Image-dependent interaction of imagery and vision

THOMAS P. REBOTIER and DAVID J. KIRSH
University of California San Diego

LARAINE MCDONOUGH
Brooklyn College and the City University of New York Graduate Center

The influence of imagery on perception depends on the content of the mental image. Sixty-three students responded to the location of the 2 hands of a clock while visualizing the correct or an incorrect clock. Reaction time was shorter with valid cueing. Could this have resulted from visual acquisition strategies such as planning visual saccades or shifting covert attention? No. In this study, a crucial control condition made participants look at rather than visualize the cue. Acquisition strategies should have affected equally both types of cueing, but we observed that the effect of the visual cue was smaller and limited to a particular subcase in which one expects visual acquisition strategies. Thus, what matters is the similarity of the content of the mental image with the visual scene. In addition, an interaction involving the hand used for responding supports the notion that composite imagery is lateralized.

Twenty years after the imagery debates between Steven Kosslyn and Zenon Pylyshyn, mental imagery is considered a byproduct of the top-down mechanisms of vision. As a consequence of those top-down mechanisms, mental imagery should be able to bias visual perception. For example, suppose that you are visiting the Louvre and get a glimpse of a painting through a closing door. Expecting to see the *Mona Lisa* in that particular location, you would be visualizing Da Vinci's masterpiece. If the *Mona Lisa* is there, that ongoing visualization should help you see it. On the other hand, had the museum swapped the *Mona Lisa* with an El Greco landscape, you would not identify this landscape because your visual pathways would be biased in favor of the *Mona Lisa*. So, in the top-down effect theory, ongoing imagery should affect vision by introducing a bias in the common representational medium.

Yet it remains unclear whether pictorial visual representations are affected by ongoing imagery. In present studies showing that imagery biases vision, the bias could also be explained by changes in perceptual strategy. Such strategies encompass changes in eye position (as pointed out by Farah, 1988), covert attention, or saccade planning. For ex-

ample, eye position and movements optimal for perceiving the *Mona Lisa* will be suboptimal for perceiving the El Greco landscape. Such strategy changes arise from the knowledge of the picture; they are a degraded form of hypothesis testing and would operate equally well if imagery were propositional. Therefore, as long as the effects observed can be explained by strategy changes, they do not strengthen the theory that imagery results from top-down effects in the visual pathways.

Previous direct studies, in which ongoing imagery is superimposed onto a visual task, can be classified in three types: studies that focus on demonstrating a task-related, content-independent interference of imagery with perception; psychophysical studies trying to reproduce with an imaged context the same perceptual alterations of a central stimulus that are observed when the same context is really seen; and studies such as ours, looking in particular at the content-dependent interaction between the mental image and visual perception.

The inhibition of vision by imagery is called the Perky effect. Initially thought to be a confusion between mental and visual images (Perky, 1910), this effect was later interpreted as a low-level mechanism whereby ongoing imagery decreases perceptual sensitivity. Initial studies (Segal & Gordon, 1969; Segal & Fusella, 1969) made use of mental images similar to the stimuli but contained possible confounds, such as uneven mental load. In most subsequent studies (Segal & Fusella, 1970; Segal, 1971; Reeves & Segal, 1973) the mental images were systematically different from the stimuli, so that the inhibition observed could be attributed to either content-specific interaction or systematic interaction. When the mental images accidentally resembled the stimuli, as in one condition of Reeves (1980), a "reverse Perky effect" was found, showing a facilitation of vision by imagery; Reeves's interpretation was different but not incompatible with content-dependent facilitation. Similar pairs of mental images and stimuli can also be found in Reeves (1981), who reported that the similarity between mental image and stimulus had no significant effect, but attentional load effects (as suggested by Farah, 1983) could be masking more subtle results. Lately, several studies (Craver-Lemley & Reeves, 1987; Craver-Lemley & Reeves, 1992; Craver-Lemley, Arterberry, & Reeves, 1997) have used Vernier acuity to measure how imagery can interfere with vision. In this paradigm, the Perky effect is a diminishing of acuity when visualizing. For nonmasked stimuli, these authors estimated the Perky effect as equivalent to a variation of about 0.24 log units of stimulus energy.[1] As found by Piéron (1920) and more recently confirmed by Mansfield (1970), reaction times (RTs) decrease as the 0.3 power of stimulus energy; thus, for bright (50 to 100 cd/m^2), nonmasked stimuli (such as that used in our experiment), we would not expect the Perky effect to induce a change of

more than 7 ms. In summary, Perky effect studies have convincingly established that when imagery is concomitant with vision, reality decision (Segal, 1971) becomes an issue. It is also likely that a mild inhibition results from interference between the processes of vision and imagery. However, this set of experiments is not conclusive about how the imagery–vision interaction depends on the similarity between mental and visual images.

Several experiments show that alterations of perception induced by visual context can be replicated using imaged context. Successful replication of context-induced illusions has been obtained for the illusions shown in Figure 1. Berbaum and Chung (1981) reproduced the famous Müller–Lyer illusion. Wallace (1984) asked high and low imagers to image different elements of the Ponzo, Wundt, and Hering illusions (Figure 1). He found that high imagers experienced these illusions with the equivalent magnitude in imagery and vision, whereas low imagers experienced the illusions only in the visual presentation. Ishai and Sagi (1995) compared the patterns of facilitation and interference onto a Gabor stimulus by flanking masks seen or imaged. They found that the imaged facilitation is observed at all distances, with the same asymptotic magnitude as that of the visual flanks. Thus, in some situations a mental image seems to affect perception of other features in the same way a visual image would.

Yet there are situations in which mental images do not produce the same effects as visual images. Typically, it appears that effects resulting from prolonged visual exposure were not replicated with prolonged imagery. Over and Broerse (1972) compared the influence on vertical line detection of viewing or visualizing vertical or horizontal gratings. Viewed gratings masked the target line in a strongly orientation-dependent fashion. Imaged gratings produced some masking, but it was homoge-

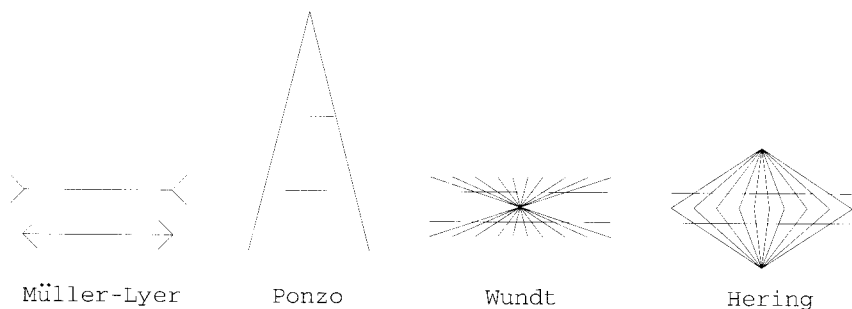Müller-Lyer   Ponzo   Wundt   Hering

Figure 1. Illusions imaged with an imaged context; these illusions are also obtained by good imagers when they visualize rather than see the illusion-inducing contexts

neous across orientations. Rhodes and O'Leary (1985) presented three experiments testing how previous imaging of gratings affected grating detection. For those experiments as well, results from imaging gratings were different from those obtained by seeing the grating.

