Selective interference between imagery and perception: Is it modality specific or relation specific?

DENNIS SCHORR, GERALD BALZANO, and EDWARD E. SMITH
Stanford University, Stanford, California 94305

When subjects perform a perceptual task while maintaining an image, there is more interference when the task and image are in the same modality than when they are in different ones (e.g., Segal & Fusella, 1970). This has been interpreted as demonstrating that the image is a perceptual representation. An alternative interpretation is that both image and task are represented propositionally with the relations expressed in the propositions being more similar when the image and task involve the same modality. To evaluate the new interpretation, subjects formed images that varied in modality—visual vs. auditory—and in the relation describing the imaged information—spatial vs. temporal. While maintaining an image, subjects performed either a spatial or a temporal visual task. Performance was poorer when the task and image were in the same modality, but was relatively unaffected by whether the task and image relations were the same or not.

There is currently a debate about the role of propositions and images in representing information. In a recent review of this debate, Kosslyn and Pomerantz (1977) point out that some of the best support for the imagery position comes from studies of selective interference. In a representative study of this ilk, Segal and Fusella (1970) demonstrated that performance on a visual detection task was interfered with more by maintaining a visual rather than an auditory image, while performance on an auditory detection task was disrupted more by maintaining an auditory rather than a visual image. Such modality-specific interference has been interpreted in the following way: (1) Detecting the task signal involves limited-capacity perceptual processes; (2) some of these processes are also needed to maintain an image; and (3) since an image is a perceptual representation, image maintenance usurps more of the processes needed to detect the signal when the image and signal are in the same modality.

As Kosslyn and Pomerantz (1977) note, it is possible to come up with a propositional interpretation of the findings in question. Specifically, (1) detecting a signal, of course, involves perceptual processes, but the outcome of these processes can be represented propositionally; (2) the so-called "image" is also represented propositionally; and (3) the propositions used to represent the perceptual task and image contain more similar predicates or relations when both task signal and image are from the same modality, since visually

presented information tends to express spatial relations, while auditorily presented information tends to express temporal relations. This interpretation suggests that modality-specific interference is really relation-specific interference in disguise.

We have, then, two interpretations of modality-specific interference, the perceptual one based on modalities per se, and the new propositional one based on relations. To decide between them, we used an interference paradigm modeled after Segal and Fusella (1970) and varied the type of relation, spatial or temporal, present in the task and image, as well as the modality of the image (see Table 1). Subjects formed four types of images: (1) visual images with a dominant above-below (spatial) relation, (2) visual images with a dominant before-after (temporal) relation, (3) auditory images with a dominant above-below (spatial) relation, and (4) auditory images with a dominant before-after (temporal) relation. While maintaining images of one of these types, subjects performed a visual task in which they tried to detect either an above-below (spatial) relation or a before-after (temporal) relation. The perceptual interpretation implies that performance in the perceptual task should be poorer when the task and image are from the same modality; it may also be consistent with poorer performance when the task and image involve the same relation if perceptual processes are relation specific. The propositional interpretation implies poorer performance when the task and image involve the same kind of relation; it may also be consistent with poorer performance for same-modality items because subjects may encode propositions about the modality itself. So only two possible patterns of results have clear-cut implications: There is modality-specific but not relation-specific interference (supporting the perceptual interpretation),

Table 1
The Eight Conditions of the Experimental Design
and Corresponding Difference Scores

| | | Task | |
		Visual Spatial	Visual Temporal
Visual Image	Spatial	2	65
	Temporal	97	165
Auditory Image	Spatial	1	22
	Temporal	27	5

Note—Difference scores = mean test-trial minus mean baseline RT in milliseconds.

or there is relation-specific but not modality-specific interference (supporting the propositional view).

In a pilot study using the design in Table 1, image modality was manipulated between subjects while image relation and task relation were varied within subjects. There was some evidence for modality-specific interference but not for relation-specific interference, as performance in the perceptual task was actually slightly better when the task and image expressed the same rather than different relations. The following study differed from the pilot in that all factors were varied within subjects, more than double the number of observations was obtained in each condition, and the images were simplified.

