Feature-Specific Perceptual Processing Dissociates Action From Recognition

Liana E. Brown, Cathleen M. Moore, and David A. Rosenbaum
Pennsylvania State University

Does visual processing differ for action and recognition? To address this question, the authors capitalized on research showing that color is preferred over binocular disparity in the ventral (recognition) stream, whereas disparity is preferred over color in the dorsal (action) stream. Participants searched for oblique targets among vertical distractors in displays defined only by color or disparity. Action-task participants stamped the target with a handheld block, whereas recognition-task participants lifted the block through a target-compatible gap. Analyses of reaction time and time-varying hand orientation showed that disparity and color displays were processed equally efficiently during action, but disparity was processed less efficiently than color during recognition. The results suggest that visual processing differs for action and recognition.

One of the most intriguing hypotheses in recent research on human perception and performance is that visual information processing is different for action and recognition. This hypothesis is intriguing because it contradicts one’s intuition that visual experience is of a piece. What we see and how we see do not seem to depend on whether we are physically interacting with objects or looking at those objects for the sake of judgment or recognition. Nonetheless, as suggested by Milner and Goodale (1995), intuitions may be wrong in this case. According to Milner and Goodale and their well-known how and what hypothesis, seeing for the sake of acting (how) is not the same as seeing for the sake of recognizing (what).

In this article, we review the evidence that has been adduced for the how and what hypothesis. We discuss some of the weaknesses of this evidence and then introduce a new method for testing the hypothesis. The method capitalizes on differential selectivity for color and binocular disparity stimuli in the dorsal (how) and ventral (what) streams. We report an experiment that involved the use of this method. The results support the how and what distinction. In the General Discussion section, we review the main points of the study and consider alternative accounts of our findings, none of which does as well as Milner and Goodale’s (1995) model.

The How and What Hypothesis

According to Milner and Goodale’s (1995) how and what hypothesis, visual information processing is different for action and recognition. The grounds for this hypothesis are both functional and neuroanatomical. At the functional level, perception for action and perception for recognition make different demands on memory. When planning and controlling object-directed action, one must continually update information about the position, size, and orientation of the target object relative to the moving body. During recognition, however, one must gather and store object information relative to other objects in the scene for later identification or judgment. Because perception for recognition stores object information, whereas perception for action continually updates object information with little need for storage, perception for recognition should not be as vulnerable to periods of visual occlusion as perception for action. Furthermore, because perception for recognition gathers information relative to other objects in the scene, whereas perception for action gathers body-centered object information, perception for recognition should be more vulnerable to context effects than perception for action. Assuming that these distinctions hold, the how and what hypothesis is a dual-representation model in that perception for action and perception for recognition produce separate, task-specific representations.

At the neuroanatomical level, the how and what theory is linked to two distinct visual pathways that project from the primary visual cortex (V1) to extrastriate visual regions. The dorsal stream projects from V1 to the posterior parietal cortex, where cells appear to code multimodal, effector-specific, and effector-centered representations of external targets (Battaglia-Mayer et al., 2000; Caminiti, Ferraina, & Battaglia-Mayer, 1998; Graziano & Gross, 1996; Rizzolatti, Riggio, & Sheliga, 1994; Sakata & Taira, 1994). These representations are ideally suited for guiding action. The
ventral stream, in contrast, projects from V1 to the inferotemporal cortex. Lesion experiments in monkeys, accidental lesions in humans, and brain imaging studies in humans all indicate that the inferotemporal cortex is critical for object and face recognition (Haxby et al., 1993; Milner & Goodale, 1995; Ungerleider & Haxby, 1994; Ungerleider & Mishkin, 1982). These findings support the view that perception for action is mainly served by the dorsal stream, whereas perception for recognition is mainly served by the ventral stream. The dorsal stream is for *how*, and the ventral stream is for *what*.

The behavioral evidence for Milner and Goodale’s (1995) hypothesis includes an observed difference between impairments among patients with optic ataxia and impairments among patients with visual form agnosia. One patient studied in detail, a woman known as DF, experienced a comprehensive visual form agnosia after suffering diffuse brain damage centered in the visual association areas (Milner et al., 1991). DF’s performance on visual detection, discrimination, and matching tasks suggested that she was unable to recognize objects or drawings. However, when she manually interacted with objects, she could orient her hand and scale her grip appropriately (Goodale & Milner, 1992; Goodale, Jakobson, Milner, et al., 1994; Milner et al., 1991). Analyses of the spatial–temporal characteristics (kinematics) of DF’s hand movements showed that she made these manual adjustments before coming into contact with the objects (Goodale, Milner, Jakobson, & Carey, 1991). Thus, although DF could not make explicit perceptual judgments about objects, she could interact with the same objects physically. Milner and Goodale (1995) attributed DF’s preserved visual processing abilities to her relatively intact dorsal (action) stream, and they attributed her impaired visual processing abilities to her damaged ventral (recognition) stream.

As convincing as these lesion-deficit patterns may be, extrapolating from neuropsychological data to normal processing can be problematic, because spared abilities in neuropsychological patients may reflect characteristics of an impaired rather than an intact system. It is important, therefore, to test the how and what hypothesis by looking for action–recognition dissociations in neurologically normal individuals.

Two approaches have been taken with neurologically normal participants: the delayed-response method and the illusion method. In the delayed-response method, a target stimulus is presented and then removed for a controlled period of time, after which the response is signalled and performed. Actions performed during zero-delay conditions are compared with actions performed after longer delays (e.g., 1 to 4 s). If the representation for action reflects only currently available information, as the how and what hypothesis claims, then this representation should degrade over the course of the delay, and consequently delayed movements should differ from nondelayed movements. Experiments have shown that the kinematics of delayed movements do indeed differ from the kinematics of nondelayed movements (Goodale, Jakobson, & Keillor, 1994; Hu, Eagleson, & Goodale, 1999; Westwood, Heath, & Roy, 2000). However, these experiments have not provided clear support for a dissociation between perception for recognition and perception for action because they have not shown that perceptual judgments are more immune to delay than are comparable action measures.