Thus, identical effects of imaged context and visual context appear to be limited to the cases in which imagery is ongoing during stimulus inspection, supporting the idea that ongoing top-down activity rather than priming of neural pathways is responsible for the influence of imagery on vision.

The last group of studies attacks directly the question of content-dependent interaction between imagery and vision. These studies use a mental image cue before a similar or a different image but lack controls for the main alternative explanation of the observed effects: changes in perceptual strategy.

Peterson and Graham (1974) compared cueing by listening with cueing by visualizing on a forced-choice recognition task. Incompatible cueing lowered the hit rate in the imaging group only. Both types of compatible cueing raised the hit rate from 50% to 80%. The authors suggested that imagery appears to interfere negatively with perception when the image formed is incongruent with the perceptual signal and that the verbal cues provide a priming effect with or without imagery. However, as Farah (1985) pointed out, it is possible that their imaging task caused eye movements or particular attention to parts of the screen, which could account for the effects.

Farah (1983) showed that auditory imagery could cause facilitation and interference to audition. The results showed that auditory imaging of the correct signal lowered the detection threshold. Farah (1985) followed up on her previous conclusions and, coming back to the visual modality, focused on the content-specificity of imagery–vision interaction. She set up two experiments, testing the influence of visualizing shape or shape and location. The results showed that the correct visualization at the correct place caused better performance than any other combination. There was also a strong main effect for the location alone. Farah concluded that imagery shares perceptual representation at the low level, the "array" level in Kosslyn's terms. Farah instructed participants to maintain central fixation; however, covert attention confounds could still be present in this study because it did not include a visual cueing control condition. A follow-up study by Farah (1989) showed that spatial attention indeed plays too big a role to be ignored as an alternative explanation of her 1985 results. Heil, Rössler, and Hennighausen (1993) replicated and extended the study of Farah (1989). Their results also emphasized the importance of spatial attention allocation.

In conclusion, this last group of studies showed a content-dependent interaction, but there could be confounds related to eye movements or spatial attention. Until the synergy of spatial attention and object imagery is better understood, one should be wary of tasks involving simultaneous imaging and covert attention. Besides the lack of this control, all these studies used a reduced set of mental images for cues. This could entangle the effects of imagery with the effects of learning and fastidiousness. The studies of Peterson and Graham (1974) and Farah (1983, 1985) stand out as the best examples of content-dependent interaction of imagery and perception, but these three studies should be considered as the background for new experiments rather than taken as conclusive.

In this context, our goal is to show that the content of imagery is critical for low-level interaction. Two different types of cue, an ongoing mental image or a prior visual image, are used before a visual task. Cue validity in the main condition, "imagery cue, then visual task," is compared with the control condition, "visual exposure, then visual task." In the main condition, simultaneous imagery and perception superimpose top-down and bottom-up sources of activation, whose similarity will distinguish between synergy and interference. This effect will not arise when the cueing is from prior visual exposure because in that case there is no superposition from two sources. Therefore, rather than trying to nullify the effects of knowing what the stimulus can look like, our experiment allowed these confounds to be present in both the main and the control conditions. The expected result is that on top of the common effect of cue validity, imagery cueing will show a supplementary effect when compared with the control condition.

A secondary goal of this study was to confirm the hemispheric lateralization of imagery. Mental imagery appeared at first to involve more heavily the left hemisphere (Goldenberg, Podreka, Steiner, & Willmes, 1987; Peronnet, Farah, Weisberg, & Monheit, 1989; Charlot, Tzourio, Zilbovicius, Mazoyer, & Denis, 1992; d'Esposito, Detre, Aguirre, Stallcup, Alsop, & Farah, 1997). However, this interpretation is mitigated, as it appears that each hemisphere assumes different tasks (Kosslyn, 1994; Loverock & Modigliani, 1995). For example, Kosslyn (1994) suggested that when the composing elements are connected by quantitative, coordinate relationships, the right hemisphere is used preferentially, and when the connections are specified in general categories (e.g., "above," "behind"), imagery requires the left hemisphere. Neuropsychological evidence increasingly supports this model (Kosslyn, Thompson, Gitelman, & Alpert, 1998; Rajj, 1999). The neuropsychological evidence is completed by behavioral evidence; for example, Michimata (1997) compared RTs and error rates for a categorical or a coordinate judg-

ment on clock stimuli, similar to ours, either seen or visualized in the left hemifield or the right hemifield. They found a right visual field advantage for the categorical judgment, an effect that is present for both vision and imagery but significant only for vision stimuli in the first of three blocks. Reciprocally, better coordinate judgment is made on stimuli appearing or visualized in the left hemifield, but significance is also marginal and limited to the imagery condition. In addition, Lempert (1987, 1989) found that imaging decreased tapping performance more strongly for the right hand than for the left hand, suggesting a competition for resource allocation. This last study hinted that we might find a difference in the pattern of results obtained from right-hand responders and left-hand responders.

EXPERIMENT

METHOD

Participants

Sixty-three volunteers aged 18–30 took part in the experiment to earn class credit. Thirty-nine were male and 24 were female. Right-handers greatly outnumbered left-handers (59 to 4). All participants but one were fluent English speakers. All participants had normal or corrected-to-normal vision.

Procedure

Each participant responded to 160 trials. Each trial was a sequence consisting of a cue, a blank screen, and a stimulus (Figure 2). The cue was either a picture of a clock or verbal instructions to image a clock (e.g., "Please visualize a clock marking one twenty"). Participants knew from the beginning what kind of dial to visualize because they had a preliminary practice with noncued trials. After the cue was shown for 1,000 ms, a blank screen was presented for 3 s. This time allowed a participant either to think about the clock picture or to visualize the clock corresponding to the verbal instructions. Then came the test stimulus, a clock dial covered by a half gray plane (Figure 2). The participant had to press a green button if both of the clock hands were in the gray
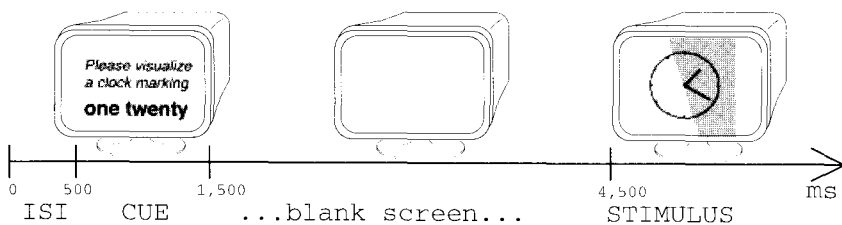

Figure 2. Timeline of a trial

plane (a "yes" response) or press a red button if one or both hands were in the white plane (a "no" response). The "yes" and "no" labels were assigned a posteriori for analysis. Careful precautions were taken to prevent the participants from using such verbal labels. The two response buttons were made of two round 3-cm pads (one red, one dark green) that were glued onto the "E" and the "O" of a keyboard stripped of other keys. They bore no inscription, nor were they ever called "yes" or "no" in front of the participants.

