THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 1988 *40A* (2) 201–237

# Features and Objects: The Fourteenth Bartlett Memorial Lecture

Anne Treisman

*University of California, Berkeley, U.S.A.*

Perception seems so effortless and instantaneous, however rich and varied the visual scene may be, that it is hard to imagine the complexity of the analysis on which our experience depends. I have been interested in finding out which operations do in fact tax the system most, and which appear to take place automatically. The idea that feature integration might pose a special problem for the perceptual system dates back at least to the 1960s. Neisser (1967), following Minsky (1961), claimed that "to deal with the whole visual input at once, and make discriminations based on any combinations of features in the field, would require too large a brain to be plausible." They suggested that the scene could first be articulated into parts, and a fixed set of pattern recognition procedures could then be applied repeatedly to each local region. In 1969 and 1973, I raised the possibility that whereas detection could be triggered by simple features, conscious awareness might depend on feature integration and that, with high attention load, errors of integration might be made (Harvey & Treisman, 1973; Treisman, 1969). Milner incorporated the same idea in his model of shape recognition (Milner, 1974). Garner's discoveries (1970, 1974) that many perceptual

Requests for reprints should be sent to Dr Anne Treisman, Department of Psychology, University of California, Berkeley, California, 94720, U.S.A.

The lecture was given on January 5th, 1987, at University College, London. Much of the research described in the lecture was supported by a grant from the Natural Sciences and Engineering Research Council of Canada and by a Fellowship from the Canadian Institute of Advanced Research. Some was supported by a grant from the U.S. Air Force Office of Scientific Research. I am grateful to these sources for making the research possible. I am also grateful to Marcia Grabowecky, Stephen Gormican, Deborah Butler, and Sharon Sato, who collaborated with me on some of the unpublished experiments, and to Daniel Kahneman and Glyn Humphreys for their comments and suggestions on an earlier draft.

Preparation of the paper was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under grant number, AFOSR 87–0125. The manuscript is submitted for publication with the understanding that the U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes, notwithstanding any copyright notation thereon.

dimensions are processed independently in ratings of similarity, in speeded classification tasks, and in absolute judgements made the question about how the dimensions are recombined to specify objects more cogent, as did the anatomical and physiological discoveries of many separate visual areas that appear to specialize in coding different properties (Cowey, 1979; Maunsell & Newsome, 1987; Van Essen & Maunsell, 1983; Zeki, 1981). In 1975 we began to collect data that confirmed these speculations. Attention did appear to be needed to ensure error-free feature integration.

## A Model for the Perception of Visual Objects

In order to provide a structure to hold together the various findings, I begin by outlining the model, shown in Figure 1 (Treisman, 1985; Treisman &

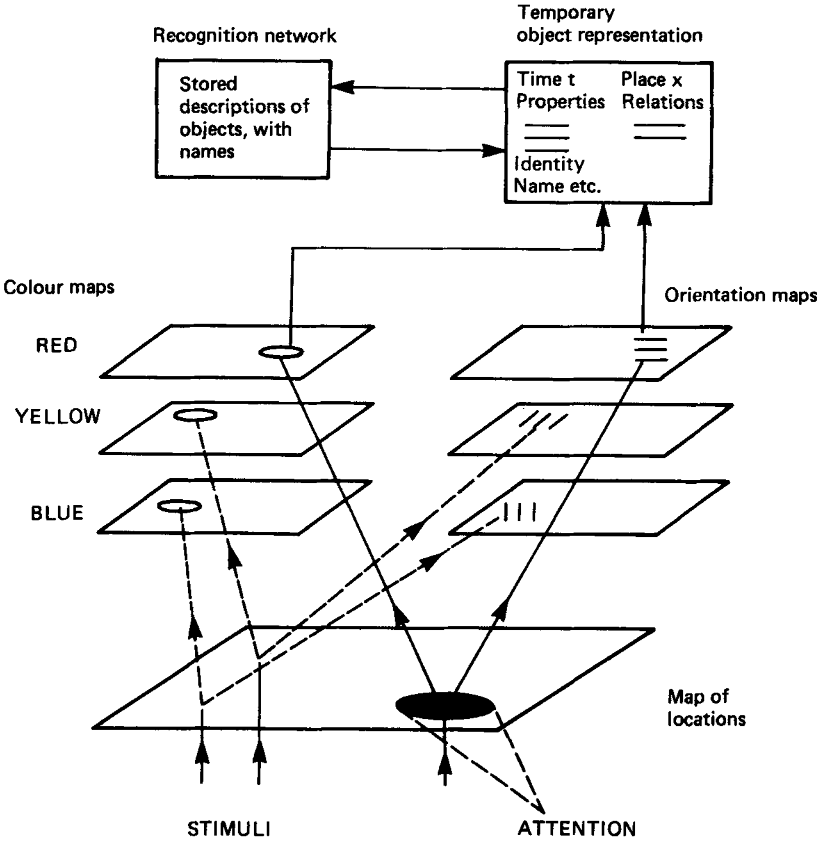


FIG. 1. General framework for perceptual processing in object perception.

Gormican, 1988; Treisman & Souther, 1985), which has developed out of a series of experiments. Think of it as a memory heuristic, a framework to give shape to the data currently available, rather than a fully-specified theory. It is certainly too simple and also certainly wrong in some respects.

The initial assumption is that different sensory features, such as colours, orientations, sizes, or directions of movement, are coded in specialized modules. (I return later to the question of how to decide what is and what is not a functional feature in the language of visual coding.) I further assume that these basic features are coded automatically, without focused attention, and spatially in parallel. Differences at this level of processing can mediate the segregation of figures and ground that sorts the visual scene into potential objects, ready for more detailed perceptual analysis. For example, it might collect the brown areas together, separating them from the green, so that we can subsequently identify a cow that is partly hidden by a bush.

Each module forms different feature maps for the different values on the dimension it codes—for example red, blue, and green within the colour module, vertical, diagonal, and horizontal within the orientation module. For dimensions like these, which vary continuously, the maps may also be functionally continuous, forming a kind of three-dimensional cube, although widely separated values would have effectively discrete representations. In search tasks, these separate maps allow the detection of targets with a unique sensory feature, simply from the presence of activity in the separate map for that feature.

When features must be located and conjoined to specify objects, attention is required. Attention selects within a "master map of locations" that shows *where* all feature boundaries are located, but not *which* features are located *where*. Thus it distinguishes "filled" from "empty" locations, where "filled" implies the presence of any discontinuity at the feature level. When attention is focused on a particular location in the master map, it allows automatic retrieval of whatever features are currently active in that location, through links to the corresponding locations in the different modular feature maps. My claim is that locations in the feature maps are made available to control responses and conscious experience only through their links to those locations in the master-map that are currently selected by the attentional "spotlight". Attention can be spread over a large or a small area; the narrower the focus, the more precisely located and accurately conjoined the features in that location will be. There is some empirical evidence (Posner, Snyder, & Davidson, 1980) suggesting that attention cannot be split between two spatially separate locations. However, some more recent results (discussed elsewhere in this paper) may modify this claim.

I have hedged my bets on where to put the master-map of locations by publishing two versions of the figure! In one of them, the location map receives the output of the feature modules (Treisman, 1986a) and in the other it is placed at an earlier stage of analysis (Treisman, 1985; Treisman & Gormican, 1988), as it is in Figure 1. Placing the master-map early implies that different dimensions are initially conjoined in a single representation before being separately analysed, dimension by dimension. Some recent research by Houck and Hoffman (1986), which I discuss later, has currently tipped the balance for me in favour of this version of the model. It is also consistent with physiological evidence that early coding by single units (for example in area VI) is selective for particular values (e.g., vertical or moving right), but combines particular values on each of several different dimensions (e.g., orientation and spatial frequency).

Given this evidence that many features are initially conjoined, we need some explanation of the need for the inferred separate analysis along different dimensions as a precursor to object recognition. Part of the answer may lie in the realization that the properties we conjoin to form objects should be real-world properties, after constancy mechanisms have operated, not properties of the retinal stimuli. The relevant conjunctions will generally characterize more complex and structured entities than the oriented bars or gratings that are apparently picked up in area V1. The early conjunctions can therefore not be directly interpreted in a form that is useful to the organism. However, we still need an explanation for why different properties should be separately analysed at an intermediate stage. Marr (1982), Cowey (1981), Ballard (1986) and Barlow (1986) have each suggested important advantages that might accrue from a specialized, modular analysis of different properties. For example, it may allow easy coding of relations within dimensions, without crosstalk from other dimensions; it may also be easier, in evolutionary terms, to develop a special-purpose module to perform a particular function, just as in computational models special subroutines are easier to debug if isolated from the main program.