With the illusion method, participants are shown stimuli that give rise to illusory percepts of size, distance, or orientation, and then they perform both perceptual judgments and actions on the stimuli. The logic behind this method is that although perception for recognition may be susceptible to the influence of context, perception for action may not be. Experiments involving this method have yielded inconsistent results. Some studies have shown that when participants physically interact with figures that give rise to false perceptions of length and size—the Müller–Lyer illusion, the Ponzo illusion, and the Titchener illusion—participants’ movements reflect the true length or size information embedded in the scene, as if the motor system resists illusory influences (Aghiota, DeSousa, & Goodale, 1995; Bridgeman, Kirch, & Sperling, 1981; Ellis, Flanagan, & Lederman, 1999; Gentilucci, Chieffi, Daprati, Saetti, & Toni, 1996; Haffenden & Goodale, 1998, 2000; Servos, Carnahan, & Fedwick, 2002; Westwood, Chapman, & Roy, 2000). However, other studies have shown that both action and recognition–judgment measures are vulnerable to size illusions (Franz, Gegenfurtner, Bulthoff, & Fahle, 2000), tilt illusions (Glover & Dixon, 2001), and position illusions (Smeets & Brenner, 1995).

A general concern about the illusion method is that direct comparisons of perceptual judgment and action measures may not be entirely valid. For instance, comparisons of recognition and action performance measures may not be valid because the way a scene is scrutinized may differ for action and recognition. Participants performing perceptual recognition–judgment tasks may be more inclined to survey the entire display than participants performing action tasks, who may inspect only the target and its immediate surround (Franz et al., 2000). Moreover, movements of the hand toward the target may confer advantages to action performance not available during recognition–judgment performance. For example, during movement, the hand may occlude the target’s surround, making the true position or size of the target easier to perceive. Also, the hand’s approach toward the target may allow the actor to make precise comparisons between hand position and target position as well as precise comparisons between grasp aperture size and target size (Ellis et al., 1999; Glover & Dixon, 2001). A strategy for addressing these problems has been to occlude the participant’s view of the hand and target display during the response (Franz et al., 2000; Haffenden & Goodale, 1998, 2000). However, this strategy may decrease the power of the recognition-task/action-task manipulation because participants are forced to view the object knowing that it will be occluded during the impending response and then rely on memory during the response. Thus, this strategy may encourage participants to draw on a neural system different from the one of primary interest.

**A New Approach**

Here we present a method for testing the how and what hypothesis that allows us to control for recognition–action task differences without removing vision during performance. It relies on the idea that if some stimulus features are processed more efficiently by one neural stream than by another, then perception of stimuli defined by those features will benefit from being embedded in tasks that engage the more efficient neural stream. Suppose, for example, that the ventral stream processes color more efficiently than binocular disparity and that recognition tasks engage the
ventral stream more than the dorsal stream. According to our reasoning, stimuli defined by color should be processed more efficiently than stimuli defined by binocular disparity when participants perform a recognition task. Conversely, if the dorsal stream processes binocular disparity more efficiently than color, and action tasks engage the dorsal stream more than the ventral stream, then stimuli defined by disparity should be processed more efficiently than stimuli defined by color when participants perform an action task.

The foregoing example finds support in neurophysiology. The stimulus selectivity of single cells and of distinct extrastriate cortical areas double dissociate the ventral stream from the dorsal stream. In the ventral stream there is a greater proportion of cells that prefer (i.e., are selective for) color-defined stimuli over binocular disparity-defined stimuli, whereas in the dorsal stream there is a greater proportion of cells that prefer disparity-defined stimuli over color-defined stimuli (Lennie, 1998; Livingstone & Hubel, 1988; Milner & Goodale, 1995; Van Essen & DeYoe, 1995; Zeki, 1978).

There is also precedent for a comparison of the perceptual processing efficiency of color-defined and binocular disparity-defined stimuli in the visual search literature. Cavanagh, Arguin, and Treisman (1990) found that targets defined by orientation were extracted more efficiently when the visual search display was defined by color than when the display was defined by disparity (see Figure 1). In the Cavanagh et al. study, participants were asked to search for an oblique (45°) target among vertical distractors and to report the target’s presence or absence by making a speeded choice response with a joystick. Participants tilted the joystick in one direction to respond yes or in the opposite direction to respond no. The slope of the function relating reaction time (RT) to number of display items was statistically indistinguishable from zero when the displays were defined by color, but the slope was significantly greater than zero when the displays were defined by disparity. Cavanagh et al. concluded that the binocular disparity processing channel is less finely tuned to orientation differences than is the color channel. To the extent that the visual search task of Cavanagh et al. primarily engaged the ventral stream, this difference in processing efficiency can be attributed to the ventral stream’s preference for color-defined stimuli over disparity-defined stimuli.

Cavanagh et al.’s (1990) finding raises an interesting question for the how and what hypothesis. Will the difference in search efficiency for color and binocular disparity displays change if participants perform an action task rather than a recognition task with the same stimuli? If recognition and action differentially engage the ventral and dorsal streams, then changing the task to an action task should alter the color–disparity difference. For example, if the relative engagement of the dorsal and ventral streams for action were completely reversed from that for recognition, then binocular disparity displays might be processed more efficiently than color displays during an action task. However, it may be too strong to assume that a complete reversal would occur in normal participants. Even if action tasks engage the dorsal stream more than do recognition tasks, they do not necessarily engage the dorsal stream more than the ventral stream. If the relative engagement of the two streams is sensitive to task characteristics at all, then the relative processing efficiency of color and binocular disparity displays may be different for the two tasks. This result would support the claim that visual processing is different for recognition and action.

To address this question, we repeated and extended Cavanagh et al.’s (1990) study by adding a condition in which participants responded by reaching out and orienting a handheld block to match the location and orientation of an on-screen target. Our aim in creating this action task was to mimic the orientation tasks performed well by DF, namely, posting a handheld card through an oblique slot (Milner et al., 1991) or stamping an oblique target with a rubber stamp (Goodale, Jakobson, Milner, et al., 1994). Milner and Goodale (1995) accounted for DF’s ability to perform these interactive tasks by arguing that the tasks engaged her intact dorsal stream more than remote matching tasks did. Insofar as this argument is correct, our action task should engage the dorsal stream more than a standard visual search task in our neurologically normal participants.

Method

Overview

Participants searched for a 30° left- or right-oblique target among vertical distractors presented in displays defined either by color or by binocular disparity. They responded to the target in one of two ways. In the recognition task, they indicated whether the target was oriented to the left or to the right by nudging a handheld block to the left or right. In the action task, they lifted the block up to the screen to match its orientation and location to the target’s orientation and location (see Figure 2). We pre-

Figure 1. Examples of displays used in the Cavanagh et al. (1990) experiments. The items in each display are defined by discontinuities in a single stimulus medium: binocular disparity (left) and color (right). The target differs from the distractors by orientation only. Therefore, target detection reaction time for each display should tap orientation processing efficiency within each stimulus medium channel.
dicted that in the recognition task, participants would search the color-defined displays more efficiently than they would search the binocular disparity-defined displays, as found by Cavanagh et al. (1990). In contrast, we predicted that the relative processing efficiency of color and binocular disparity displays would adopt a different pattern during the action task.