Eighty clock images generated randomly were used so that each image was used once in the vision condition and once in the imagery condition (condition order was varied between participants and was a factor in the analysis). Different images were generated for the practice and baseline trials. The orientation of the gray plane was varied randomly from trial to trial to prevent "solve-on-cue-and-match" strategies. When the gray plane was added to generate the stimuli, the clocks were assigned randomly in the "yes" or "no" conditions under two constraints: that "yes/no" in the vision block was independent of "yes/no" in the imagery block and that the response could not be predicted from the angle of the hands. So, unlike in perfectly random generation, a clock showing 12:30 (where the hands are five and a half divisions apart) had the same .5 probability of being a "yes" as a clock showing 1:10 (where the hands are less than one division apart). This was done to negate a possible bias from participant strategy or implicit learning.

Half of the cues were valid (showing the same time as the clock stimulus) and half invalid; this proportion was chosen because unequal proportions would create nonhomogeneous variance between the primed and the distracted conditions.[2] Maxwell and Delaney (1990, p. 471) explicitly warned that "for within-subject designs, there is a homogeneity of treatment-difference variances assumption." Also, a 50% cue validity is noninformative only when the cue regards a binary distribution. In the case of this experiment, the cueing indicated that the corresponding clock would occur 50% of the time, and any of the other clocks (potentially, 143 possible clocks) would share equally the remaining 50%. Therefore, it was felt that a 50% cue validity would be sufficient to motivate the participants.

Participants attended to the stimulus with one hand ready between the green and the red button and had to press either button with that same hand and position it back between the buttons during the interstimulus interval. About half the participants were instructed to use their right hand and the other half their left hand. The position of the buttons was the same for all participants: green button (two clock hands in the gray) on the right, red button on the left. The cases with a valid cue were called primed trials, whereas the cases in which cue and stimulus times differed were called distracted trials. Figure 3 shows examples of all possible trial types. The order in which the trials were presented was a between-subject factor: For a minority of participants,[3] trials were interleaved, alternating the cue type; for most participants, trials were blocked by cue type, and order of block administration was varied between participants. Clock dials were 52 mm in diameter, the gray plane was a random half of a 77 mm square, and participants were seated approximately 70 cm from the screen.

There were no explicit controls for eye movements because we hoped that eye

Figure 3. The eight cases of trials

movement strategies would develop identically for both cueing conditions. There was a concern that eye movements related to the process of visualization might interfere with the movements planned strategically. However, because of the 3-s delay between cue and stimulus, a delay calibrated by several pilot experiments, we think that this interference was over by the time the stimulus came up.

The experiment was presented on a PowerMac using *Superlab* software. The data were collected on the PowerMac in Excel spreadsheets and analyzed using SPSS software. Before the experiment, the participants were given three pretests inspired by Craik and Dirkx (1992) and then were given a practice of 20 trials with the stimulus only (no cue) to familiarize them with the task. The practice stimuli were not reused in the experiment, but after the experiment they were run again to get an estimate of the noncued response.[4] We knew from the pilot study that without cueing the RTs would be much shorter and did not want to waste too many trials on collecting baseline data.

The participants were interviewed after the experiment. The interview had no fixed format; only after interviewing several participants did we realize that where they were looking on the stimulus (the gray, the white, or the hands) could be a relevant factor, and we then systematically asked the participants what part of the stimulus they were looking at.

Because the base RTs varied with each participant and task, the dependent measure was the difference in RT between the primed and distracted conditions. Differences were calculated by obtaining a global RT for each[5] of the eight

cells described in Figure 3, then subtracting the primed RT from the distract-ed RT to get the effect in each of the four conditions of cue and response. Because it was not possible for the cue to be neutral, we did not use a neutral cue, and the dependent variable was an unspecified combination of facilitation in the primed case and inhibition in the distracted case.

RESULTS

The difference in RT[6] between primed and "distracted" cases was analyzed in a mixed-design analysis of variance with cue (vision, imag-ery) and response (yes, no) as within-subject variables and hand used (right, left), task order (vision first, imagery first, interleaved), sex (male, female), and fatigue (midterm, exam week) as the between-sub-ject variables.

Imagery cues had more effect than visual cues

A main effect for cue was found, $F(1, 42) = 4.57$, $p < .05$, showing significantly more cueing in the imagery than the vision condition. The direction of this finding is counterintuitive: Visual exposure to the cor-rect clock gives a smaller advantage (10.8 ms) than ongoing imagery of this clock (20.6 ms). This main effect is qualified by two three-way in-teractions, one to which we will return later, involving response and hand use, $F(1, 42) = 7.33$, $p < .05$, the other involving sex and hand use, $F(1, 42) = 4.22$, $p = .05$.

Additional analysis of the two cases of "no"

The effect of cueing becomes clearer when the "no" category is split into two. As shown in Figure 4, in one case a "no" response can hap-pen when one hand is in the white area and the other hand is in the
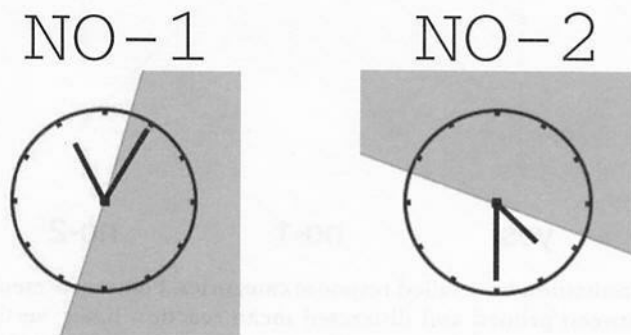

Figure 4. The two "no" responses: "no-1" (clock hands astride the plane bound-ary) and "no-2" (both clock hands out of the shaded plane)