The final level of perceptual coding shown in Figure 1 is one at which the different properties abstracted by specialized modules are recombined to allow the perception of objects, scenes, and events. I assume that conscious perception depends on temporary object representations in which the different features are collected from the dimensional modules and inter-related, then matched to stored descriptions in a long-term visual memory to allow recognition.

## Visual Search for Features and Conjunctions

Next I outline the evidence that led to these ideas, and describe new findings. Some fit the framework and others will lead to modifications. Our initial experiments showed that when subjects look for a target defined only by a conjunction of properties (e.g. a green "T" among green "X"s and brown "T"s), their search latencies increase linearly with the number of non-target or distractor items (Treisman & Gelade 1980; Treisman, Sykes, & Gelade, 1977). On the other hand, when two disjunctive targets are defined by separate unique features like a particular colour or the presence of curvature (e.g., a blue letter or an "S" among green "X"s and brown "T"s), detection times showed no systematic effect of the number of distractors. The contrasting results suggested that attention must be focused serially on each object in turn to verify how its properties are conjoined, but that distractors can be rejected in parallel, whenever the targets have unique features that none of them shares. The target with the unique feature will then normally "call attention" to its location (see discussion on pages 226–230).

We obtained the same patterns of search with local elements of shapes: for example search for an "R" among "P"s and "Q"s appeared to be serial, whereas search for an "R" among "P"s and "B"s was faster and the functions increased less steeply and less linearly with the number of distractors. Note that in terms of similarity, "R" is less discriminable from "B" than from "Q"; in a control experiment, it was found more slowly when it was embedded in "B"s alone than in "Q"s alone. We attributed the difficulty when "P"s were mixed with "Q"s rather than "B"s to the fact that "R" has a unique feature (the diagonal line) that distinguishes it from "P" and "B", whereas the same diagonal line is shared by the "Q" distractors.

My hypothesis is that "pop-out" for a target defined by a single distinctive feature is mediated by the unique activity it generates in the relevant feature map. If activity is detected there, the target must be present; if not, a negative response is made. Note that this criterion requires the relevant features to be unique to the target. There are two ways in which this criterion could be violated: first, the relevant features could be shared to differing degrees by the target and the distractors; second, the relevant features could be present in the distractors and absent from the target. We have recently shown that both these conditions give apparently serial search, even though no conjunction process is involved (Treisman, 1985; Treisman & Gormican, 1988; Treisman & Souther, 1985). For example, when both the target and the distractors are lines differing only in length (see Figure 2), search times increase with display size with a slope that depends on the discriminability of the target (Treisman & Gormican, 1988). The suggestion is that the more similar the activity generated in the relevant feature map by the target and the distractors, the longer a serial search will take.

In addition, even when the features are highly discriminable, a target that *lacks* the relevant feature appears to require serial search. Thus a target circle *without* an intersecting line among distractor circles that all have the lines gives latencies that increase linearly with display size, even though the reverse arrangement allows parallel detection, (i.e. a target circle *with* an intersecting line does pop out of a display of distractor circles without; see Figure 3 and Treisman & Souther, 1985). In sum, unless activity from a *unique* feature signals the *presence* of a target, attention seems to be focused serially on items or on groups of items. In each case, I suggest, attention is narrowed sufficiently for the target-induced activity to differ discriminably within the attended area from the activity generated by distractors alone (Treisman & Gormican, 1988). The more items that can be checked together without error within each "fixation" of attention, the faster the rate at which the display will be scanned.

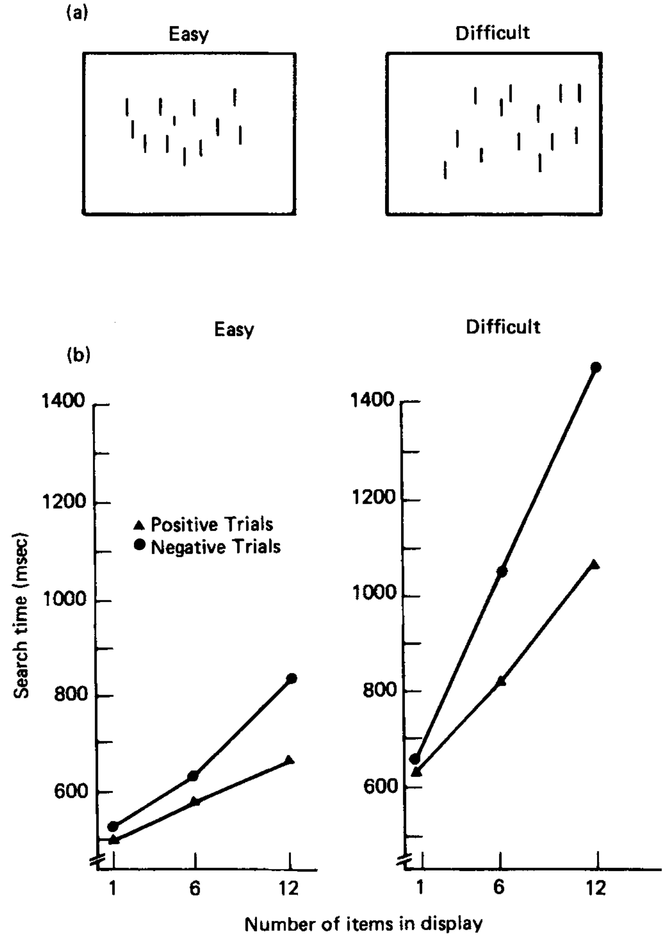


FIG. 2. Examples of displays and mean search times for a target line differing in length from the distractors.

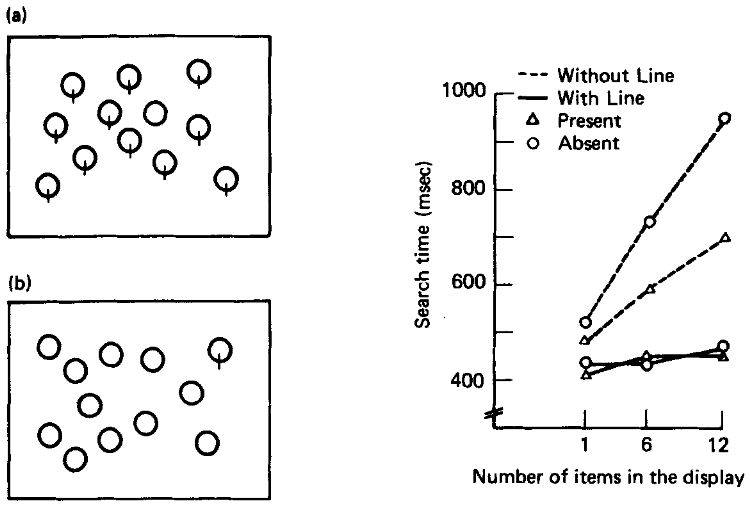


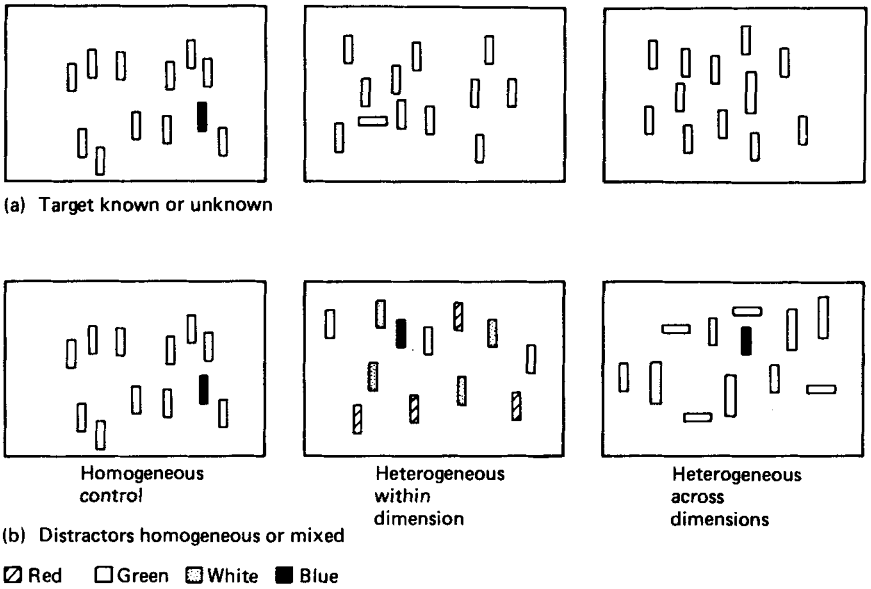
FIG. 3. Examples of displays and mean search times for a target circle with and without an intersecting line.