Participants

Twenty-eight neurologically normal individuals participated (mean age = 23.3 years, range = 17–37 years). Two participants reported being left-handed. Participants received monetary payment or course credit for

Figure 2. The recognition-task responses (top) and the action-task responses (bottom). Participants sat facing the screen, and the response box was placed on the table such that it sat between the participant and the screen, aligned with the participant’s midsagittal axis and the screen’s vertical midline. Each trial began when the participant depressed the home switch located on the floor of the response box below the light of the flashlight. The flashlight served as a divider between the left and right sides of the response box. When the display appeared in the recognition task, participants released the home key and lifted the bar through the response box opening compatible with target orientation (the left opening for a left-oriented target or the right opening for a right-oriented target). This movement was just large enough to break the photocircuit. In contrast, when the display appeared in the action task, participants released the home switch and brought the handheld block to the screen and matched its end to the location and orientation of the target on the screen. IREDs = infrared-light-emitting diodes.
taking part. All participants had normal or corrected-to-normal visual acuity. One participant was replaced because he was red–green color blind. Three other participants were replaced because they were apparently stereo blind, having failed the stereopsis screening test described subsequently.

**Apparatus**

Steroscopic depth stimuli were presented with a Stereographics CrystalEyes II stereo-display system. This system simulates depth by displaying alternating left- and right-eye images to the wearer of liquid-crystal shuttered glasses. Presentation of the computer display was synchronized with the shutters at a rate of 60 Hz.

Participants responded by manipulating a 3-cm × 1-cm × 10-cm wooden block that had a layer of rubber foam glued to the end opposite the grip. RT and response choice were recorded through the use of a custom-built response box that consisted of a small flashlight (Hubbellite, Model 2202-3A) mounted above the center of an open 15.0-cm × 10.0-cm electronic project box. The flashlight divided the box such that two 4.0-cm openings were created (see Figure 2). Each opening was flanked on the outer wall by a photocell that formed a closed photocircuit with light emitted by the flashlight. Participants began each trial by using the block to depress a 0.8-cm-diameter home switch mounted in the base of the box, directly beneath the bulb of the flashlight (see Figure 2). Release of the home switch triggered a signal used to measure RT. As the handheld block moved upward, it passed by the photocell and broke the photocircuit, enabling us to record through which of the two openings the block passed and when this occurred. The signals from the response box were transmitted to the game port of a Pentium-class IBM-compatible personal computer that was used to control all experimental events.

The apparatus was positioned such that, during the experiment, the participant’s right side faced the cameras of an OPTOTRAK 3200 3D motion analysis system (Northern Digital, Inc., Waterloo, Ontario, Canada). Two infrared-light-emitting diodes (IREDs) were positioned on the right side of the handheld wooden block, and a third IRED was positioned on the side of the computer monitor to provide a constant reference position. All IREDs were visible to the OPTOTRAK’s cameras at all times. The OPTOTRAK recorded the three-dimensional positions of the three IREDs at a frequency of 50 Hz. Although the IREDs were attached to the handheld block for all participants, IRED positions were recorded only for participants in the action task.

**Displays**

Participants were shown rectangular bar targets oriented either 30° to the left or to the right of vertical, along with one or five vertical distractors. Target and distractor items were presented at six fixed locations in a circular arrangement (diameter = 6.9°). A circular fixation point, 0.5° in diameter, was presented in the center of the display.

Two display types were presented that differed with regard to stimulus medium. Color displays consisted of red rectangular bars drawn on a 21.6° × 18.6° equiluminant green background that in turn was drawn on the black background of the monitor. Each bar subtended 1.0° × 3.0° from a typical viewing distance of 60 cm. Red–green equiluminance was determined with a pretest for minimum rate of perceived flicker (a description of which follows). Residual luminance differences between the red and green hues that may have arisen from imperfect matching or color adaptation were masked with 13% contrast random luminance noise (Cavanagh et al., 1990). The color displays were viewed through the liquid crystal display glasses, but the displays were created so that there was no disparity between the left- and right-eye views.

Analogous binocular disparity displays were defined by random-dot stereograms. Again, a large rectangular 21.6° × 18.6° grid of randomly filled light (2.71 cd/m²), medium (1.70 cd/m²), and dark (0.76 cd/m²) gray 1-pixel × 1-pixel dots was drawn on the black background of the monitor. There was 71.9% contrast between the light- and dark-gray fill values, and the overall luminance of this stereogram (1.72 cd/m²) matched, as closely as possible, the overall luminance of the color displays (1.60 cd/m² at the point at which red and green were physically equiluminant). The targets and distractors were drawn as rectangular bars overlaying the background rectangle. These bars differed from the background rectangle in stereoscopically defined depth only. The depth was created by presenting separate left- and right-eye images of the display such that all of the pixels except those defining the regions of the targets and distractors had an uncrossed disparity of 0.2°. This technique placed the target and distractor bars on the horopter and the background on a depth plane behind the horopter. The procedure ensured that the targets and distractors appeared on the horopter regardless of display type.

**Tasks**

Each participant completed one of two visual search tasks depending on the group to which he or she was randomly assigned (except for the constraint that there be an equal number of participants in both groups). The tasks differed only with respect to how the response was made. In both tasks, participants were asked to search for an oblique target bar among vertical distractor bars and to indicate whether the target was oriented to the left or to the right. All participants were asked to make their responses as quickly and as accurately as possible. Participants initiated each trial by depressing the home switch. They were asked to press the home switch until the display appeared.

In the recognition task, participants indicated target orientation by nudging the handheld block through the left opening of the response box if the target was oriented to the left or through the right opening of the response box if the target was oriented to the right (see Figure 2). Participants were not required to match the orientation of the bar with the target. This nudging response enabled participants to release the home switch and pass the bar by the photocell switch, allowing us to record RT and target orientation choice, respectively. We used the nudging response for the recognition task because it could be performed quickly and easily, because it was highly compatible with stimulus orientation, because it was similar to the joystick response performed in the Cavanagh et al. (1990) study, and because it was the same as the early portion of the action task, as discussed subsequently.

In the action task, participants were asked to indicate target orientation by lifting the block up to the screen and pressing it against the target, matching the position and orientation of the block’s end to the position and orientation of the target (see Figure 2). To perform this task, participants always exited the response box through the opening corresponding to the dominant hand. We designed the action task this way because we did not want to force participants to choose between exiting one opening or the other depending on the target stimulus. This would have transformed the action task into a recognition task.