METHOD

Tasks and Images

Tasks. Both tasks were visual, but the spatial task required detection of an above-below relation, while the temporal task required detection of a before-after relation. In both tasks, two dots were presented tachistoscopically on every trial, where each dot was 7/16 in. in diameter and the horizontal distance between the dots was 3-3/4 in.

In the spatial task, the two dots were presented simultaneously, and subjects made an "above-below" judgment. They pressed a button (with the left forefinger) if one dot was above the other and made no response if the two dots were equal in height. The two types of trials occurred equally often, and on trials where one dot was above the other, the vertical separation between them was 1/8 in. Exposure duration was determined separately for each subject so that he or she responded with approximately 85% accuracy. Durations ranged from 10 to 45 msec across the eight subjects.

In the temporal task, the dots were presented successively on two-thirds of the trials. Subjects made a "before-after" judgment. They pressed a button (with the left forefinger) if one dot occurred before the other and made no response if the dots occurred simultaneously. The two trial types occurred equally often. On those trials where one dot occurred before the other, the temporal separation between them was constant for each subject, this constant ranging from 30 to 55 msec across subjects to insure responding at approximately 85% accuracy.

Images. Both visual and auditory images could be either spatial or temporal. Visual images involved two colors drawn from the set of red, blue, and green (subjects were prefamiliarized with these three colors). To induce a spatial image, the experimenter said two color names (e.g., red and blue); subjects had been previously instructed to image the first-named color

at the "top" of their "image field" and the second-named color at the "bottom," with a blank region between them. (The imagery instructions used the words "top" and "bottom," rather than "above" and "below," which were used in the task instructions, to prevent interference between image and perceptual task at the level of specific words.) Subjects were told the shape of each color was irrelevant, but both colors had to be imaged simultaneously for a period of a few seconds. To induce a temporal image, the experimenter named two colors; the subject had been instructed to image the colors successively at the same spatial location, with the first-named color occurring "first" and the second-named color occurring "next" (with instructions using the words in quotation marks). Again, the shape of the colors was irrelevant, and subjects were instructed to recycle the images of the successive colors, roughly at the rate of 1-2 sec/cycle.

The auditory images involved two sounds: single notes from a flute, violin, or baritone (subjects were prefamiliarized with these sounds). To induce a spatial image, the experimenter named two sounds (e.g., flute and violin), and subjects had been previously instructed to image the first-named sound as emanating "from the ceiling" and the second-named sound as coming "from the floor." Subjects were instructed to image both sounds simultaneously for a period of a few seconds. To induce a temporal image, the experimenter named two sounds; subjects had been instructed to image the sounds as successively emanating from the same location, with the first-named sound occurring "first" and the second-named sound occurring "next." Subjects were further instructed to recycle their images of the successive sounds, roughly at the rate of 1.5-3 sec/cycle.

Only one of the four kinds of images was required in any block of trials, and subjects generated each image while looking at an illuminated, blank, tachistoscopic screen.

Procedure

The experiment included three sessions. In the first, subjects practiced generating the four types of images. In the second and third sessions, subjects performed one of the visual tasks while maintaining the various kinds of images. Half the subjects had the spatial task in the second session and the temporal task in the third session, while the remaining subjects had the reverse assignment.

Session 1. To teach the subjects visual-spatial images, two color patches (drawn from red, blue, and green) were presented simultaneously in the tachistoscope, with one color at the top and the other at the bottom. To teach visual-temporal images, two colors were presented in rapid succession, the pair being recycled every 1-2 sec. For auditory-spatial images, a pair of sounds (produced by a flute, violin, or baritone) was presented simultaneously on a stereophonic tape recorder. The sounds were produced so that one would be localized as coming from the ceiling, the other from the floor. To teach auditory-temporal images, a pair of sounds, localized as coming from the same location, was presented in rapid succession; the pair was recycled every 2 sec. After being presented a physical manifestation of a particular kind of image, subjects tried to generate that image themselves. They rated each of their images on a 5-point scale, with 1 corresponding to "no image" and 5 to "as vivid as if the colors or sounds were actually present." If a subject's average rating for any of the four kinds of images was less than 3, that subject was dropped from the experiment. This procedure resulted in the loss of half of the subjects, but gives some confidence that the remaining subjects were able to perform the imagery part of the task.