gray area ("no-1" response); in the other case, a "no" response can occur when both hands are in the white area ("no-2" response). We had not anticipated that these two kinds of "no" responses would differ, but an examination of the data in terms of the imagery and vision conditions (shown in Figure 5) made it clear that they did. We asked two questions of these data: Is the facilitation effect of imagery greater than vision for the three different responses? Are the effects significantly different from zero, which would indicate facilitation? Looking first at the "yes" responses, imagery showed greater facilitation than vision, $F(1, 62) = 5.37$, $p < .05$; however, only imagery showed significant facilitation, $p < .01$. The "no-1" response showed that imagery and vision did not differ significantly, $p = .88$, and both showed facilitation effects, $ps < .01$. However, even though some facilitation was shown in the "no-2" response (the pattern is similar to that of the "yes" response), the scores for the imagery and vision conditions did not differ significantly from zero or from each other, $ps > .20$.
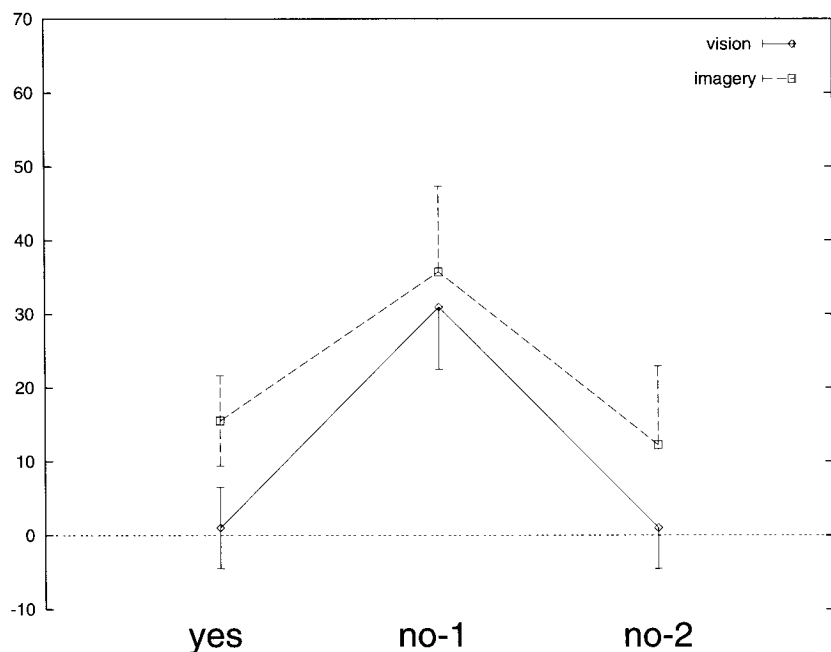
ms

Figure 5. Facilitation by detailed response categories. Points represent the difference between primed and distracted mean reaction times; vertical lines depict standard errors of the means (for readability, only half the bar is shown when overlap would occur)

Interaction involving the use of the right or left hand

As can be seen in Figure 6, the first reported interaction (cue × response × hand use) is that for imagery cueing, in the "no" situation, there is a large difference of effect between the right hand responding and left hand responding. Participants who used their left hand were hindered by valid cues by an average of 5 ms, whereas participants who used the right hand were facilitated by 50 ms. Follow-up analyses (Keppel, 1991) were conducted by separating the imagery and vision data into two separate two-way ANOVAs with response as the within-subject factor and hand used as the between-subject factor. For the imagery data, a significant interaction was found for response and hand used, $F(1, 61) = 13.4$, $p < .001$. One-way analyses of variance showed a significant effect for the "no" response condition, $F(1, 62) = 15.97$, $p < .001$, indicating significantly more facilitation with the right ($M = 50.2$ ms) than the left hand ($M = -5.1$ ms). For the "yes" responses in the imag-
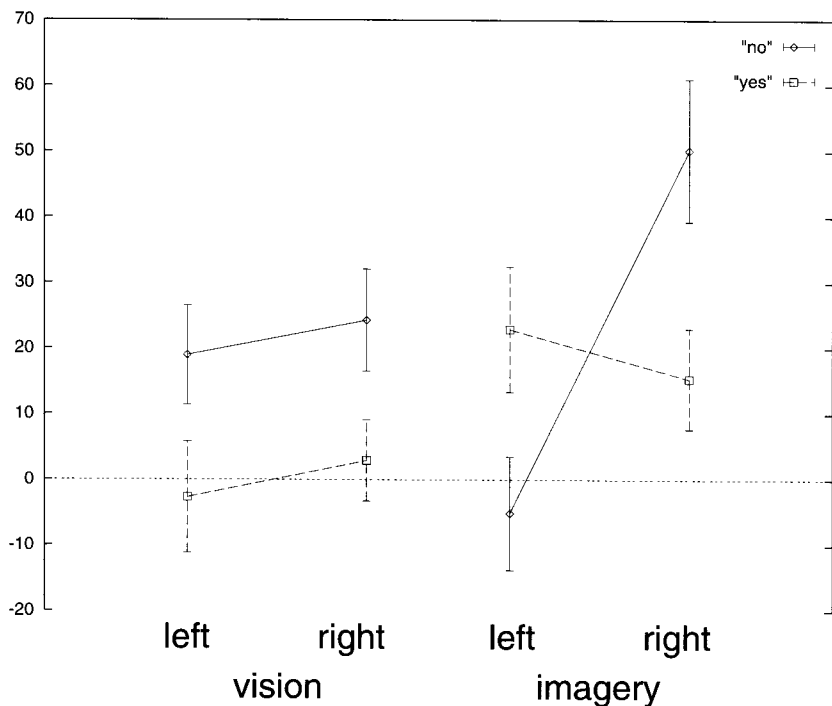

Figure 6. Three-way interaction between cue type, response, and hand used. Points represent the difference between primed and distracted mean reaction times; vertical lines depict standard errors of the means

ery condition, right and left hand use did not differ ($M = 15.3$ ms and 22.9 ms, respectively; $p = .54$). Results of the follow-up analyses on the vision data showed an overall greater facilitation for the "no" ($M = 21.6$ ms) than for the "yes" responses ($M = 0.1$ ms), $F(1, 61) = 7.78$, $p < .01$. This finding did not interact with hand used ($p > .10$), indicating a consistent pattern in the vision condition using either the right or the left hand.

Comparing "no-1" and "no-2" responses informs us that this three-way interaction must occur at or after the decision level: The interaction of hand used, response, and cue type appears to affect both the "no-1" and "no-2" cases. As shown in Figure 7, the "no-1" and "no-2" have parallel increments from the group using their left hand to the group using their right hand. The main effect in favor of the right hand is significant, $p = .006$, and so is the response–hand interaction, $p = .033$. Because both "no-1" and "no-2" are affected by the gain in facilitation from the left
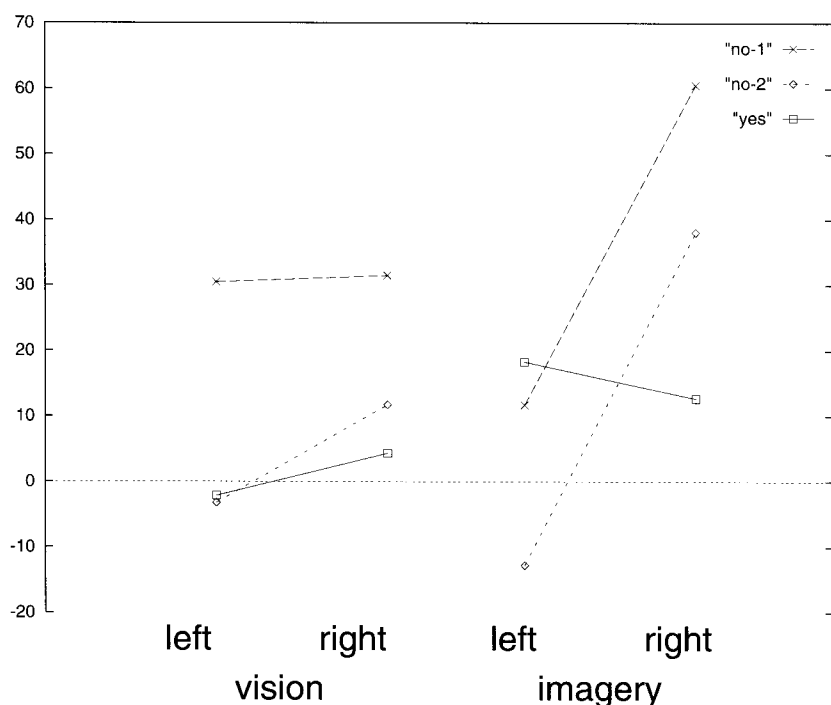
Figure 7. Response × hand used × cue type interaction with detailed categories of response. Points represent the difference between primed and distracted mean reaction times; error bars are omitted for readability. Compare to Figure 8 to see this is not a ceiling effect

to the right hand, it is likely that the interaction occurs at a level where the image is integrated in a representation no longer dependent of the difference between "no-1" and "no-2."