We recognized, however, that a difficulty was created by forcing an immediate binary choice in the recognition task but not in the action task. The difficulty was that participants in the action task had the opportunity to defer target identification until after response initiation, whereas participants in the recognition task could not do so. We took two precautions to minimize this potential problem.

First, we framed the instructions in a special way. In the action task, we told participants not to make corrective submovements as they approached the screen, but instead to complete the originally planned movement as quickly as possible, note any error that was made, and try to be more accurate next time. Research on visual–manual aiming has shown that corrective submovements are likely to occur at the end of movement and to increase total movement time (see Meyer, Smith, Kornblum, Abrams, & Wright, 1990, for a review). Thus, urging participants to move as quickly as possible without making corrective movements was expected to promote early target commitments. Indeed, as shown in a study requiring rapid
placements of a handheld object up against a vertical surface, participants chose final hand positions even before they began their movements (Rosenbaum, Vaughan, Barnes, & Jorgensen, 1992). This early preplanning of final positions is often observed during manual positioning performance (Rosenbaum, Meulenbroek, Vaughan, & Jansen, 2001). The second precaution we took was a planned analysis of participants’ movements. We planned to analyze the orientation of the handheld block during its exit from the response box to determine when hand orientation significantly differed for left- and right-oblique targets. Underlying this analysis was the assumption that distinct hand orientations during movement reflect registration of the target’s orientation. The earlier the orientation of the hand differed for targets oriented to the left or right, the earlier one could be certain that the target had been found.

Design

A 2 (task type: recognition vs. action) × 2 (display type: color vs. binocular disparity) × 2 (target orientation: left oblique vs. right oblique) × 2 (display size: 2 vs. 6 items) mixed-factorial design was used. Task type was manipulated between participants, whereas all other variables were manipulated within participants. Each participant completed one 108-trial color block and one 108-trial binocular disparity block. The order of these blocks was counterbalanced across participants. Within each block, target orientation and display size were mixed pseudo-randomly so that an equal number of left and right targets and an equal number of two- and six-item displays were presented. A 10-trial practice session preceded each block. Blocks were divided into three 36-trial segments, and participants were permitted to rest for as long as they wished between segments. This design yielded 27 observations per condition and participant.

Procedure

Each participant completed a single 90-min session. Before beginning the visual search blocks, the participant completed two short preliminary tests, one that screened for stereo blindness and a second that screened for red–green color blindness and also determined the participant’s red–green equiluminance point. A description of each test follows.

Stereopsis screening pretest. Each participant performed a short test of stereovision. Displays that contained adjacent red and blue rectangles (each rectangle subtended approximately 1.0° × 0.9°) were presented. These rectangles were presented with 0.2° crossed or uncrossed disparity such that if the participant was not stereo blind, one rectangle appeared to be on a nearer depth plane than the other. The participant pressed a keyboard letter to report whether the left or the right rectangle appeared on the nearer depth plane (a or a /keypress, respectively). Two blocks of 48 trials were administered, one practice block and one test block. Participants whose accuracy fell below 70% on the test block were excluded from the study. There were 3 such participants.

Red–green equiluminance pretest. Individual red–green perceptual equiluminance points were determined with a minimum-rate-of-perceived-flicker test (Cavanagh, MacLeod, & Anstis, 1987). This test was always administered just before the block of color trials. Each participant wore the shuttered glasses during this test, but the displays were created so that there was no disparity between the left- and right-eye views. The participant viewed six rectangular bars (1.0° × 3.0°) on a black background, each occupying one of six fixed locations in the same circular arrangement used for the visual search displays. The bars were alternately filled with red or green color at a rate of approximately 30 Hz. Whereas the luminance of the green bars never changed, the luminance of the red bars was adjusted until a minimum rate of flicker was observed. The participant made adjustments by pressing the up and down arrow keys. These adjustment trials were untimed and were repeated until the participant achieved three consecutive trials with a red–green–blue (RGB) palette red value spread of no greater than two units. The RGB palette red value closest to the mean of these three red values was designated the red–green equiluminant point for that participant for the remainder of the experiment.

Visual search test. The events of the visual search trials are shown in Figure 3. A trial was initiated when the participant depressed the response box home switch. The computer then presented a gray random-dot luminance noise mask identical in appearance to the background rectangle of the binocular disparity displays. This mask remained on the screen for 3,000 ms and was followed by the presentation of a fixation circle in the center of the display. On color trials, the mask was replaced with the green background of the color displays and a red fixation circle. On binocular disparity trials, the fixation circle was defined by depth only. After 1,000 ms, the target and distractor(s) appeared. The full display remained on the screen for the next 3,000 ms and then was replaced by the gray random-dot mask, which remained on the screen until the next trial was initiated (see Figure 3). The random-dot mask was presented to impede adaptation to the color stimuli, thereby reducing fluctuations in the individual’s red–green equiluminance point. Participants could see whether or not they made the correct response on each trial, but they were not given explicit response time feedback.

Data Analysis

Recognition task performance was assessed by three dependent measures. One was choice RT, defined as the time from the appearance of the target and distractor(s) to the release of the home switch. The second measure was departure time (DT), defined as the time for the handheld block to travel from the home switch past the photocell that registered the response choice. The third recognition measure was choice error rate, defined as the percentage of incorrect target orientation choices (i.e., the percentage of trials in which the handheld block went out of the left exit for a right-oblique target or went out of the right exit for a left-oblique target).

Action task performance was measured by the same RT and DT measures as well as by three additional measures: movement time (MT), hand orientation accuracy at movement end, and hand orientation over time. A movement velocity threshold was used to determine movement initiation and termination. Movement initiation time was defined as the moment when movement velocity first surpassed 0.05 m/s minus two time samples (40 ms). Movement termination was defined as the first time at which movement velocity subsequently fell below 0.05 m/s plus two time samples (40 ms). The difference between these two times yielded the measured MT for each trial (see Figure 4). RT termination did not always coincide with MT initiation, because the participant’s hand speed sometimes reached criterion for MT initiation before the home switch was fully released (see Figure 3).

Hand orientation was measured by calculating the angle of the handheld block. The arctangent of the line subtended by the block’s two IREDs with respect to horizontal was calculated at each point in time between movement initiation and movement end. With this method, if block orientation was perfectly matched with the target, its angles at movement end were 60° for the right-oblique target and 120° for the left-oblique target. Orientation error was defined as the absolute value of the difference between the measured hand orientation at the last frame of movement and the respective perfect match.

