation/no AS/information:* Medium performance because of the removal of verbal overshadowing, which reveals a degraded visual representation due to the short inspection time and verbal recoding (confirmed by our data).

3. *2-sec presentation/AS/no information:* High performance because of inhibition of verbal recoding in STM (confirmed by our data).

4. *2-sec presentation/AS/information:* High performance because of inhibition of verbal recoding in STM; no effect of information, hence no difference from Prediction 3 (confirmed by our data).

5. *5-sec presentation/no AS/no information:* Low performance because of verbal overshadowing (see data from Chambers & Reisberg, 1985; Hyman & Neisser, 1991; Peterson et al., 1992).

6. *5-sec presentation/no AS/information:* High performance because of the removal of verbal overshadowing, which reveals a detailed visual representation (see data from Hyman & Neisser, 1991; Peterson et al., 1992; and our data).

7. *5-sec presentation/AS/no information:* High performance because of inhibition of verbal recoding in STM (no existing data).

8. *5-sec presentation/AS/information:* High performance because of inhibition of verbal recoding in STM; no difference from Predictions 6 and 7 (our data).

This model, although partially speculative, explains a large body of results. It accounts not only for the lack of reversal reported by Chambers and Reisberg (1985), Hyman and Neisser (1991), and Peterson et al. (1992), but also for the positive results reported by Hyman and Neisser (1991), Peterson et al. (1992), and, most recently, by Reisberg and Chambers (1991) and D. Chambers and D. Reisberg (1992).

In general, our results agree with the hypothesis that image reversal is constrained by an interpretive bias that limits its occurrence. This interpretive bias can be pre-

vented by using manipulations that are effective before inspection (implicit hints) or during inspection (AS); alternatively, it can be removed by providing instructions that act on people's understanding of the specification of orientation of their images (explicit hints). The first case would reflect prevention of STM verbal recoding, whereas the second case would reflect removal of the verbal overshadowing effect.

The present research clearly indicates that the AS technique acts at an earlier stage of the process. This is shown by the fact that the presence of AS is sufficient for image reversal to improve, whereas providing explicit hints has little effect when a 2-sec presentation is used and AS is not present.

When considered in the context of other recent findings on mental-image transformations, our experiments raise a general question about the role of verbal recoding in imagery tasks. There is an apparent discrepancy between the present results and Brandimonte et al.'s (1992a) outcome showing that AS has no effect when the task is performed in STM. In that study, AS had no effect when the imagery task was performed on an image of a just-presented stimulus, rather than on an image of a learned representation in LTM. This result has been interpreted as indicating that performance was spontaneously based on visual STM (Brandimonte et al., 1992a).[1]

However, this conclusion should not be generalized to every imagery task involving coding and manipulation of a single image in STM, for a number of reasons. First, when the subjects were required to subtract a part of an image, they were forewarned of the exact type of imagery task. Second, the subjects were first shown a composite picture and, immediately after, the part to take away. It is likely that this procedure induced use of the visual code in that the first presentation was immediately followed by the other visual input specifying which part of the previous picture to take away. Third, the part to take away was not easily nameable. This fact may have contributed to induce visual rather than verbal coding. It is, then, conceivable that under different conditions, a verbal form of encoding takes place even in short-term imagery tasks. This view is supported by recent findings showing that use of the visual code is hard to control voluntarily and that it is susceptible to variations in the external conditions (Brandimonte et al., 1992a). For example, it has been shown that when a series of easily nameable stimuli has to be learned in LTM, knowing the nature of the task to be performed is not sufficient to induce use of a visual form of encoding. Indeed, subjects tend to spontaneously recode visual stimuli into a verbal form and rehearse their names during initial learning (Brandimonte et al., 1992a), given that in terms of attentional resources the speech-based code is less consuming (Baddeley, 1986).

The present investigation indicates that verbal overshadowing can be responsible for difficulties in image reversal. This is not to say that the only reason for a poor performance in image-reversal tasks is the competition

between visual and verbal codes. It is likely that the duck/rabbit configuration is unique, as it involves a simple graphical pattern and two easily nameable interpretations. Similar results might not be obtained with other bistable configurations that imply maintenance of a larger amount of visual information (e.g., the young girl/old woman) or the completion of missing details in degraded patterns (e.g., the rat/man). Therefore, limitations because of the complexity of the visual representation might explain the possible unambiguity of such images.

However, the uniqueness of the stimulus does not weaken the basic conclusion that, we think, has to be drawn from this line of research: Whenever people are forced to exploit their visual resources, mental images appear to be inherently ambiguous, and ducks become rabbits.

REFERENCES

BADDELEY, A. D. (1986). *Working memory*. Oxford: Clarendon.

BADDELEY, A. D., & HITCH, J. G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.

BAHRICK, H. P., & BOUCHER, P. (1968). Retention of visual and verbal codes of the same stimuli. *Journal of Experimental Psychology*, 78, 417-422.

BORING, E. G. (1930). A new ambiguous figure. *American Journal of Psychology*, 42, 444-445.

BRANDIMONTE, M. A., HITCH, G. J., & BISHOP, D. V. M. (1992a). Influence of short-term memory codes on visual image processing: Evidence from image transformation tasks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18, 157-165.