## Modularity in Feature Analysis

Can one search for several different feature targets at the same time? Our current research suggests that this is possible, but there is sometimes a cost. Figure 4a shows examples of displays with different targets. Subjects either knew in advance which target, if any, would be presented, or had to search for any of the three. If the disjunctive targets were all defined within the same dimension (a blue, red, or white bar among green bars; or a horizontal, left diagonal, or right diagonal bar among vertical bars) there was little increase in latency (19 msec) relative to search for a single known target (the blue bar in the colour condition, or the horizontal bar in the orientation condition). The first two graphs in Figure 5a show the results of eight subjects searching for a known and an unknown target differing from the distractors either in colour or in orientation. In each case, search also remains spatially parallel, in the sense that display size has no effect on latencies. However, if the disjunctive targets are defined by values on *different* dimensions, there is a significant increase in latency, (as shown in the third graph of Figure 5a). It took subjects an extra 90 msec to find a blue target among green vertical bars when it could instead have been horizontal or larger than the distractors, and an extra 91 msec to find a horizontal target when it could instead have been blue or large. Thus the "odd one out" pops out *within* a single, pre-specified dimensional module, but each different module may need to be separately checked to determine which of them contains it.

The opposite prediction holds when we make the distractor items heterogeneous. If different features are analysed by functionally autonomous modules, it should not matter how varied the distractor (non-target) items are, provided that they vary on irrelevant dimensions and differ by the same fixed amount from the target on the one relevant dimension. Subjects should simply check for activity signalling a contrasting item in the relevant target-defining module, and ignore the others. On the other hand, heterogeneity of the distractors *within* the relevant module might be expected to slow search, both because the distractors would contrast with each other as well as with the target, making it necessary to locate the specific map for the target within the relevant module, and because the more different maps are activated, the more similar to the target the nearer distractor value is likely to be.

The predictions were confirmed by the results of three search tasks, illustrated in Figure 4b, in which subjects looked for a fixed target value (blue in some blocks, horizontal in others) against a background of either homogeneous (green vertical) distractors, or randomly mixed distractors. The mixed distractors differed either on irrelevant dimensions or on the same dimension as the target. Figure 5b shows the search latencies for homo-


(a) Target known or unknown

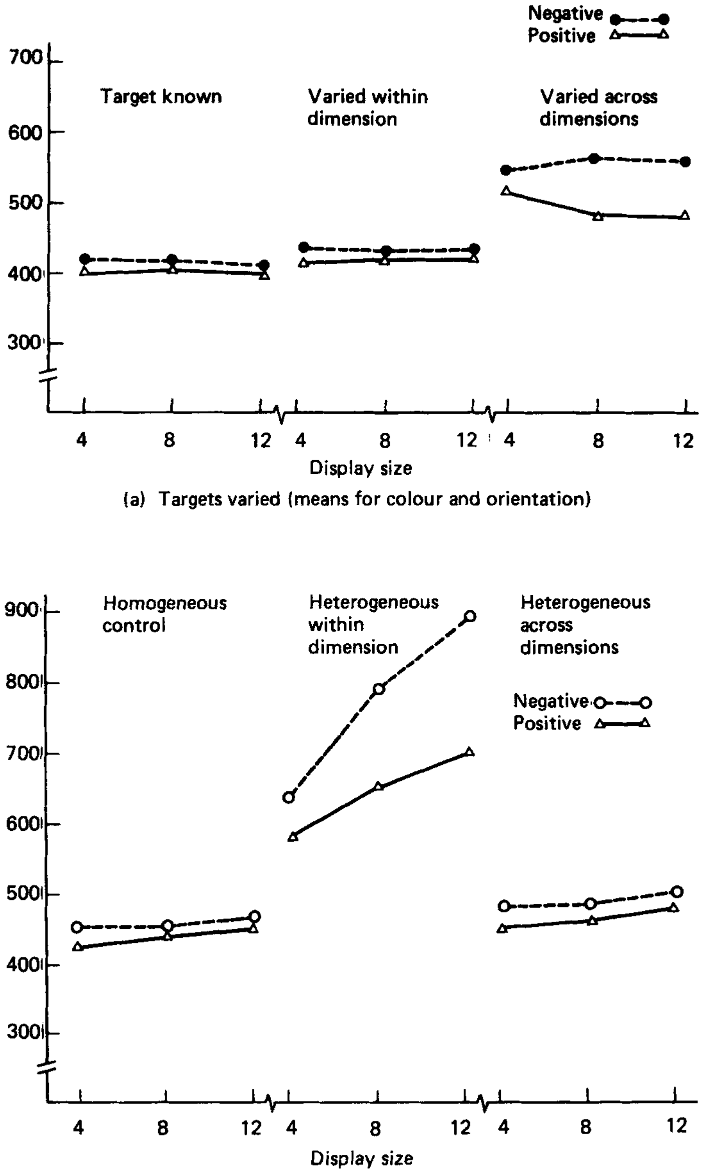
Homogeneous control

Heterogeneous within dimension

Heterogeneous across dimensions

(b) Distractors homogeneous or mixed

FIG. 4. Examples of displays testing effects of heterogeneity of targets and of distractors. (a) Targets defined on different dimensions (colour, orientation, size); distractors homogeneous. (b) Target known; distractors homogeneous or varied within the relevant dimension, or between dimensions.



(a) Targets varied (means for colour and orientation)

(b) Distractors varied (means for colour, orientation and size)

FIG. 5. (a) Search times for a single known target or for any of three targets. (b) Search times for targets defined by one unique feature among homogeneous or varied distractors.

geneous and heterogeneous displays. In the mixed, "across-dimensions" condition, the distractors were green bars in three orientations and three sizes for the blue target; vertical bars in three colours and three sizes for the horizontal target. Latencies were not significantly longer in these conditions than in the control conditions with homogeneous green vertical distractors. In the mixed, "within-dimension" condition, the distractors varied on the relevant, target-defining dimension (red, green, and white bars for the blue target; vertical, left diagonal, and right diagonal bars for the horizontal target). Search here was significantly slower for both colour and orientation targets, and was no longer spatially parallel for the orientation targets. Variations in the number of distractors generated slopes of 16 and 26 msec per item for target present and target absent displays, respectively. The results provide additional support for the idea of separate analysis by specialized modules when features are defined on different dimensions of variation.

## The Role of Attention in Feature Integration

I return now to the conjunction part of my story. I suggested that objects characterized by conjunctions of separable features are correctly perceived only through serial focusing of attention on each item in turn. We looked for more direct evidence that attention is involved, using a number of different converging operations to test the hypothesis.

*Pre-cueing a Spatial Location.* First, we explored the effect of pre-cueing the location of the target (Treisman, 1985). If attention is needed to detect conjunction targets, a valid precue should eliminate the serial checking phase. On the other hand, when the target is defined by a single feature, a cue to its location should have little effect; separate features can be detected in parallel anyway. We used displays like those in Figure 6a, containing objects that varied in shape, size, colour, and whether they were outline or filled. The target was defined either by a conjunction of properties, for example a large, brown, outline triangle, or by a single property like red (or large, or filled). We precued the location at which the target would occur, if it was present, by flashing a pointer to that location 100 msec before presenting the display. The precue was valid on 75% of the trials on which the target occurred; in other words it correctly predicted where the target would be. It was invalid on 25% of trials; in these cases the target occurred somewhere other than at the cued location. On invalid trials, attention would be directed to the wrong location rather than distributed across the whole display. An invalid cue might therefore give rise to costs rather than benefits relative to a condition with no cue (Posner & Snyder, 1975). On neutral trials, no advance information was given about the target location, although a temporal warning signal was given to equate the general level of preparation. We

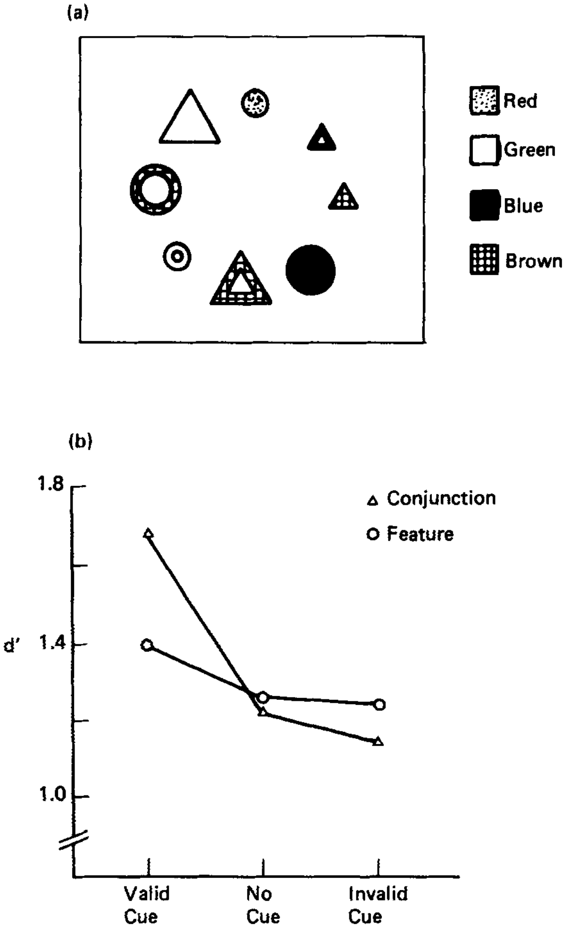


FIG. 6. (a) Example of displays used to measure costs and benefits of advance cues to the location of a target defined by any of four single features or by a conjunction of four features. (b) Accuracy (*d'*) in detecting the target with and without cues to its location.

matched the accuracy of performance for feature targets and for conjunction targets by presenting the display for a longer duration for the conjunction targets (though never more than 150 msec, to minimize the effect of eye movements). The question we asked was whether the effect of the cue would be greater for conjunction than for feature targets.