Other interactions

The second three-way interaction, involving cue, sex, and hand used, can be seen in Figure 9. Follow-up analyses were conducted separately for the women and men with response and hand used as the factors. For the male participants, imagery is more sensitive to hand used for the response, $F(1, 36) = 8.05$, $p < .01$. That is, imagery showed more facilitation for the men who were instructed to respond with their right hand, $M = 37$ ms, than their left hand, $M = 3$ ms. Hand used in the vision condition did not significantly differ, $p = .67$. The corresponding differences are smaller for the female participants, 26 ms right hand and 17 ms left hand, and not significant. Men respond more quickly than

ms
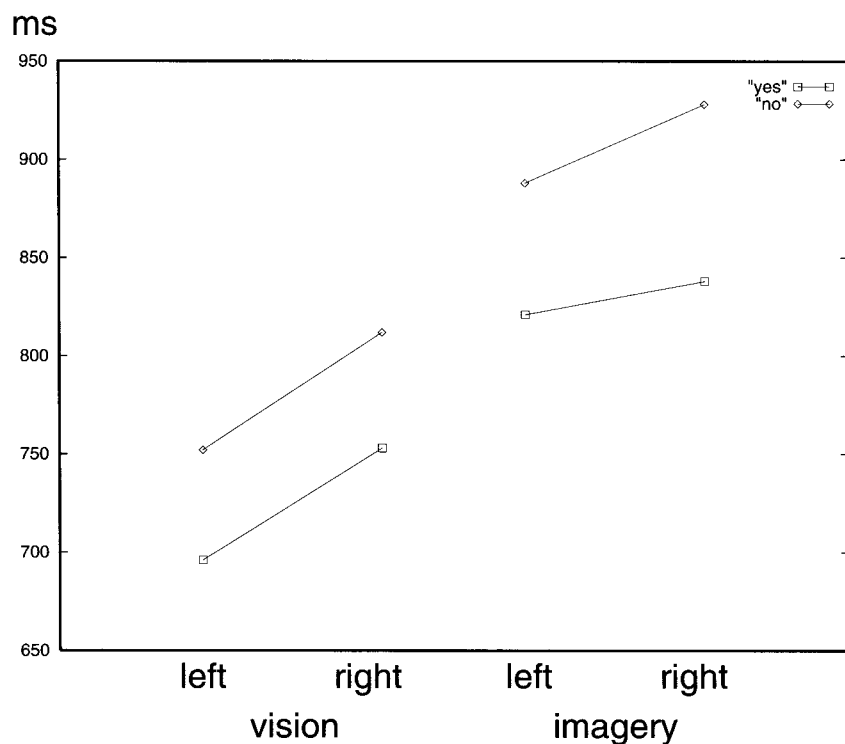
Figure 8. Base reaction times and error rates by cue type, response, and hand used. Points represent the mean reaction times, with primed and distracted combined

ms
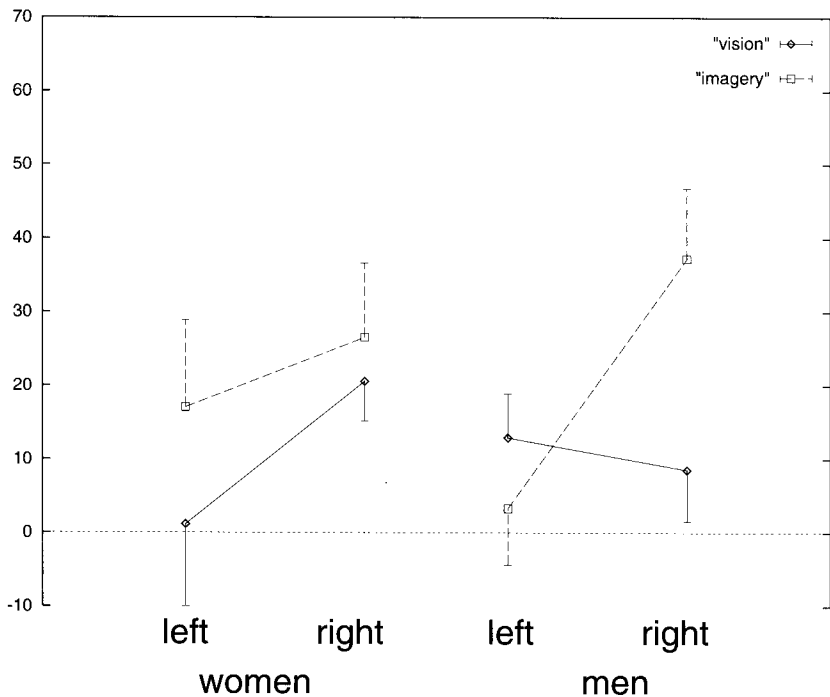
Figure 9. Interaction of sex, cue type, and hand used. Points represent the difference between primed and distracted mean reaction times; vertical lines depict standard errors of the means

women on tasks involving visuospatial working memory (Lorig-Meier & Halpern, 1999). It is possible that the effect we observed reflects a tendency for men to have a stronger preference for forming composite images using their left hemisphere, a preference that would result in faster performance; however, more evidence is needed before such a conclusion can be accepted.

A two-way interaction was also found involving order and fatigue, $F(2, 42) = 4.52$, $p < .05$. Follow-up analyses showed that the participants who received the vision task first and who were tested during the term showed more overall facilitation for the imagery than for the vision condition, $M = 27$ ms, than those who were tested during finals, $M = 5$ ms; $F(1, 26) = 5.12$, $p < .05$. No differences in terms of fatigue were found for the participants who were tested first in the imagery or the interleaved conditions, $ps > .30$. Overall, participants during the exam week had less facilitation, $M = 12.4$ ms, than participants in the term, $M = 19.6$

ms; $F(1, 42) = 5.4$, $p < .05$. A marginally significant interaction, $p = .06$, between fatigue and cue type shows that participants in the exam week had no particular advantage of imagery over vision.