Sessions 2 and 3. Eight Stanford undergraduates participated in the two remaining 2-h sessions. Each session was limited to one task (spatial or temporal) but included all four imagery conditions. Subjects were first familiarized with the perceptual

task and given at least 15 practice trials without any images. These trials were used to determine the proper exposure duration (spatial task) or interstimulus interval (temporal task) for each subject. The sequence of events on each trial was as follows. After a verbal "ready" signal from the experimenter, the subject started the trial by pressing a start key on a two-key panel. Then, 500 msec later, a pair of dots appeared and a millisecond timer was activated. Subjects indicated whether one dot was above the other (spatial task) or whether one dot occurred first (temporal task) by pressing a response key as rapidly as possible. Depression of a key stopped the timer, and the experimenter recorded the reaction time (RT) in milliseconds if the response was correct. (If the dots in the spatial task appeared on the same vertical level, or the dots in the temporal task appeared at the same time, no response was to be made.)

Then came four blocks of experimental trials, each block requiring one of the four types of images. Every block began with 15 nonimagery trials. Performance on these trials was used to establish a baseline, against which the interfering effects of images could be measured. Next, subjects were given six refresher trials in generating images. Finally, subjects were required to maintain an image while performing the perceptual task. They were asked to form each image vividly before pressing the start key and to maintain the image (in the case of spatial images) or recycle it (in the case of temporal images) throughout the time they performed the task. After the trial, subjects rated the vividness of their images on a 1-5 scale (as in Session 1) and reported on the stability of their images. There were 24 such image-plus-task or test trials in a block, with all possible images required equally often. After completing these trials, the subject began the next block. The order of the four blocks in a session was counterbalanced across subjects.

RESULTS

For each condition, the subject's mean baseline RT was subtracted from his or her mean test-trial RT. The means of these difference scores appear in Table 1, and they are used in the following analyses.

There was substantial modality-specific interference. For all four possible comparisons (the top four cells in Table 1 vs. the bottom four cells), there was more interference when a visual rather than an auditory image was combined with the visual task. The average interference effect was 82 msec for same-modality trials and 14 msec for different-modality trials, with seven of the eight subjects showing more interference on same-modality trials. This modality-specific effect was significant at the .01 level with a one-tailed Wilcoxon matched-pairs signed-ranks test (T = 2, N = 8). There was no evidence for relation-specific interference (T = 14, N = 8, p > .05, in the Wilcoxon test). The average interference effect was 43 msec for same-relation trials and 53 msec for different-relation trials, with five of eight subjects showing more interference on different-relation trials.

An analysis of variance for the complete 2 by 2 by 2 design (task relation by image relation by image modality) was performed on z scores. For every subject, a z score was computed for each of the eight conditions by subtracting the subject's overall mean RT from his or her mean condition RT and dividing the difference by the subject's pooled standard deviation. (When the data are transformed this way, all conditions have equal variance.) The main effect of image modality, that is, modality-specific interference, was significant [F(1,7) = 7.08, p < .05]. The Task Relation by Image Relation interaction was in the wrong direction and unreliable [F(1,7) = 1.65, p > .05]. No other effects approached significance.