Figure 6b shows the results: for conjunction targets, there was a substantial benefit from a valid cue, whereas for feature targets the cue had very little effect. The direction of spatial attention seems to be irrelevant when a target is defined by a single easily discriminable feature, but has a large effect when the target is defined by a conjunction of equally discriminable features (see also Treisman, 1979).

Similar results were obtained by Prinzmetal, Presti, and Posner (1986). They explored the effects of a pre-cue that indicated the general area in which a four-letter display would appear. They found a small (3%) but significant reduction in the latency to detect targets defined by a unique colour or shape (feature targets), and a larger (12%) benefit in detection of targets defined by a conjunction of colour and shape. As I said earlier, feature targets that are not highly discriminable from the distractors may require narrowed attention to increase the signal-to-noise ratio, so that search becomes serial at least across groups of items (Treisman, 1985; Treisman & Gormican, 1988). When Prinzmetal et al. eliminated feature errors by making the feature targets more discriminable, conjunction errors remained high and still showed benefits (5.5%) from a spatial pre-cue. Note, however, that their pre-cue indicated only the general location of the display, not the location of the target within the display. According to feature integration theory, attention would have to be further narrowed to eliminate conjunction errors. Presumably this could begin earlier if attention was already in the right area (see also Appendix 1, p. 237).

*Dependence of Identification on Localization.* The second source of converging evidence for the role of attention in feature integration depends on the idea that visual attention operates by selecting stimuli in particular locations; its medium is a representation of space rather than of properties. Conjunctions of properties should, then, be correctly identified only when they are also correctly localized. We asked subjects to decide which of two conjunction targets was presented (a red "O" or a blue "X" among red "X"s and blue "O"s), and also to locate it in a 2 × 6 matrix of positions (Treisman & Gelade, 1980). We then looked at the number of correct identifications on trials on which the target was incorrectly localized (by at least two positions in the matrix). We found that performance was at chance. Again, this was not the case for feature targets defined by a unique shape or colour (an "H" for an orange letter); with these separate feature targets, subjects identified substantially more than the chance expectation even when they mislocated the target by more than one cell in the matrix. Separate features can apparently be identified without first being accurately localized. The converse almost never occurred in this experiment: if the location was correct, so was the identity.

Of course, the identification task (with only two alternatives) may have been easier than the localization task (with six possible locations). If subjects had been asked which half of the display contained the target and which of six features defined the target on each trial, the results would probably have been different. Only one or two fixations of attention would be needed to locate the target to the right or left of centre (one per half-display). On the other hand, features that are less discriminable (because target and distractors share them to differing degrees) may require more narrowly focused attention. The important point in the present results was the marked difference between features and conjunctions in the interdependence of location and identity when the number of alternatives was the same for the feature and for the conjunction task and when discriminability at the feature level was actually higher for the conjunction targets (red vs. blue, and "X" vs. "O") than for the feature targets (orange vs. red and blue, and "H" vs. "X" and "O").

*Illusory Conjunctions with Divided Attention.* Another source of converging evidence appeared when we forced subjects to divide their attention across several different objects (Treisman & Schmidt, 1982). The primary task was to report two black digits, one placed at each end of a row of three coloured letters. As a secondary task, subjects were to report the colour and shape of any letters they were reasonably confident they had seen. As predicted, the conjunction process broke down in this distributed attention task. Subjects saw many "illusory conjunctions," recombining properties of presented objects. Given a red "X", a blue "T" and a green "O", they might report a green "T" or a red "O". At least some of these conjunction errors do appear to be perceptual illusions rather than memory errors or guesses. Several subjects even broke off in the course of the experiment with comments such as "Oh, you are fooling me: the digits were coloured that time" (an event that never actually occurred). We found the same type of errors with components of shapes: for example, given displays containing "S"s and straight lines, subjects reported many illusory dollar signs. They saw these even when the straight line had to be taken from a different figure—an arrow or a triangle. It seems that unattended objects can exchange parts as well as properties. We inferred that simple, highly discriminable features (whether parts or properties) can be automatically identified with or without focused attention, but that they are accurately located and conjoined only when attention is narrowed to exclude the features of other objects also present in the display.

*Iconic Memory and Conjunctions of Features.* One could still argue, perhaps, that conjunctions are present initially, but that they decay rapidly from iconic memory. More recently, Marcia Grabowecky and I have used a partial report procedure, cueing one coloured letter in a clockface display of eight, at various intervals, immediately before and up to one second after presentation. In one condition subjects reported only the shape of the cued item; in another they reported only its colour; and in a third they reported both. The question we asked was whether, at any interval we tested, report of the conjunction of colour and shape would exceed the probability predicted from independent reports of the colour and of the shape. If colour and shape are independently detected and no additional conjunction information is present, the probability of reporting the conjunction of colour and shape should simply be the product of the probability of reporting the colour and the probability of reporting the shape, when each was the only task required. However, if any holistic code of the conjunction were initially laid down, report of the conjunction should exceed the prediction from feature independence. In fact we found no significant excess at any interval (see Figure 7), suggesting that both immediate report and retrieval from iconic memory depend on separate identification of each of the two properties in the cued location. There seems to be no additional Gestalt of "blue T-ness" or "O-ish redness"!

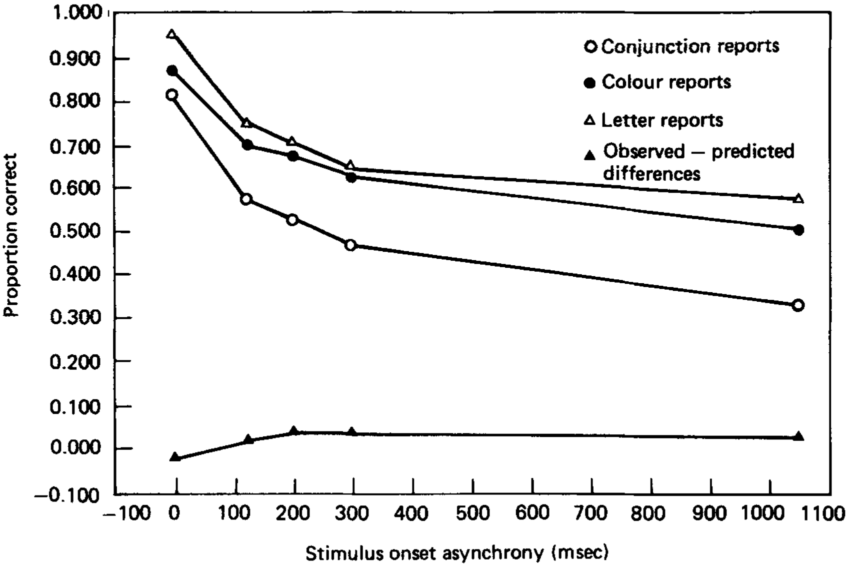


FIG. 7. Mean probability of correctly reporting the letter, the colour and the conjunction at different cue delays. The line joining the filled triangles shows the difference between observed performance on the conjunction and the predicted performance based on independent identification of each feature separately.

## Top-down Effects in Object Perception

We rarely experience illusory conjunctions in the so-called "real world" outside the laboratory—or rarely notice that we have. My friends and research assistants do sometimes come with anecdotes to please me: for example, there was the occasion on which a friend turned to greet his colleague in the street, only to find that the bald head and glasses belonged to one face and the black beard to another. But even I have to admit that these experiences are few and far between. Perhaps we constrain the conjunctions we form to fit our knowledge of familiar objects in the world; we rule out furry eggs and purple dogs before they reach conscious awareness. In some of my talks, I used to flash a slide of a magazine picture of a woman in a red blouse sitting in a flowered chair on a striped rug in a room with a yellow lamp, to demonstrate to the audience that natural scenes are not immune to conjunction errors. Each of those properties would frequently migrate to another object. Unfortunately I lost the slide during my move to Berkeley. I thought it would be easy to replace, but so far, to my surprise, I have failed to find a picture with more than one or two arbitrary, exchangeable properties. The moral I draw is that the risk of illusory conjunctions that we face in the real world in any single glance may be quite low. But this assumes that we use top-down information to rule out nonsensical object-property combinations.