The effects observed are not ceiling effects

The basic RTs (Figure 8) for imagery cues are longer than for visual cues (868 ms vs. 753 ms), so it could have been the case that the magnitude of the observed facilitation was related to the mean primed and distracted RTs in that condition. There are two possibilities: a strong hypothesis that there was an absolute correlation between facilitation and RT and a weak hypothesis that there was a person-dependent correlation between facilitation and RT. An index of correlation is the coefficient of correlation between the facilitation (distracted RT – primed RT) and the mean RT in the corresponding condition ([distracted RT + primed RT]/2). The correlation coefficient of .07 is too small to support the strong hypothesis that the magnitude of facilitation depended on the base RT. However, there could have been a trend within each participant's data, masked in the global correlation by between-subject variance. To test numerically the weak hypothesis, correlation coefficients were computed individually for each participant and subjected to a t test to find out whether their mean would be significantly positive. The mean of these coefficients is .047, about a seventh of their standard deviation, $t = 1.045$, $p > .5$. Thus, differences of facilitation observed in this study can be trusted to reflect true effects, even when the conditions compared have very dissimilar base RTs.

What can we learn from the difference in basic RTs? Analyzing the grand means,[7] with response ("yes," "no-1," "no-2") as the within-subject factor and order as a between-subject factor, we find that the baseline RT is 725 ms, a little shorter (marginally significant, $p = .053$) than the average RT with visual cues, 753 ms, and much shorter than the average RT with imagery cues, 868 ms, $p < .001$. When considering only the data from participants with interleaved trials, the pattern changes: The baseline is still significantly shorter (673 ms, $p = .024$), but vision and imagery means are much more similar (790 ms and 806 ms, respectively). This pattern of results suggests that the difficulty of the imaging condition mobilizes resources and slows down the response. When the trials are interleaved, the resource is mobilized for all trials, resulting in a general slowdown. When the trials are blocked by cue type, only the imagery cue block shows the slowdown. One of the reasons that could account for that slowdown is the task of reading the time.[8] In this context, it is good to note that our main result also holds in the interleaved condition.

Error rates

Table 1 shows the relevant error rate data. The effects observed do not approach significance. Nevertheless, they confirm the main effect observed on RTs. Valid cues diminish the error rates more for imagery than for vision, and this is particularly true of the "yes" and "no-2" subcases.

It should also be noted that the error rates generally followed the RT patterns. When RTs were shorter, errors were lower. The pretests did not correlate with any of the effects but instead correlated with the baseline RT, suggesting that they measured alertness more than any particular skill. An examination of these data showed that the participants who were tested during exam week showed a weaker performance (longer overall RTs and weaker effect) than those tested early in the term. This tendency is the reason for including a fatigue factor in the analyses, with fatigue operationalized as one of two stages: finals week and any other time.

DISCUSSION

In this study, we measured the difference of RT between a primed and a distracted response to a visual task, which consisted of detecting whether two hands of a clock dial fell into a gray plane. The gray plane appeared at random orientations, nullifying the usefulness of response anticipation. The dependent variable was the gain in RT in the case of a valid cue. We compared the facilitation induced by visual priming with that from imagery priming. Therefore, the argument stems from a difference between differences. Taking RT(distracted) − RT(primed) as a

Table 1. Basic error rates (ER), followed by the effect: ER(distracted) −ER(primed), with the "no-1"/"no-2" split

	Basic error rates		
	Yes	No-1	No-2
Baseline	.008	.048	.020
Vision cues	.167	.151	.095
Imagery cues	.349	.246	.103

	Effect of cueing on error rate		
	Yes	No-1	No-2
Vision cues	−.016	.048	−.032
Imagery cues	.0635	−.048	.143

dependent variable aims at reducing subject-dependent variations. Comparing priming by a visual image with priming by a mental image creates a situation in which most possible causes for the observed effects are common to both types of cueing. Different effects for visual and imagery cueing thereafter must be attributed to causes that are present in one condition and not the other, such as low-level interaction between imagery and vision.

We observed that imaging a valid cue produces significantly more facilitation than seeing a valid cue. For "yes" stimuli, valid imagery cues show 19 ms of improvement, whereas valid visual cues show no improvement. In addition, the effects of visualization interact strongly with the hand used to respond, confirming the present theories of imagery lateralization. In the imagery condition only, hand and response strongly interacted, revealing a difference of 55 ms between the facilitation obtained with the left hand (–5 ms) and the right hand (+50 ms). This very robust effect holds for both women and men and for the four left-handed participants and suggests a strong hemispheric asymmetry in imaging.

The task is indeed what it pretends to be

Imagery cueing did require imagery. First, it did not rely exclusively on memory. The stripped-down clock images are not exactly like real clocks, and interviews during pilot versions of the experiment made it clear that participants were building the mental image rather than remembering past images. The clock dial is a stimulus of reasonable complexity, and by varying the time shown we avoided the effects of overlearning. Second, the clock dial is not an overpracticed symbol, such as a letter. When the item to visualize is overused as a symbol, it is easy for participants to believe they are visualizing it when they think very hard about it. (In a previous unpublished experiment we found that the majority of participants who had extensive exposure to a set of four symbols, including a six-branch asterisk, would draw an eight-branch asterisk in the postexperimental debriefing.) Finally, the image to be visualized was complicated enough to tap imagery but not complicated enough that visualization would be impossible for all but the most talented participants. For all these reasons, we are confident that the imagery cueing really involved imagery.

The response to the stimulus was as free as possible from various interferences. Being as much a visual judgment as possible, the response excluded verbal behavior and the possible interferences that could have resulted from it. The response buttons bore no inscription, nor were they ever called "yes" or "no" in front of the participants. Pilot studies showed that when participants cast the task in terms of a verbal yes–no

question, they would tend to assign a "yes" to the primed situation and a "no" to the distracted situation, and this yes–no reaction was interfering with the task. In our final data, the interference pattern observed in those pilot studies did not show up, showing that the instructions prevented the participants from phrasing the task explicitly as a yes–no question. The task also stayed clear from any type of matching task. Had the cue created a statistical bias in favor of one particular response, participants might have accomplished the task ahead of time, on the cue, then performed a simple match-to-cue to find out whether the prepared response was sufficient; only if the match failed would the participants accomplish the task using the stimulus. To avoid this, the position of the shadow was varied so that the response depended on the whole stimulus, including the shadow, rather than on the clock dial alone, and the planes were carefully distributed to give an equal probability of "yes" and "no" response regardless of the clock geometry.

Discarding possible confounds: Ceiling effect, afterimage, and semantic code

One possible source of the effect would be that the slower condition shows more facilitation if the effect is proportional to the RT. Inversely, the effect could appear in the fastest conditions and be swamped in the slower ones. In either case, any monotonic relationship between the facilitation and the basic RTs would increase the correlation coefficient between the dependent variable (facilitation) and the basic RTs. This coefficient is shown to be very small (.047; see *Results*), arguing against this confound.

Is it possible that the stronger facilitation with imagery is observed because in the vision condition a negative afterimage interferes with perception of the stimulus? If so, the facilitation would be consistently stronger in different subcategories. However, the difference in facilitation can be broken down and traced to two particular effects: a difference on the "yes" and "no-2" but not in the "no-1" and a strong difference between the right-hand responders and the left-hand responders for both "nos" but not for "yes." The afterimage hypothesis cannot account for these effects. Therefore, afterimage interference did not play a substantial part in our result.