Results and Discussion

Pretests

The average score on the stereo-screening test for the 28 participants who participated was 95.7% ± 6.5% (range = 80%–100%) correct. With respect to red–green equiluminance, the background green (RGB values = 0.52, 1.60, and 0.52 cd/m²) was, on average, matched with this perceptually equiluminant red
These red values ranged from low red (RGB values = 1.32, 0.52, and 0.52 cd/m²) to high red (RGB values = 1.71, 0.52, and 0.52 cd/m²).

**RT**

The RT results are summarized in Figure 5. RTs were submitted to a Task Type (recognition vs. action) × Display Type (color vs. binocular disparity) × Orientation (left vs. right) × Set Size (2 vs. 6 items) mixed-factorial analysis of variance (ANOVA). This analysis revealed a three-way interaction among task type, display type, and set size, $F(1, 26) = 12.33, MSE = 783.39, p < .01$. Other sources of variance were also significant: task type, $F(1, 26) = 11.48, MSE = 35,593.12, p < .01$; display type, $F(1, 26) = 49.79, MSE = 7,186.04, p < .01$; set size, $F(1, 26) = 30.00, MSE = 843.93, p < .01$; Task Type × Set Size, $F(1, 26) = 14.44, MSE = 843.93, p < .01$; and Display Type × Set Size, $F(1, 26) = 28.49, MSE = 783.39, p < .01$. The interaction between task type and display type was not significant, $F(1, 26) = 2.34$.

The three-way interaction was investigated further by separating the analysis by task type. Two separate 2 (display type) × 2 (set size) ANOVAs were conducted for recognition and action tasks. For recognition tasks, the interaction between display type and set size was significant, $F(1, 26) = 4.29, MSE = 783.39, p < .05$. For action tasks, the interaction between display type and set size was not significant, $F(1, 26) = 1.34, MSE = 783.39, p > .05$. These results suggest that display type and set size have different effects on RTs depending on whether the task is a recognition or action task.
size) repeated measures ANOVAs were performed, one for each task type. This procedure revealed an interaction between display type and set size for the recognition task, $F(1, 13) = 27.63, MSE = 1,110.01, p < .01$, replicating the findings of Cavanagh et al. (1990). In contrast, this interaction was not reliable for the action task, $F(1, 13) = 2.86, ns$.

The foregoing analyses reveal that participants who performed the recognition task responded more efficiently to displays defined by color than to displays defined by binocular disparity, wherein efficiency is defined by the slope of the function relating RT to set size. In contrast, participants who performed the action task processed the color and the disparity displays equally efficiently. Whereas 11 of the 14 participants who performed the recognition task processed color displays more efficiently than disparity displays, only 3 of the 14 participants who performed the action task conformed to this pattern. Planned comparisons (two-tailed $t$ tests) confirmed that when recognition task search slopes were compared with zero, the disparity search slope was significantly greater than zero, $t(13) = 7.35, p < .01$, whereas the color search slope was not, $t(13) = 0.48, ns$. In contrast, for the action task, neither the

![Figure 4](image-url) Kinematic profile from a single randomly chosen action trial. The solid circles represent the unfiltered velocity profile. A velocity cutoff of 0.05 m/s minus two time samples was used to determine the start of movement, and a velocity cutoff of 0.05 m/s plus two time samples was used to determine the end of movement. The vertical black lines reflect start and end points for this trial. The open squares illustrate the hand orientation profile over time. In this trial, the hand orientation at the end of the movement was 55.9°. This response was to a right-oblique target for which the perfect final hand orientation was 60°.

![Figure 5](image-url) Mean reaction times as a function of set size for color and binocular disparity displays. Error bars represent within-subject 95% confidence intervals for comparing display type within each task type. During the recognition task (left), color displays were processed significantly more efficiently than binocular disparity displays. In the action task (right), binocular disparity displays were processed no less efficiently than color displays.
disparity, t(13) = 1.94, ns, nor the color, t(13) = −0.08, ns, search slope was significantly different from zero.

During the recognition task, target location was an irrelevant feature that may have affected the time to report target identity, whereas, during the action task, both target location and orientation were relevant to the appropriate response. This situation left open the possibility that the observed search efficiency differences between recognition and action were driven by stimulus–response compatibility effects in the recognition task (Kornblum & Lee, 1995; Simon, 1990). To address this possibility, we coded target location as appearing either to the right or left of fixation, and we analyzed RT separately for trials in which target location and orientation were compatible (e.g., a left-oblique target that appeared left of fixation) and for trials in which target location and orientation were incompatible (e.g., a left-oblique target that appeared right of fixation). There was a compatibility benefit of 35 ms for the recognition task (Kornblum & Lee, 1995; Simon, 1990). To address this possibility, we coded target location as appearing either to the right or left of fixation, and we analyzed RT separately for trials in which target location and orientation were compatible (e.g., a left-oblique target that appeared left of fixation) and for trials in which target location and orientation were incompatible (e.g., a left-oblique target that appeared right of fixation). There was a compatibility benefit of 35 ms for the recognition task, F(1, 13) = 17.23, MSE = 11,898.57, p = .001. However, there was no interaction between compatibility and pattern of search efficiency: Color displays were processed efficiently during both compatible (3.5 ms/item) and incompatible (∼2.5 ms/item) trials, whereas binocular disparity displays were processed inefficiently during both compatible (23.7 ms/item) and incompatible (25.3 ms/item) trials. In contrast, we did not find a compatibility benefit for the action task (∼2 ms), F(1, 13) = 0.14, ns, nor were there any interactions involving compatibility. During action, color displays were processed efficiently during both compatible (∼1.75 ms/item) and incompatible (0.2 ms/item) trials. Similarly, binocular disparity displays were processed efficiently during both compatible (6.5 ms/item) and incompatible (6.5 ms/item) trials. Hence, although recognition was affected by target identity–location compatibility, this compatibility effect does not explain why color displays were processed more efficiently than disparity displays during recognition.

DT

DT data are summarized in Figure 6. Participant means were analyzed in a 2 (task type: recognition vs. action) × 2 (display type: color vs. binocular disparity) × 2 (orientation: left vs. right) × 2 (set size: 2 vs. 6) mixed-factorial ANOVA that revealed no significant sources of variance. Participants required an average of 151 ± 11.5 ms to move the bar from the home switch past the photocell switch.