*Feature Integration with Familiar Objects.* Together with Deborah Butler, I tested the effects of expectancy on illusory conjunctions using displays like those with coloured letters that had earlier given rise to many illusory conjunctions (see Figure 8a). The twist was that we could call the stimuli either non-committal names (ellipse, bars, triangle, arrow, ring) or names that constrained the expected colours, (lake, logs, carrot, tree, and tyre). We found that the constraining labels did indeed eliminate conjunction errors when the objects were presented in their expected colours. Subjects were no more likely to report, for example, an orange lake when a blue lake and an orange carrot were present than when no orange or no lake was present in the display.

We then asked a further question: does this constraint reflect a top-down influence on the selection of which features to conjoin, or does it simply rule out unacceptable conjunctions if any are formed accidentally. For example, when expecting a carrot, do we set up a "frame" (Minsky, 1975), with slots for orange and for elongated triangle, which guides the conjunction of features? Or is the conjunction a bottom-up process, constrained only by spatial attention, with a subsequent check for familiarity once the perceptual object has been assembled? We ran a further experiment in which we presented the same objects with familiar labels but occasionally switched their colours (see Figure 8b). The question was whether subjects would form

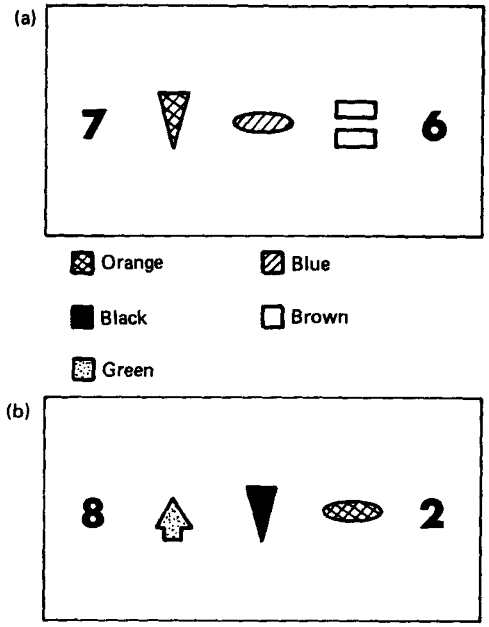


FIG. 8. Examples of displays used to test top-down effects on the conjoining of features. (a) Display containing stimuli with expected associations of colours and shapes labelled as tree, lake, tyre, logs, carrot. (b) Display containing mispaired colours and shapes, to test whether illusory conjunctions would be generated to correct the anomaly.

illusory conjunctions to *correct* the anomaly. In fact, there was no evidence that they did, although they did misperceive the individual features to fit their expectations. In other words, subjects were no *more* likely to misperceive a green triangle as an orange triangle (a carrot) when the colour orange was present elsewhere in the display than when it was not; (this was our measure of true illusory conjunctions).

If these results can be trusted (and I will try to replicate them), they suggest that the process of conjoining is a bottom-up one, controlled only by spatial attention. Once a set of features has been selected, expectations may bias the names we give them or constrain our guesses to fit the pre-specified description of familiar objects. But it seems that these pre-specified object schemes do not "hunt" through the scene for the physical features to match their slots, collecting them into the appropriate bundles regardless of their true locations. In the model in Figure 1, I show long-term memory as interacting only with the object level, after the features have already been conjoined.

*Constraints from Figure–Ground Relations.* We have recently explored one constraint that may be at least partly observed—the distinction between figure and ground. In one experiment, we looked for illusory conjunctions that might exchange colours or shapes between a background and a superimposed shape (as in Figure 9). The method was the same as in the previous experiments, except that the item to be reported was cued immediately after the display, together with the mask. The cue was equiprobably a pointer on the left for the left figure, on the right for the right figure, pointers above and below for the ground, and the word "digits" when the digits were to be reported. As in the earlier experiments, the instructions were to attend to the digits, to ensure that attention would be divided over the whole display rather than focused on any one figure. However, report of the figures and ground were not delayed by prior recall of the digits, as these were tested alone, and only on a quarter of the trials.

Subjects again made a substantial number of conjunction errors, averaging 10.1%. Significantly more of these exchanged features between the two small figures than between one figure and the background (14.1 compared to 2.2). There was, however, an interesting difference between colour and shape

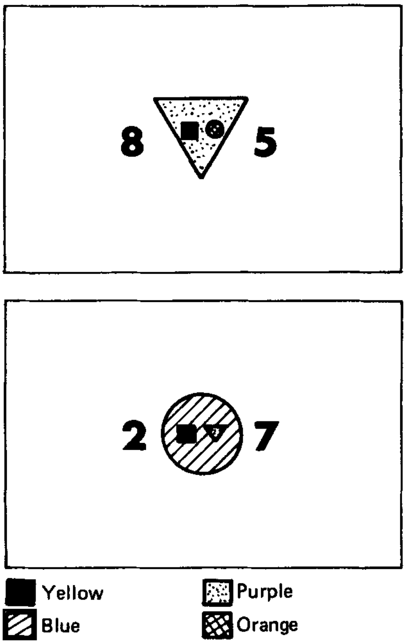


FIG. 9. Examples of displays used to test whether illusory conjunctions would be formed between figure and background.

in the extent to which the figure–ground relation constrained the migration of features. Whereas both colour and shape migrated between the two figures, subjects never exchanged colours between a figure and its background. In fact, they made almost no errors of any kind in reporting the background colour.

Shapes, however, did migrate between figure and ground. Despite the difference in scale and the hierarchical structure of the figure-ground relationship, the shapes migrated only slightly less often between figure and ground than between two figures; (the difference was not significant). We had previously found no effect of similarity between two objects in constraining the rate of illusory conjunctions (Treisman & Schmidt, 1982). A small blue circle would, for example, as often lend its colour to a large filled triangle as to another small circle. Our new result extends this principle across a larger difference of size, and across the roles of figure and ground for exchanges of shape but not for exchanges of colour. The colour of the background may have been too salient to be misperceived at all. However, there might be an alternative explanation, to which I return at the end of this paper.

## "Object Files" in Perceptual Representation

The experiments so far have dealt with the question of how we select the features to be conjoined. Can we say any more about the object representations into which the selected features are entered? The experiment with familiar labels suggested that features are conjoined *before* being compared to stored descriptions for identification. These temporary assemblies of features with their spatial relations must therefore be distinct from the nodes in semantic memory whose activation mediates perception in many current information-processing models. Daniel Kahneman and I have argued that many perceptual phenomena depend on some such "episodic" representations (to borrow a term from the theory of memory proposed by Tulving, 1972). They collect the incoming sensory data from the currently attended object and update them when changes are detected (Kahneman & Treisman, 1984). We called them "object files", by analogy with a file the police might open to record and assemble all the accruing information relating to a particular crime or accident. One important source of evidence comes from divided attention tasks. Attention load seems to be measured in terms of the number of *objects* present. Thus, the ease of dividing attention to code two different properties depends on whether they are seen as properties of the same object or of different objects (Treisman, Kahneman, & Burkell, 1983). Response latency and errors in locating a gap in the contour of a shape while concurrently naming a word are greater when the word and the shape form two separate objects (see Figure 10b), than when they can be seen as parts of the same global object (see Figure 10a). Exactly the same number of

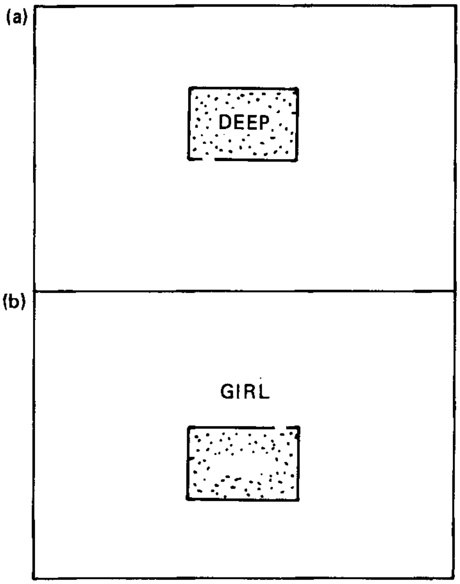


FIG. 10. Examples of displays used to test the efficiency of divided attention (a) to two parts of one composite perceptual object and (b) to two separate perceptual objects.

labels would apply in both cases (e.g. word, colour, rectangle, gap, left or right). Only the number of separate perceptual objects differs. A natural inference is that attention load is determined by the number of separate representations (object files) that must be set up. (See Duncan, 1984, for a similar set of results). Another essential function that object files would serve is to individuate different, otherwise identical, replicas of the same object when more than one is present in the field. Norman (1986) pointed out the need for some such distinction between types and tokens. Kanwisher (1987) has recently demonstrated the difficulty subjects have in seeing both of two identical stimuli in a rapidly presented sequence, and interpreted it as a failure to set up separate tokens of the same type when the maximum rate of token individuation is exceeded for any one type.