The pattern of results could not be accounted for by processing at a symbolic level. The pattern of results implies interaction at two levels: a lower level, evidenced by the "yes," "no-1," and "no-2" analysis, and a higher level, shown by the three-way interaction of hand used, cue, and "yes" or "no" response. If symbolic coding is involved, it should be at the higher level. Another mechanism must still account for the interaction at the lower level.

"No-1" specificity

For correctly primed stimuli, the mean RTs for "no-1" and "no-2" are almost equal (vision, 774 and 775 ms; imagery, 898 and 899 ms), but when the cue is misleading, the response to "no-1" is slowed much more (vision, 809 and 778 ms; imagery, 931 and 912 ms). Therefore, there is a reason why seeing a distracted stimulus will systematically slow the response to a "no-1" but not a "yes" or "no-2" stimulus. There has to be a particular type of expectation, knowledge driven because it appears with both types of cues, that when disappointed affects only "no-1." By most subjective reports, the "no-1" stimulus is visually the most complicated. This suggests that the reason for the slowing of "no-1" is that only this type of stimulus makes the cognitive load of dealing simultaneously with the disappointed expectation and the stimulus too complex.

We infer two things from this: that there is a knowledge-driven effect and that it does not affect "yes" and "no-2" stimuli. We speculate that the expectation created by the cue is based on spatial attention (e.g., reshaping or shifting covert attention or planning such shifts, or saccades). Whether this speculation is true or not, we have shown on the visual cue control that in some situations the effect of a congruent cue does not result from mental imagery. That could have happened in studies previously cited, which did not have such a control. Finally, lack of effect from a visual cue shows that the knowledge-driven effect does not affect "yes" and "no-2" stimuli.

Is it imagery that interacts with vision?

Having observed that for "yes" and "no-2" stimuli, visualizing the valid cue makes judgment faster than visualizing the invalid cue, we can use the control condition to argue that it is really imagery that causes this effect.

A first possible confound would have been classic priming. A priming effect could be obtained if a representation used for the task is easier to reactivate than to activate. For example, the semantic representation of time could be primed by the cue, and, being more easily reactivated, it would facilitate judgment on the stimulus clock. Obviously, such a mechanism would have the same effect, if not a better effect, with the visual cue, because seeing a clock activates all representations used for processing the stimulus. Imagery could at best do just as much. Therefore, observing a stronger effect for imagery cannot be explained by classic priming.

A second possible confound is the use of various strategies for acquiring information. Besides eye movements, pointed out by Farah (1985), covert attention phenomena also come to mind, such as the preallocation of spatial attention between the expected position of the clock

hands. However, because all of these strategies are based on expecting a particular clock, they should facilitate the judgment with a visual cue as much as the judgment with an imaged cue. This is not the case, so the observed facilitation with an imaged cue cannot be explained by information acquisition strategies.

Because the facilitation with imaged cues appears genuinely stronger than that with visual cues, and because that difference cannot be accounted for by other mechanisms, the remaining explanation is that imagery indeed affects the judgment more than prior vision. Thus, the interaction between imagery and judgment depends genuinely on whether the cue imaged is the same clock as the stimulus clock.

What is the balance between facilitation and inhibition?

Is there true facilitation from congruent stimuli? Perception during imaging should be compared with the situation without imaging; when this is done, perception during imaging is always worse, presumably because the mental load of imaging hinders the response. Including a neutral condition with an equal mental load, especially a mental load equally taxing to the same subsystem, looks almost impossible. In our experiment the content-dependent effect favors congruent imaged cues, but the RT is longer in the imaging condition than in the visual one, even with congruent cues (753 ms vs. 869 ms). However, that difference is smaller when the trials are interleaved rather than blocked (770 ms vs. 790 ms). For participants tested with interleaved trials, vision and imagery times come much closer, and in the primed situation the RT is the same for both types of cueing (804 ms), an equality that holds in subpatterns caused by response and hand used. Thus, if the cueing clock was the correct clock, whether those participants had seen or visualized that clock did not affect their time. All the difference came from the distracted priming situation. Therefore, we believe that the balance is tipped toward inhibition in the case of bad cues rather than toward facilitation in the case of good cues.

Observing only inhibition in our experiment does not mean that true top-down facilitation does not occur in other situations. We believe that imagery does facilitate perception in difficult or ambiguous settings; however, that is not what this experiment shows. Such facilitation probably will be very difficult to isolate experimentally from various kinds of priming or biases. Obtaining the answer here may require more invasive techniques, such as event-related potentials, magnetoencephalography, or functional magnetic resonance imaging.

Hemispheric difference

Using one hand or the other to press the buttons was unimportant, except for imagery cueing, for the "no" responses, where it caused a

strong effect. With imaged cues, on "no" stimuli, much higher facilitation was observed for participants using their right hand than for participants using their left hand. A more detailed analysis also shows that the effect influences "no-1" and "no-2" equally. The major line of explanation remaining is that this difference originates from content-dependent interaction of imagery at higher, nonpictorial representations.

Lateralization of imagery, favoring the left hemisphere, has been studied particularly since the review of cases made by Farah (1984), who looked at the localization of brain lesions in patients having lost imaging ability. Many reviews and studies since have nuanced the idea that the left hemisphere is predominant in all cases of imagery generation, but at least it is predominant when images are made of composite parts. In his review of the lateralization of imagery in the left hemisphere, Kosslyn (1994, pp. 311–319) wrote, "The left hemisphere could perform imagery tasks that required adding high-resolution parts to an image . . . and could also perform imagery tasks that required only the global shape of an object. . . . In contrast, the right hemisphere was impaired when parts had to be added." Michimata (1997), also using clock stimuli, found in some conditions that imaging in the right hemifield allowed better categorical judgment. Considering that the imaging part of the task in our experiment is a typical task of assembling parts (the two hands, almost independently specified, plus the dial), our results are strongly in line with Kosslyn's conclusions. Somehow, the left hemisphere mediation of image generation must account for having content-dependent interaction strongly detectable with a right-hand response but invisible with a left-hand response.

It is tempting to attribute the interaction observed here to a response resource conflict, in line with what Lempert (1987, 1989) found when studying the effect of imagery on tapping. The "no" response, more complicated because it is a disjunction of cases, could require more resources to process. Participants using their right hand would put the highest demands on the left hemisphere, and inconsistencies between the cue and the stimulus could then take a higher toll. This response resource account cannot explain the decrement in left-hand response facilitation, but it could constitute part of the explanation.

Conclusion

This experiment shows that mental images interact with visual images. The interaction is between the contents, rather than just between the processes. Because valid cue facilitation is greater when the cues are imaged than when they are seen, the effect cannot come from confounds based only on stimulus knowledge. This suggests that the interaction results from the simultaneity of input from bottom-up vision and

top-down imagery, and that the similarity of the mental image to the visual image determines whether their interaction is synergistic or antagonistic.