Errors

Choice errors on the recognition task. Choice error rates for the recognition task appear in Table 1. Error rates were very low, reflecting the ease of the task. The arcsine-transformed error rates were submitted to a Display Type (color vs. binocular disparity) × Orientation (left vs. right) × Set Size (2 vs. 6) repeated measures ANOVA. No significant sources of variance were found.

Hand orientation errors on the action task. A summary of the hand orientation error data is presented in Figure 7. A Display Type (color vs. binocular disparity) × Orientation (left vs. right) × Set Size (2 vs. 6) repeated measures ANOVA for these data yielded no significant main effects or interactions.

Action-Task MT

A summary of the MT data appears in Figure 8. MTs were submitted to a 2 (display type: color vs. binocular disparity) × 2 (target orientation: left vs. right) × 2 (set size: 2 vs. 6) repeated measures ANOVA. There was no effect of display type, F(1, 13) = 0.10, ns. There was a small but significant increase in MT as a function of set size (2.2 ms/item), F(1, 13) = 8.87, MSE = 126.34, p < .05, but this increase was not reliably different across the two display types, F(1, 13) = 1.27, ns.

Further Comparisons of Performance Across the Two Tasks

As described in the Method section, comparing performance across the action and recognition tasks is complicated by potentially different response initiation criteria. Recognition-task participants were required to find the target before response initiation, because the task demands a short movement in the target-compatible direction. In contrast, action-task participants were not required to find the target before response initiation because the task demanded that hand orientation match target orientation only.

Figure 6. Mean departure times (DTs) as a function of task type, display type, and set size. DT reflects the time required to move from the home switch to the photocell switch. As such, it reflects the time participants needed to negotiate the response box. Error bars represent within-subject confidence intervals for comparing display types within each task type. The left and right graphs show DT for the recognition and action tasks, respectively. There were no significant sources of variance.
at movement end. It is possible, therefore, that participants in the action task initiated their responses immediately after detecting the onset of the display and then deferred completion of target perception until the movement was under way. This deferment strategy would have been an especially plausible possibility given that a target was presented and a response was required on every trial, meaning that there may have been nothing to discourage early response initiation based on display onset only. We assessed this possibility with the following analysis.

**Hand orientation over time.** Our test of the deferment hypothesis assessed participants’ knowledge of target orientation by studying their hand orientation at and shortly after movement initiation. The earliest moment at which hand orientation reliably indicates target orientation provides an estimate of (a) the latest time by which the target was found and (b) the time by which the target’s identity (orientation) was processed sufficiently to influence ongoing behavior correctly. To determine the time at which individuals’ hand paths discriminated left- from right-oblique targets, we relied on a method developed previously for detecting the onset of the lateralized readiness potential, a motor-related electroencephalogram slow-wave component (de Jong, Wierda, Mulder, & Mulder, 1988). We used this method to compare hand-orientation profiles for movements to left-oblique targets with hand-orientation profiles for movements to right-oblique targets. A Wilcoxon rank sum (W) test was performed within each participant, within each display-type condition, within each set-size condition, and at each of the first 200 time samples (within the first 400 ms) of the movement to determine whether the distribution of hand angles associated with responding to left-oblique targets was different from the distribution of hand angles associated with responding to right-oblique targets (see Figure 9). The latency at which the participant could distinguish a left-oblique target from a right-oblique target was defined as the first time sample at which the one-tailed W statistic was significantly different from zero ($\alpha = .01$) and continued to be significantly different from zero for the remainder of the time series (de Jong et al., 1988).

Target orientation discrimination latencies are summarized in Figure 10. These data were submitted to a Display Type (color vs. disparity) × Set Size (2 vs. 6) repeated measures ANOVA that yielded a main effect for set size such that MT was longer when six items were presented than when two items were presented.

---

**Table 1**

*Response Choice Errors (Mean Percentages Incorrect, With Standard Errors of the Mean) for the Recognition Task*

<table>
<thead>
<tr>
<th>Display type</th>
<th>Set size</th>
<th>Color</th>
<th>Binocular disparity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 items</td>
<td>0.90 (0.30)</td>
<td>0.52 (0.20)</td>
</tr>
<tr>
<td></td>
<td>6 items</td>
<td>1.03 (0.50)</td>
<td>1.01 (0.40)</td>
</tr>
</tbody>
</table>

---

**Figure 7.** Action-task hand orientation errors at movement end as a function of display type and set size. Error bars represent within-subject confidence intervals for comparing display types. There were no significant sources of variance.

**Figure 8.** Movement times (MTs) as a function of display type and set size. Error bars represent within-subject confidence intervals for comparing display types. There was a significant main effect for set size such that MT was longer when six items were presented than when two items were presented.

**Figure 9.** A single participant’s action-task hand orientation profiles over time. Each curve reflects a single trial performed by this participant while viewing color displays with two items. At each of the first 200 movement time samples, a Wilcoxon rank sum (W) statistic was calculated. The sample at which the W statistic was significant and remained significant for subsequent samples defined the time at which the participant’s hand orientation correctly reflected target orientation. W statistics were calculated within each participant, for each Display Type × Set Size combination.
ded in a two-item display (115.7 ms) than when it was embedded in a six-item display (93.6 ms). There were no other significant main effects or interactions, suggesting that display type was not a critical determinant of processing time once the action response was initiated.

The target orientation discrimination latencies showed that participants’ hand orientations reliably reflected the presented target orientation within 104 ms of movement onset, on average. Although it is conceivable that the action-task participants initiated their responses before finding the target and then continued to search for it during the first 100 ms of MT, this interpretation is highly unlikely in that it usually requires 100 ms or more for visually presented information to affect ongoing arm and hand movements (Castiello, Jeannerod, & Prablanc, 1991; Georgopoulos, Kalaska, & Massey, 1991; Paulignan, MacKenzie, Marteniuk, & Jeannerod, 1991).

Total time to distinguish target orientation. The preceding analysis of hand orientation over time showed that action-task participants made responses that distinguished target orientation by 104 ms after movement onset, on average. Recognition-task participants, in contrast, were required to make responses that distinguished target orientation immediately on movement onset. To ensure that our measure of visual search time captured the same perceptual events for recognition and action, we reanalyzed our visual search data using an estimate of the total time to distinguish target orientation (TTD). For the recognition task, TTD was equal to RT. For the action task, TTD was equated with the sum of RT and the target orientation discrimination latency at which the hand began to rotate in the correct direction.