A final reason for distinguishing episodic object files from semantic nodes is to account for our ability to maintain the perceptual unity and continuity of objects as they move and change. A distant aeroplane retains its continuity as a single perceptual object, even when we see it flap its wings and alight on a nearby tree, thus forcing us to change the label we initially assigned. A new node, for bird instead of aeroplane, becomes active, but we see a single, continuing object. "Object files," as we conceive them, are addressed by their spatial and temporal co-ordinates rather than by labels for their identity or for any properties they may contain. Kahneman, Gibbs, and I (1983; Treisman & Kahneman, 1983) have recently explored the unity and continuity of objects across time and space, using stimuli that move and change within displays. In several studies, we asked subjects to identify an object (e.g. name a letter) and looked at the effects of its immediate past history on the speed with which they could respond. We discovered a form of object-specific priming that we call "re-viewing". For example, in one experiment we presented two successive pairs of letters, with the second pair displaced to the right or left of the first (see Figure 11). When the onsets of the two displays are separated by an interval of 130 msec, the perception is of one pair of letters moving to the left or right. This global apparent motion is similar to the effect studied by Ternus (1926). The direction of motion of the whole display is determined by the location of the peripheral letter in the second display. If it appears to the left, *both* letters are seen to move left; if it is to the right, both move right. The other letter appears at the fixation point in either case. Subjects were asked to name the letter in the second display that was presented at fixation and cued by bar markers. The naming latency was reduced when the target letter matched the initial letter that appeared to move into it, but not at all when it matched the other letter from the initial display. So, what seems to be critical is whether the priming letter and the target letter to be named are integrated into the same object representation. If they are seen as separate objects, no re-viewing advantage is observed, although the letter nodes in the hypothetical recognition network must on average have been equally primed in both cases.

Priming that is *not* object-specific (whether repetition priming across

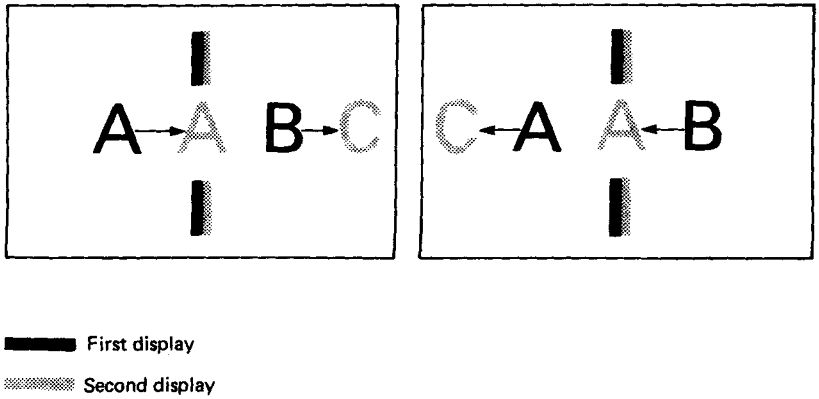


FIG. 11. Example of displays used to test the "re-viewing effect". The latency of naming the cued letter is reduced only when it matches the initial letter that is perceptually integrated with it.

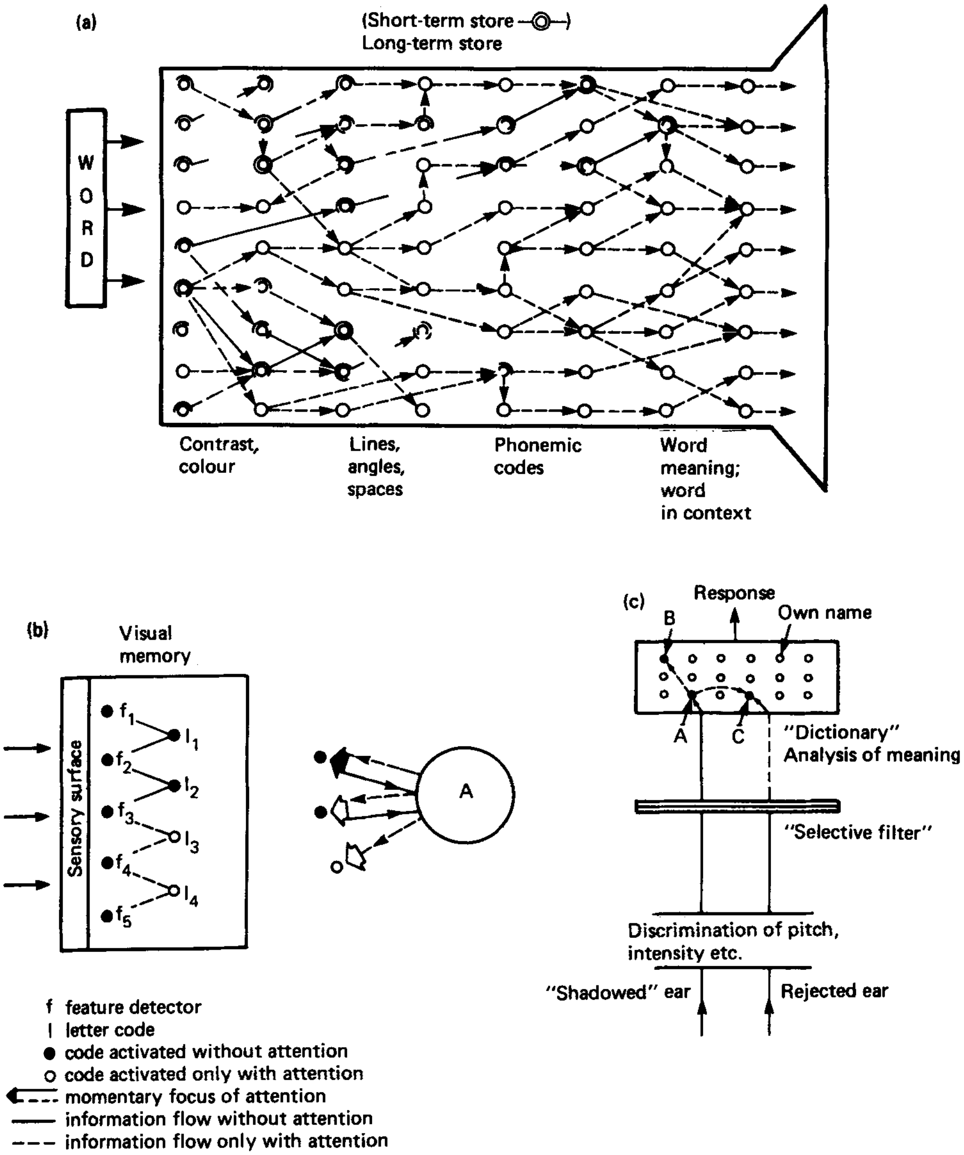


FIG. 12. Examples of "display board" models that equate perception with the activation of nodes in a recognition network. (a) An analogic depiction of short-term store embedded within long-term store (from Shiffrin, 1976). (b) Model of perceptual processing showing two states of coding visual letter patterns. Arrows from the Attention Centre (A) to solid dot codes denote that attention can activate these codes, and in turn be activated (attracted) by them (from LaBerge, 1975). (c) The thresholds of words B and C are lowered by their high transition probability after word A. Word C is also activated by this "attenuated" signal from the rejected ear and is sometimes heard (from Treisman, 1960).

longer intervals or associative priming between different objects) would still, we assume, be mediated by the activation of nodes within a recognition network, as in the "display board models" of mental life (Kahneman & Treisman, 1984; see examples in Figure 12). Having proposed one of the early display board models myself (Treisman, 1960), I still believe they have an important explanatory role to play. However, I would use them in explaining identification and classification, but not directly to mediate "seeing". The re-viewing results confirm the need to separate the recognition network from the temporary object files, as shown in the model in Figure 1.