Notes

1. We thank an anonymous reviewer for suggesting this quantitative estimate.

2. The within-subject noise, from trial to trial, is averaged out, but the magnitude of the residual error caused by that noise decreases in each cell as the square root of the number of trials in the average.

3. Out of 63. This condition was dropped after a participant complained about heavy difficulties.

4. Baselines were obtained for only 62 of the 63 participants because of a time constraint for one of them.

5. Preliminary analyses were conducted using three measures of the global RT for each cell: the median, the raw mean, and the mean of the RTs lying within two standard deviations of the raw mean (trimmed mean). Because all three measures showed the same general effects, only the trimmed means, which are less noisy, are presented.

6. A table with the full results, listing RTs for primed and distracted cases in all conditions and subconditions, can be found at http://cogsci.ucsd.edu/~rebotier/Clocksresults.html.

7. The one participant for whom the baseline data were missing has been excluded from these averages and the ANOVA.

8. We thank an anonymous reviewer for this suggestion.

References

Berbaum, K., & Chung, C. S. (1981). Müller–Lyer illusion induced by imagination. *Journal of Mental Imagery, 5,* 125–128.

Charlot, V., Tzourio, N., Zilbovicius, M., Mazoyer, B., & Denis, M. (1992). Different mental imagery abilities result in different regional cerebral blood flow activation patterns during cognitive tasks. *Neuropsychologia, 30,* 565–580.

Craik, F., & Dirkx, E. (1992). Age-related differences in three tests of visual imagery. *Psychology and Aging, 4,* 661–665.

Craver-Lemley, C., Arterberry, M., & Reeves, A. (1997). Effects of imagery on Vernier acuity under conditions of induced depth. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 3–13.

Craver-Lemley, C., & Reeves, A. (1987). Visual imagery reduces Vernier acuity. *Perception, 16,* 599–614.

Craver-Lemley, C., & Reeves, A. (1992). How visual imagery interferes with vision. *Psychological Review, 99,* 633–649.

d'Esposito, M., Detre, J., Aguirre, G., Stallcup, M., Alsop, D., Tippet, L., & Farah, M. (1997). A functional MRI study of mental image generation. *Neuropsychologia, 37,* 725–730.

Farah, M. (1983). Perceptual interference and facilitation with auditory imagery. *Perception & Psychophysics, 33,* 475–478.

Farah, M. (1984). The neurological basis of mental imagery: A componential analysis. *Cognition, 18,* 245–272.

Farah, M. (1985). Psychophysical evidence for a shared representational medium for mental images and percepts. *Journal of Experimental Psychology: General, 114,* 91–103.

Farah, M. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review, 95,* 307–317.

Farah, M. (1989). Mechanisms of imagery–perception interaction. *Journal of Experimental Psychology: Human Attention and Performance, 15,* 203–211.

Goldenberg, G., Podreka, I., Steiner, M., & Willmes, K. (1987). Patterns of regional cerebral blood flow related to memorizing of high- and low-imagery words: An emission computed tomography study. *Neuropsychologia, 25,* 473–485.

Heil, M., Rössler, F., & Hennighausen, E. (1993). Imagery–perception interaction depends on the shape of the image: A reply to Farah (1989). *Journal of Experimental Psychology: Human Perception and Performance, 19,* 1313–1320.

Ishai, A., & Sagi, D. (1995). Common mechanisms of visual imagery and perception. *Science, 268,* 1772–1774.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Kosslyn, S. (1994). *Image and brain: The resolution of the imagery debate.* Cambridge, MA: MIT Press.

Kosslyn, S., Thompson, W., Gitelman, D., & Alpert, N. (1998). Neural systems that encode categorical versus coordinate spatial relations: PET investigations. *Psychobiology, 26,* 333–347.

Lempert, H. (1987). Effect of imaging sentences on concurrent unimanual performance. *Neuropsychologia, 25,* 835–839.

Lempert, H. (1989). Effect of imaging vs. silently rehearsing sentences on concurrent unimanual tapping: A follow-up. *Neuropsychologia, 27,* 575–579.

Loring-Meier, S., & Halpern, D. (1999). Sex differences in visuospatial memory: Components of cognitive processing. *Psychonomic Bulletin & Review, 6,* 464–471.

Loverock, D., & Modigliani, V. (1995). Visual imagery and the brain: A review. *Journal of Mental Imagery, 19,* 91–132.

Mansfield, R. (1970). Latency functions in human vision. *Vision Research, 13,* 2219–2234.

Maxwell, S., & Delaney, H. (1990). *Designing experiments and analysing data: A model comparison perspective.* Belmont, CA: Wadsworth.

Michimata, C. (1997). Hemispheric processing of categorical and coordinate spatial relations in vision and visual imagery. *Brain and Cognition, 33,* 370–387.

Over, R., & Broerse, J. (1972). Imaged lines fail to induce contour masking. *Psychonomic Science, 29,* 203–204.

Perky, C. (1910). An experimental study of imagination. *American Journal of Psychology, 21,* 422–452.

Peronnet, F., Farah, M., Weisberg, L., & Monheit, M. (1989). Brain activity underlying mental imagery: Event-related potentials during mental image generation. *Journal of Cognitive Neuroscience, 1,* 302–316.

Peterson, M., & Graham, S. (1974). Visual detection and visual imagery. *Journal of Experimental Psychology, 103,* 509–514.

Piéron, H. (1920). Analyse du temps de latence sensorielle [Analysis of sensory latency times]. *L'Année Psychologique, 22,* 58–142.

Rajj, T. (1999). Patterns of brain activity during visual imagery of letters. *Journal of Cognitive Neuroscience, 11,* 282–299.

Reeves, A. (1980). Visual imagery in backward masking. *Perception & Psychophysics, 28,* 118–124.

Reeves, A. (1981). Visual imagery lowers sensitivity to hue-varying, but not to luminance-varying, visual stimuli. *Perception & Psychophysics, 29,* 247–250.

Reeves, A., & Segal, S. (1973). Effects of visual imagery in visual sensitivity and pupil diameter. *Perceptual and Motor Skills, 36,* 1091–1098.

Rhodes, G., & O'Leary, A. (1985). Imagery effects on early visual processing. *Perception & Psychophysics, 37,* 382–388.

Segal, S. (1971). Processing of the stimulus in imagery and perception. In S. J. Segal (Ed.), *Imagery: Current cognitive approaches* (pp. 69–100). New York: Academic Press.

Segal, S., & Fusella, V. (1969). Effects of imaging on signal-to-noise ratio, with varying signal conditions. *British Journal of Psychology, 60,* 459–464.

Segal, S., & Fusella, V. (1970). Influence of imaged pictures and sounds on detection of visual and auditory signals. *Journal of Experimental Psychology, 83,* 458–464.

Segal, S., & Gordon, P. (1969). The Perky effect revisited: Blocking of visual signals by imagery. *Perceptual and Motor Skills, 28,* 791–797.

Wallace, B. (1984). Apparent equivalence between perception and imagery in the production of various visual illusions. *Memory & Cognition, 12,* 156–162.