The results of the TTD analysis are presented in Figure 11. The TTDs were submitted to a 2 (task type: recognition vs. action) × 2 (display type: color vs. binocular disparity) × 2 (set size: 2 vs. 6) mixed-factorial ANOVA. The analysis revealed an interaction of task, display type, and set size, $F(1, 26) = 11.26, MSE = 1,469.83, p < .01$. Other sources of variance were also significant: display type, $F(1, 26) = 29.80, MSE = 10,698.90, p < .01$; set size, $F(1, 26) = 5.75, MSE = 2,120.70, p < .05$; Task Type × Set Size, $F(1, 26) = 11.94, MSE = 2,120.70, p < .01$; and Display Type × Set Size, $F(1, 26) = 9.65, MSE = 1,469.83, p < .01$. Neither the main effect for task, $F(1, 26) = 0.11$, nor the interaction between task type and display type, $F(1, 26) = 2.49$, was significant.

The three-way interaction was investigated further by separating the analysis by task type. Two separate 2 (display type) × 2 (set size) repeated measures ANOVAs were performed, one analysis for each task type. This procedure revealed an interaction between display type and set size for the recognition task, $F(1, 13) = 27.63, MSE = 1,110.01, p < .01$, but not for the action task, $F(1, 13) =$.
Planned comparisons (two-tailed $t$ tests) confirmed this pattern. When recognition task search slopes were compared with zero, $t(13) = 7.35, p < .01$, whereas the color search slope was not, $t(13) = 0.48, ns$. In contrast, for the action task, neither the disparity, $t(13) = −.76, ns$, nor the color, $t(13) = −0.26, ns$, search slope was significantly different from zero.

The foregoing analysis of TTD provided two important results. First, the effect of display type on visual search efficiency during action was the same both for TTD and RT, indicating that, during action, color and binocular disparity displays were processed equally efficiently. Second, when TTD was used as the dependent measure instead of RT, there was no longer a main effect of task type, indicating that the total time to discriminate left- from right-oblique targets was the same for recognition and action.

**General Discussion**

The goal of this study was to test whether visual information processing for action is different from visual information processing for recognition in neurologically normal individuals. Our method focused on the relative processing efficiency of displays defined by color versus binocular disparity during visual search tasks that involved recognition or action responses. The most important result was that when participants physically interacted with targets, orientation was extracted efficiently from both color- and disparity-defined displays. Conversely, when participants indicated target recognition, orientation was extracted more efficiently from color-defined displays than from disparity-defined displays. The difference in search efficiency between the action and recognition tasks is consistent with the view that visual processing differs for action and recognition.

The test of visual information processing for action and recognition used here has several advantages over previous tests designed to assess the how and what hypothesis. First, by comparing responses to color- and disparity-defined stimuli within rather than between tasks, we controlled for the online-correction advantage that experimental action tasks often have over recognition tasks. Other studies have tried to control for this task difference by limiting vision of the display and hand during both recognition and action tasks, but that manipulation forces participants to rely on memory of the stimuli during performance. This reliance on memory may reduce the difference between perception for action and perception for recognition, rendering measures of these perceptual processes indistinguishable. Thus, a second advantage of the test used here is that vision of the hand and of the display was continually available throughout both tasks. Moreover, the instructions for both tasks stressed fast and accurate responding, which made response latency the main dependent variable in each task. Using a latency measure allowed us to focus on the visual processing that must be performed to find the target for recognition or for action.

*The Link Between Differential Visual Processing Across Tasks and Properties of Cells in the Dorsal and Ventral Streams*

As noted in the introduction, the pattern of differential processing across the recognition and action tasks observed in this study can be related to known differences in single-cell selectivity in the dorsal and ventral streams. This link provided one of the motivations for the present study, and the results conformed well to expectations.

That color displays were processed more efficiently than binocular disparity displays during the recognition task replicates the results of Cavanagh et al. (1990). As noted by those authors, such results suggest that whereas the binocular disparity channel can process orientation, it is less finely tuned to orientation than is the color channel. This difference between binocular disparity and color processing may be attributable to two factors in combination: (a) the greater proportion of cells in the ventral stream with a preference for color-defined stimuli relative to cells with a preference for binocular disparity-defined stimuli and (b) the greater engagement of ventral stream processing during recognition tasks than during action tasks. To illustrate how these factors may combine to influence search efficiency, we present the following analogy to retinal resolution. Spatial resolution of the retinal image is greater at the fovea than at peripheral locations because there is a higher concentration of photoreceptors at the fovea than in the periphery. By analogy, orientation resolution may be greater for color-defined stimuli than for disparity-defined stimuli in recognition tasks because there is a higher concentration of color-sensitive cells than binocular disparity-sensitive cells in the ventral stream, and recognition tasks depend selectively on ventral stream processing. Consistent with this interpretation, a pilot study involving our recognition task (a standard visual search task) indicated that the resolution of orientation perception was higher for color-defined stimuli than for disparity-defined stimuli.

Small differences between target and distractor angle ($10^\circ$) were still processed efficiently (7.5 ms/item) when the display was defined by color but were processed substantially less efficiently (128.5 ms/item) when the display was defined by binocular disparity.

By applying analogous logic, the elimination of the processing advantage for color over binocular disparity in the action task can be attributed to two factors in combination: (a) the greater proportion of cells in the dorsal stream with a preference for binocular disparity-defined stimuli than for color-defined stimuli and (b) the greater relative engagement of dorsal stream processing during action tasks than during recognition tasks.

By these lines of argument, the difference in the relative efficiency with which color and binocular disparity were processed for action and for recognition can be linked to known differences in cell selectivity across the dorsal and ventral neural streams. Cavanagh et al. (1990) speculated that visual search mechanisms could have access to stimulus-specific representations or that these mechanisms could have access only to a common representation of the display for which the contributions from each stimulus have already been integrated. Although the present experiment was not designed to distinguish between these alternatives, our results suggest that the construction of stimulus-specific representations is responsive to task conditions and that visual search mechanisms can access these stimulus- and task-specific representations. On the whole, this evidence is consistent with Milner and Goodale’s (1995) distinction between visual processing for action and visual processing for recognition.
**Possible Alternative Explanations**

Although the results obtained here are consistent with Milner and Goodale’s (1995) how and what hypothesis, it is important to consider other possible accounts of the results. Here we consider three such possibilities: (a) the premotor theory of attention, (b) a common representation account, and (c) an account based on the where and what distinction.

**Premotor theory of attention.** An alternative framework for the interpretation of our results, the premotor theory of attention (Rizzolatti et al., 1994), also proposes multiple visual representations, not only for recognition and action but also for different types of action. The premotor theory suggests that spatial attention mechanisms, like those responsible for visual search, are tied to spatial maps that are functionally related to the intended action system. According to the theory, the intention to perform an action with a particular effector facilitates perception of the target’s location in the appropriate effector-centered spatial map. For example, if the goal is to look at a target, the target’s location is facilitated within an eye-centered spatial map, whereas if the goal is to reach for a target, the target’s location is facilitated within an arm-centered map.