Another demonstration that we relate to object files involves the apparent integration across time and spatial displacement of separately presented components to form a composite shape when the components are presented within the same perceptual object (Treisman & Kahneman, 1983). A horizontal or vertical line was briefly presented in each of four squares in an initial display (see Figure 13). The squares then moved to new locations, and a second line was presented in each square. Subjects were asked to detect a plus in any of the squares in the second display. A plus replaced the line in one square on half the trials. In one condition, the lines in the final display would form pluses if they were superimposed on the lines in the same moving shape

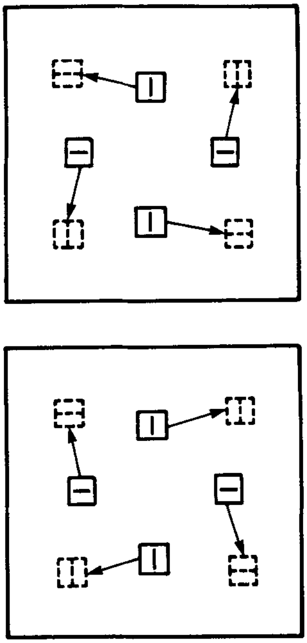


FIG. 13. Examples of displays used to demonstrate the object-specific integration of parts (two lines) into compound shapes (pluses).

from the initial display. In another condition, the lines in each shape were in the same orientation in the initial and in the final display. In this case no plus could be generated by perceptually integrating the two lines that appeared in the same shape across the two displays. On negative trials with no plus presented, subjects were slower to say that no plus was present when the components to make a plus were both presented in the same square, one before and one after it moved to its new location. Again, what linked the two lines seems to be the fact that both were entered into the same object representation, despite the change in the object's location between the appearance of the first and second lines.

## Recent Developments and Their Implications for the Theory

The story so far seemed reasonably coherent, and it fitted the data we had collected. The rest of the paper outlines some other new findings, which raise problems for the theory, and suggest some possible ways to modify the framework I proposed.

*Contingent Aftereffects and Attention* The first finding is by Houck and Hoffman (1986). They used the McCollough effect (McCollough, 1965) to test whether the coding of conjunctions always depends on focused attention. Subjects looked at displays of 4, 8, or 12 patches of alternating green or red, horizontal or vertical stripes. Any one patch alternated between two complementary stimuli, either red vertical with green horizontal, or green vertical with red horizontal. McCollough has shown that after staring at one such alternating pair, subjects experience a "contingent aftereffect". For example, after adapting to red vertical alternating with green horizontal, they see black and white vertical stripes tinged with green (the complementary colour to the red vertical) and black and white horizontal stripes tinged with red. Houck and Hoffman varied the deployment of attention and the number of McCollough patches. Their subjects focused either on a central monitoring task (to detect a missing dot in a 3 × 3 matrix), or on a peripheral monitoring task (to detect a differently oriented "C" among a set of "C"s), or they divided their attention between the central and the peripheral tasks. The results were clearcut: neither the attention tasks nor the different display sizes made any difference to the size of the McCollough aftereffect. Since the aftereffect reflects adaptation to a conjunction of properties, the result seems to challenge the claim that conjoining properties requires attention.

The most likely explanation, I believe, is one suggested by Houck and Hoffman—that the McCollough effect reflects a very early stage of processing, either before or at what I've called the master-map of locations, and preceding the analysis of separable features by specialized modules. The McCollough effect is both monocular and tied to particular retinal locations,

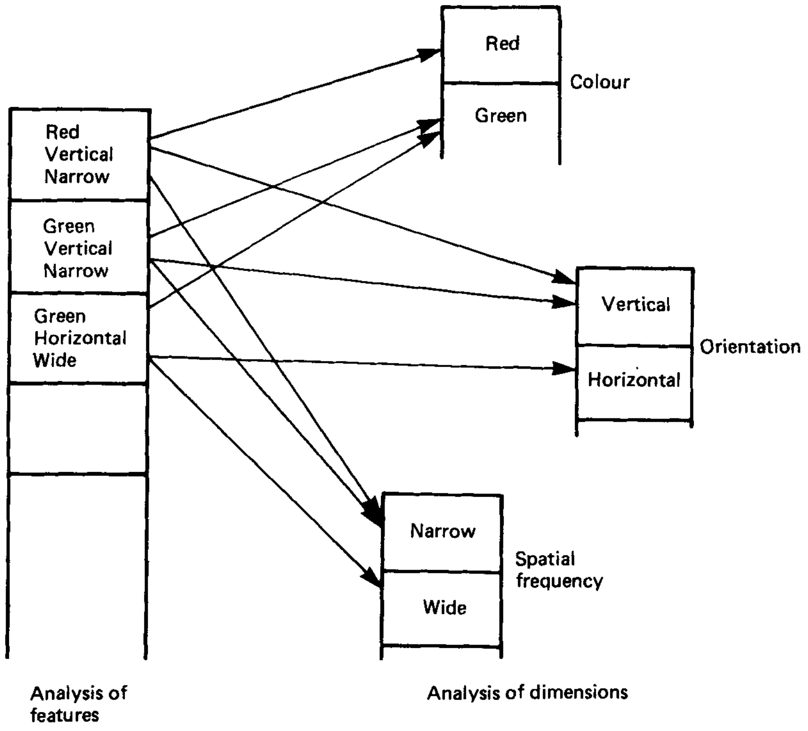


FIG. 14. Functional scheme showing specialization of coding first for specific values on different dimensions and next for whole dimensions.

which suggests early vision. Physiological recordings also show early specialization for particular *values* on dimensions but little segregation of different dimensions as such. Most cells in V1 have preferred values of orientation, spatial frequency, and ocular dominance, but any one cell is specialized along a number of dimensions. Many of the dimensions appear to be organized in orthogonal patterns of layers, alternating columns and hypercolumns, which group neighbouring values in neighbouring cells. This organized structure may be an initial step towards further specialization in other visual areas, as shown schematically in Figure 14.

*Parellel Coding of Conjunctions?* The next findings to raise problems for feature integration theory were made by Nakayama (in preparation). He has recently tested visual search for conjunction targets defined on several dimensions that I had not previously explored. I had data for conjunctions of colour with line orientation (using the letters "T" vs. "X"), colour with curvature (using the letters "X" vs. "O"), colour with line length, and conjunctions of different parts of shapes (the components of dollar signs, "R"s, "T"s, and arrows). I found apparently serial search with all these conjunctions as targets. Nakayama has also found serial search for colour with orientation, for spatial frequency or size with orientation and initially with colour and motion. The first exceptions he found were conjunctions of binocular disparity with motion and with colour, both of which yielded parallel detection (Nakayama & Silverman, 1986). This could be reconciled with my account if we assume that attention can select a depth plane in the same way as it selects an area in the frontal plane. Within the selected plane or area, a feature search on one dimension would reveal the conjunction target without any need for serial scanning, as if the distractors that shared its distinctive property had not been presented.

More recently, however, Nakayama has discovered a display of colour and motion that also allows the conjunction target to "pop out" perceptually in parallel, and he has found several other pairings of dimensions that produce the same result. In addition to disparity with all other dimensions tested, these include size or spatial frequency with direction of contrast (black vs. white) and with colour. What are we to make of these findings? The first surprise is that the properties which are most clearly conjoined physiologically in area V1 (spatial frequency and orientation) are not among those allowing pop-out, whereas the properties that seem physiologically most separable from each other (colour and motion) are. An account in terms of conjunction detectors at the level of single units in striate cortex does not fit the data on relative difficulty. Some alternative way of integrating the findings must be found.

Phenomenologically, Nakayama describes the displays that do allow conjunctions to be detected as forming clear and salient planes, segregating the two types of distractors. Just as disparity segregates one plane in depth perceptually from the other, so the version of colour–movement conjunctions that allows pop-out seems to allow selective attention to either of two perceptually segregated planes. For example, subjects can ignore the imaginary plane to which the red squares oscillating up and down are attached, and attend to the plane with the green squares oscillating left and right. In fact, the motion seems to create one global green figure, on which the odd red one stands out from the rest. If these introspective reports can be linked to objective measures of perceptual segregation, we can retain the theory that attention is required in conjunction search. Certainly if the two sets of distractors in Nakayama's tasks were spatially grouped in the left and right half fields of the frontal plane, there would be no problem in attending selectively to the left or right half and finding the red item among the green ones (Treisman, 1982). The difficulty in applying the same account to Nakayama's results with distractors that are spatially intermingled in the frontal plane is the assumption that attention can never be spatially divided between two or more areas. The belief that spatial attention is unitary derives from findings by Posner et al. (1980) testing the detection of single light flashes in an otherwise empty field. Generalization from these results to all other attention tasks may have been premature. Several studies have shown that attention can be quite effectively directed to one of two superimposed shapes (Rock & Gutman, 1981; Tipper, 1985) or scenes (Neisser and Becklen, 1975).