No significant effects were obtained with either error percentages or vividness ratings, although errors were less frequent on baseline than on test trials (12% vs. 16%). The average vividness rating over all conditions was 3.7 (where 5.0 corresponded to "as vivid as if the colors or sounds were actually present"), suggesting that subjects generated vivid images. An analysis of subjects' reports on image stability further suggested that they tended to maintain their images during the perceptual task. On 73% of all test trials, subjects reported maintaining the image without disruption; on 16% of the trials, subjects reported the image was disrupted when the dots appeared but was quickly "recovered"; on the remaining 11% of the trials, subjects reported the image was disrupted when the dots appeared and was not recovered. These disruption percentages did not show any significant variation with the conditions of interest.

DISCUSSION

The results provided evidence for modality-specific interference, but not for relation-specific interference. The null result cannot readily be attributed to an insensitivity in the design. For one thing, same-relation trials actually produced less interference than different-relation trials (the opposite of the prediction), and this reversal also occurred in the pilot study. For another, in the z-score analysis, the error mean square for the relation effect (.33) was less than that for the modality effect (.35), so the null effect of relation type does not reflect greater variability. These results, then, support the perceptual interpretation of earlier selective-interference studies over the propositional one.

The results also extend Segal and Fusella's (1970) findings by demonstrating modality-specific interference on RT performance (Segal and Fusella used an accuracy measure). Although we used only half of the Segal and Fusella design (they used an auditory as well as a visual task), our extension of their work is of some import, given the inconsistent findings in studies of modality-specific interference. Thus Byrne (1974) and Paivio (1971, Footnote, p. 374) both reported difficulties in trying to replicate Atwood's (1971) interference effect. And Elliott (1973), along with Kosslyn, Holyoak, and Huffman (1976), found no evidence of modality-specific interference when information learned under imagery instructions was followed by a visual task. One reason why the Segal-Fusella paradigm may lead to more consistent modality-specific interference is that it requires an image to be actively maintained at the same time that visual processing occurs; in the paradigms of the other studies mentioned in this paragraph, imaginal information is followed by a visual task. Another difference between the Segal-Fusella paradigm and the less successful paradigms is that the former uses impoverished perceptual displays, while the latter use highly discriminable displays. Perhaps modality-specific interference can only be obtained when subjects' perceptual resources are truly taxed, as when dealing with impoverished displays.

Finally, two problems with the present study deserve mention. First, the present experiment, like that of Segal and

Fusella (1970), employed subjects that were prescreened to be good imagers, thereby casting doubt on the generality of the results. Second, although most of our subjects did not show relation-specific interference, there was a marginal correlation across subjects between the magnitude of their relation-specific interference effect and the magnitude of their modality-specific effect ($p = .57$, $df = 6$, $.05 < p < .10$). Subjects who showed the most modality-specific interference tended to show the most relation-specific interference (such as it was), while subjects who showed the least modality-specific interference tended to show the greatest reversals in relation-specific interference. While we are unsure how seriously to take this marginal correlation, even at face value it may be compatible with the perceptual position. There is some evidence that better imagers show more modality-specific interference than poorer imagers (Segal & Fusella, 1970), and better imagers may encode more information about relations into their images, thereby making same-relation trials more interfering than different-relation ones.

REFERENCES

ATWOOD, G. An experimental study of visual imagination and memory. *Cognitive Psychology*, 1971, **2**, 290-299.

BYRNE, B. Item concreteness versus spatial organization as predictors of visual imagery. *Memory & Cognition*, 1974, **2**, 53-59.

ELLIOTT, L. Imagery versus repetition encoding in short- and long-term memory. *Journal of Experimental Psychology*, 1973, **100**, 270-276.

KOSSLYN, S. M., HOLYOAK, K. J., & HUFFMAN, C. S. A processing approach to the dual coding hypothesis. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, **2**, 223-233.

KOSSLYN, S. M., & POMERANTZ, J. R. Imagery, propositions, and the form of internal representations. *Cognitive Psychology*, 1977, **9**, 52-76.

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

SEGAL, S. J., & FUSELLA, V. Influence of imaged pictures and sounds on detection of visual and auditory signals. *Journal of Experimental Psychology*, 1970, **83**, 458-464.