How can we evaluate the relevance of the premotor theory to our data? Because our recognition response was performed with the wrist primarily, whereas our action response was performed with the shoulder, elbow, and wrist, these different action requirements may have invoked different effector-centered spatial maps. Thus, the premotor theory might explain our results by proposing that the recognition and action tasks were served by different visuospatial representations. For the premotor theory to fully explain our data, however, one would need to assume that the wrist-centered spatial map and the arm-centered spatial map are differentially sensitive to color and binocular-disparity stimuli. At this point in time, we cannot rule out this possibility.

**A common representation.** An explanation that requires no appeal to dual representations also appears to account for our results; on closer inspection, however, it does not do so completely. Suppose that a single representation of the scene receives information relevant to both action and recognition and that this representation is updated by continued sampling over time (Di Lollo, Enns, & Rensink, 2000; Wolfe, 1994). Because action requires continually updated, viewer-centered information about the state of the limb and target, whereas recognition requires storage of object characteristics and comparisons with existing object representations, action might be triggered by an earlier version of this common representation than is recognition. This explanation appears to account both for the relative search efficiency differences between recognition and action and for the observation that action-task responses were initiated earlier than recognition-task responses.

On closer examination, however, our results do not support this single representation account. During the action task, the target was always extracted efficiently from the display, and this was true regardless of whether target extraction time was equated with RT or with TTD. In contrast, during the recognition task, the target was not always extracted efficiently (especially when it was defined by binocular disparity), even though the recognition-task response was initiated later than the action-task response. If the recognition response was initiated later than the action response because visual search mechanisms acted on a later, improved representation of the display, one would not expect orientation to be extracted less efficiently at this later time. Early, efficient display processing for action coupled with late, inefficient display processing for recognition cannot be explained easily by a common representation that improves over time.

A variant of the common representation model proposes that whereas movement planning and recognition rely on a single representation, perception during action relies on a separate representation of the visual scene from the one used for recognition (Glover & Dixon, 2001). In support of this possibility, Glover and Dixon showed that when participants reached for and grasped cylindrical objects presented on a tilted vertical grating background, participants’ perceptual judgments of the cylinder’s orientation and their early hand orientation were susceptible to a tilt illusion. In contrast, as prehension unfolded, the effect of the illusion on hand orientation weakened over time until it was negligible by movement end. On the basis of this result, Glover and Dixon argued that movement planning is based on a visual representation that also serves perceptual judgment and recognition, but online movement control is based on a separate representation that is less susceptible to context effects. Consistent with this planning–control distinction is evidence that neglect patients’ grasping responses are less vulnerable to unilateral neglect when they are given online movement feedback than when they were denied feedback (Edwards & Humphreys, 1999). With respect to our experiment, however, the planning–control model predicts that the pattern of visual search efficiency should be very similar for both recognition and action. Our results do not fit such a prediction, however, and so they argue against this alternative explanation.

**Where and what.** In the action task participants’ responses indicated both target location and identity, whereas in the recognition task their responses indicated target identity only. This difference leaves open the possibility that the action task called on a dorsal stream location-finding mechanism and that this may have benefited perception of binocular disparity information. This explanation is consistent with the classic where and what hypothesis, which proposes that the dorsal stream is a specialized module for perceiving location, whereas the ventral stream is specialized for perceiving objects (Ungerleider & Haxby, 1994; Ungerleider & Mishkin, 1982).

Two lines of evidence cast doubt on the where and what account, however. First, evidence from the visual search literature indicates that perception of target identity requires target localization (Johnston & Pashler, 1990; Wolfe, 1994). This evidence suggests that the target’s distinguishing feature stimulates a unique location on a map of activations and that attention is drawn to this location to determine target identity (Wolfe, 1994). Second, as reported in the Results and Discussion section, location information influenced recognition-task RT in the form of a Simon effect (Kornblum & Lee, 1995; Simon, 1990). Therefore, we are confident that perception for recognition and perception for action required that both target location and identity be determined. Because the where and what explanation of our results hinges on the assumption that perception of target location was required only during action, it cannot explain our data.
Why Was Color Always Processed Efficiently?

A final question is why color was always processed efficiently in the present experiment. If the relative engagement of the ventral and dorsal streams by action tasks was a complete reversal of that by recognition tasks, then we would predict that color-defined stimuli would be processed less efficiently than disparity-defined stimuli during the action task. Instead, binocular disparity and color displays were processed equally efficiently during action. This observation suggests that the relative engagement of the ventral stream and dorsal stream for action may not be a complete reversal of that for recognition in our normal participants.

It is possible that color is important to both recognition and action because it is a first-order stimulus (Cavanagh, 1988, 1989; Cavanagh et al., 1990). First-order stimuli (those defined by color or luminance) form the basis for higher order stimuli (those defined by texture, motion, and binocular disparity) and therefore may be shared more freely between the perception-for-recognition and perception-for-action systems. For example, color discontinuities can provide more reliable edge information than do luminance discontinuities, because luminance can signal both object and nonobject (shadow) edges (Cavanagh & Leclerc, 1989). Evidence suggests that shapes defined by equiluminant color can be used to guide prehension as effectively as shapes defined by luminance (Franz, Gegenfurtner, Bulthoff, & Fahey, 1998).

Not only were color displays always processed efficiently in the present study, but participants also responded to color displays sooner, on average, than they responded to binocular disparity displays. This color advantage was observed during both recognition and action tasks. Cavanagh et al. (1990) also found this difference and attributed it to the additional time required to resolve depth. This interpretation fits with the results of the present experiment as well. Participants always needed more time to resolve depth from binocular disparity; once they had done so, however, they were able to extract orientation efficiently for use in the action task but not the recognition task.

Conclusion

In conclusion, the present results are consistent with the notion that visual representation for action is different from visual representation for recognition in neurologically normal participants. Our results do not support, or at least do not demand explanation in terms of, the premotor theory of attention (Rizzolatti et al., 1994), a common representation for both action and recognition that continues to improve over time, or the distinction between a where and a what system (Ungerleider & Haxby, 1994; Ungerleider & Mishkin, 1982). Instead, the efficiency with which people process stimuli seems to depend on whether they are looking for the sake of acting or looking for the sake of recognizing and judging.

References


Received November 15, 2000
Revision received April 11, 2002
Accepted April 16, 2002