BRANDIMONTE, M. A., HITCH, G. J., & BISHOP, D. V. M. (1992b). Verbal recoding of visual stimuli impairs mental image transformations. *Memory & Cognition*, 20, 449-455.

BRANDIMONTE, M. A., HITCH, G. J., & BISHOP, D. V. M. (1992c). Manipulation of visual mental images in children and adults. *Journal of Experimental Child Psychology*, 53, 300-312.

BUGELSKI, B. R., & ALAMPAY, D. A. (1961). The role of frequency in developing perceptual set. *Canadian Journal of Psychology*, 15, 205-211.

CARMICHAEL, L., HOGAN, H. P., & WALTER, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived forms. *Journal of Experimental Psychology*, 15, 73-86.

CHAMBERS, D., & REISBERG, D. (1985). Can mental images be ambiguous? *Journal of Experimental Psychology: Human Perception & Performance*, 3, 317-328.

CHAMBERS, D., & REISBERG, D. (1992). What an image depicts depends on what an image means. *Cognitive Psychology*, 24, 165-176.

FINKE, R. A. (1989). *Principles of mental imagery*. Cambridge: MIT Press.

FINKE, R. A. (1990). *Creative imagery*. Hillsdale, NJ: Erlbaum.

FINKE, R. A., PINKER, S., & FARAH, M. (1989). Reinterpreting visual patterns in mental imagery. *Cognitive Science*, 13, 51-78.

GIRGUS, J. J., ROCK, I., & EGATZ, R. (1977). The effect of knowledge of reversibility on the reversibility of ambiguous figures. *Perception & Psychophysics*, 22, 550-556.

HITCH, G. J., HALLIDAY, S., SCHAFSTAAL, A. M., & SCHRAAGEN, J. M. C. (1988). Visual working memory in young children. *Memory & Cognition*, 16, 120-132.

HITCH, G. J., WOODIN, M. E., & BAKER, S. (1989). Visual and phonological components of working memory in children. *Memory & Cognition*, 17, 175-185.

HYMAN, I. E., & NEISSER, U. (1991). *Reconstruing mental images: Problems of method* (Emory Cognition Project Rep. No. 19). Atlanta: Emory University.

JASTROW, J. (1900). *Fact and fable in psychology*. Boston: Houghton Mifflin.

KOSSLYN, S. M. (1980). *Image and mind*. Cambridge: Harvard University Press.

KOSSLYN, S. M., BRUNN, J., CAVE, K. R., & WALLACH, R. W. (1984). Individual differences in mental imagery ability: A computational analysis. *Cognition*, **18**, 195-243.

KOSSLYN, S. M., HOLTZMAN, J. D., FARAH, M. J., & GAZZANIGA, M. S. (1985). A computational analysis of mental image generation: Evidence from functional dissociations in split-brain patients. *Journal of Experimental Psychology: General*, **114**, 311-341.

MURRAY, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, **78**, 679-684.

NELSON, D. L., & BROOKS, D. H. (1973). Functional independence of pictures and their verbal memory codes. *Journal of Experimental Psychology*, **98**, 44-46.

PAIVIO, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.

PAIVIO, A. (1975). Imagery and long-term memory. In A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory* (pp. 57-85). New York: Wiley.

PETERSON, M. A., & HOCHBERG, J. (1983). Opposed-set measurement procedure: A quantitative analysis of the role of local cues and intention in form perception. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 183-193.

PETERSON, M. A., KIHLSTROM, J. F., ROSE, P. M., & GLISKY, M. L. (1992). Mental images can be ambiguous: Reconstrual and reference-frame reversals. *Memory & Cognition*, **20**, 107-123.

POSNER, M. I. (1969). Abstraction and the process of recognition. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 3, pp. 63-100). New York: Academic Press.

POSNER, M. I., BOIES, S. J., EICHELMAN, W. H., & TAYLOR, R. L. (1969). Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monographs*, **79**, 1-16.

REISBERG, D., & CHAMBERS, D. (1991). Neither pictures nor propositions: What can we learn from a mental image? *Canadian Journal of Psychology*, **45**, 336-352.

REISBERG, D., SMITH, D. J., BAXTER, D. A., & SONENSHINE, M. (1989). "Enacted" auditory images are ambiguous; "pure" auditory images are not. *Quarterly Journal of Experimental Psychology*, **3**, 619-641.

ROCK, I. (1983). *The logic of perception*. Cambridge: MIT Press.

SCHOOLER, J. W., & ENGSTLER-SCHOOLER, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, **22**, 36-71.

SLOWIACZEK, M. L., & CLIFTON, C. (1980). Subvocalization and reading for meaning. *Journal of Verbal Learning & Verbal Behavior*, **19**, 573-582.

TSAL, Y., & KOLBET, L. (1985). Disambiguating ambiguous figures by selective attention. *Quarterly Journal of Experimental Psychology*, **37A**, 25-37.

NOTE

1. This result and similar findings with difficult-to-name pictures also make it unlikely that increased effort or attention are promoted by AS because of increased motivation.