The question then becomes under what conditions and how can we attend selectively to one spatially intermingled group as a unit and reject another? At this point, I can only speculate; further research will be needed to test the ideas. One starting point is the observation that an item with a highly distinctive feature tends to "call attention" to itself; this is the basis of the "pop-out" phenomenology. In previous accounts, I have left open the question how attention is controlled. The model I sketched in Figure 1 shows the selection of attended features to be made through a common spatial reference in the master-map of locations. The selection is made serially, either following a scan-path in search or directed by a spatial cue given in advance. The "calling" of attention by a salient feature suggests the possibility that locations in the master-map might also be selectively activated or inhibited through links downward from particular feature maps. Thus, if a highly distinctive value is present against a background for which all the other activity is concentrated in another widely separated feature map, mutual inhibition between the replicated features might be fed down to the master-map locations that contain them. The resulting stronger activation of the location of the unique feature in the master-map might be indistinguishable from the activation that would be produced by focused attention to that location. If so, it would produce the same consequence: namely that the features linked to the active master-map location in all the other feature maps would be automatically accessed and conjoined with the salient feature, just as if attention had reached that location in the course of a normal serial search, or had been voluntarily directed to it in response to an advance cue.

This account can be extended to cover Nakayama's results if we assume that it is possible voluntarily to inhibit master map locations that contain *non*-target features through downward links from the feature-maps, when the features of the target and the distractors are highly discriminable. So, for example, if the target is a red bar moving up and down, among green bars moving up and down and red bars moving left and right, all locations containing green might be inhibited, and/or all locations containing motion left and right (see Figure 15). Whereas spatial selection by focused attention to a cued location seems to be limited to one region at a time (Posner et al.,

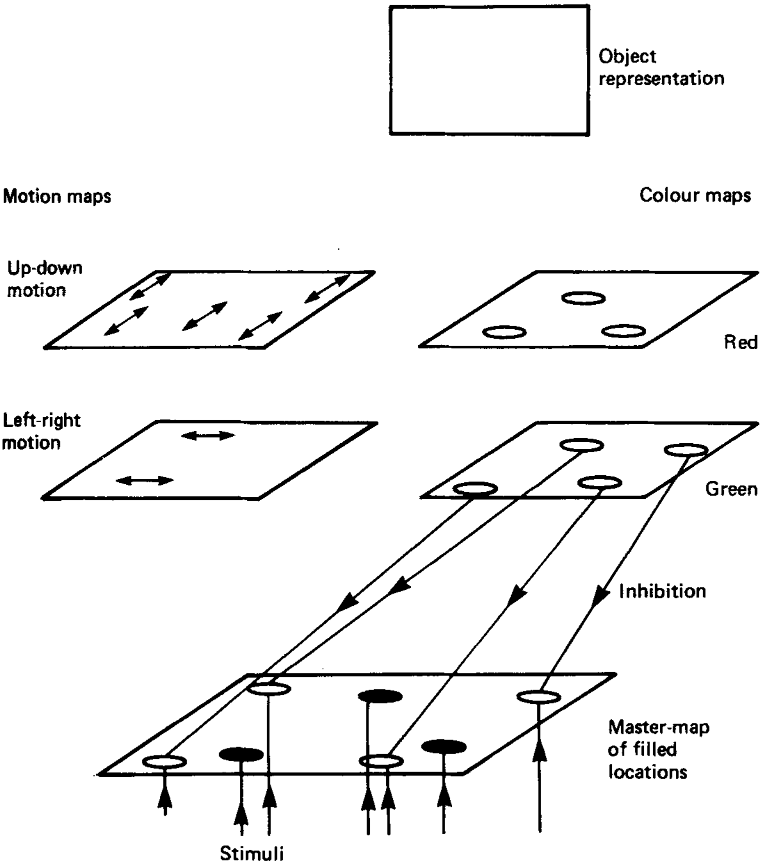


FIG. 15. A possible mechanism for parallel access to a subset of spatially intermingled objects. In this example, locations containing green are inhibited, leaving locations containing red more strongly activated. Any of these locations that is also linked to activity in the up-down motion map must contain the target (red bar moving up and down). For the sake of clarity, the upward links between the active master map locations, the feature maps, and the object representation are omitted. The attention spotlight could select either all locations or a subset of the remaining active locations, the number depending on how effectively the interspersed distractor locations have been inhibited.

1980), selective inhibition from a feature-map need not be restricted in the same way.

Wolfe, Franzel, and Cave (1988) recently reported similar results with highly discriminable feature conjunctions, and proposed a similar account based on inhibition at the feature level. They pointed out that conjunction targets logically need not be detected by a process of conjoining features;

instead one might simply reject any item that had a mismatching feature, leaving the target as the only surviving item. They showed that a triple conjunction target can generate completely flat search functions when it differs from the distractors in two of its features. For example, a large red "O" pops out among small red "X"s, small green "O"s, and large green "X"s. Within the framework shown in Figure 1, inhibition from two different feature maps would converge on each distractor location in the master-map, increasing the difference in activation between the target and the distractors. On the other hand, when each distractor had both target features and the target was defined only by the particular *arrangement* of its features (target "T" among distractor "L"s), Wolfe et al. confirmed that search remains serial. In this case, subjects would be forced to conjoin features in order to identify the target, using a serial scan with focused attention.

What remains unexplained is the range of different search rates that can be obtained with conjunction targets. Table 1 shows the mean slopes, intercepts, and measures of linearity that we have recently obtained in a study replicating Nakayama's experiments. We used bars with highly discriminable features on four different dimensions [pink vs. green, orientations 45° left vs. 45° right, large (1.7°) vs. small (0.9°), and motion oscillating 0.4° up and down vs. left and right]. We tested search for conjunctions of values on every possible pair of these dimensions (e.g., colour with size; colour with motion). When a dimension was not one of the pair being tested it took on a neutral value (grey, vertical, medium-sized (1.3°), and stationary). We controlled density by presenting displays of 4 or 9 within a randomly selected 2 × 2 or 3 × 3 array of the 4 × 4 matrix (7.6° by 7.6°) used for the largest display of 16

TABLE 1
Results of Experiment on Search for Targets Defined by Conjunction of Colour, Size, Motion and Orientation

| | *Slope* | | *Intercept* | | *% Linearity*[a] | |
|---|---|---|---|---|---|---|
| | *Positive* | *Negative* | *Positive* | *Negative* | *Positive* | *Negative* |
| Colour-Size | 6.9 | 12.2 | 434 | 447 | 99.90 | 98.40 |
| Colour-Motion | 9.8 | 20.1 | 590 | 567 | 98.80 | 99.60 |
| Colour-Orientation | 16.6 | 27.5 | 505 | 457 | 100.00 | 100.00 |
| Size-Motion | 8.6 | 17.9 | 598 | 572 | 95.30 | 97.80 |
| Size-Orientation | 12.9 | 25.8 | 529 | 481 | 99.50 | 100.00 |
| Motion-Orientation | 13.9 | 43.5 | 903 | 802 | 96.60 | 99.60 |

[a]Note: This is the percentage of the variance due to display size which is contributed by a linear component ($r^2$).

items. The slopes (obtained from 16 subjects after one hour of practice) range from about 10 msec to about 40 msec per item for targets defined by the different combinations of features. All, however, are linear and they approximate the two-to-one ratio of negative to positive slopes that suggests a serial self-terminating scan.

We must explain, then, how feature-based inhibition would generate these varied linear search functions. In fact, it seems possible to use the same framework as we proposed to deal with the results obtained in search for targets defined by a single feature, in conditions in which the targets either *lacked* the relevant features or shared it to differing degrees with the distractors (Treisman & Gormican, 1988; Treisman & Souther, 1985). As I said earlier, we explained the wide range of search rates we obtained in these less discriminable feature search conditions with the assumption that subjects could search groups of items serially. The size of the group would be determined by the discriminability of the target. So instead of a dichotomy between parallel, pre-attentive search and serial search with focused attention to each item individually, we proposed a continuum from narrowly focused attention to broadly divided attention. In conjunction search with feature-based inhibition, some locations in the master-map would be more highly-activated than the others. Either half of the distractor locations (if inhibition is controlled by only one feature) or all the distractor locations (if both distractor features generate inhibition) would be transmitting little activity. The serial scan could therefore use a wider aperture to distinguish a group that contained the target from a group that did not. At the extreme, successive attention to the two halves of the display, regardless of the number of items, could be sufficient to distinguish a display containing a target (asymmetrical activity) from a display that did not (more uniform activity).

If this account is correct, the difference between displays that allow rapid detection of conjunctions and those that do not is determined not so much by which dimensions are tested but by whether the values used on those dimensions are sufficiently separable to allow the selective control of master-map locations. Nakayama's bright red and green patches on a dark background and our pink and green bars were much more discriminable than the thin red and green lines of the letters that gave serial search functions in our earlier experiment (Treisman & Gelade, 1980; Experiment 2).

The feature inhibition hypothesis also provides a mechanism for figure–ground segregation. The relative activation of different areas within the master-map of locations can be modulated either by selective inhibition from a feature map (e.g. for red areas leaving green areas active), or by attention to one particular area when no feature-based control is possible. The phenomenology would be the same in the two cases; a set of items would stand out perceptually *either* when they are spatially grouped and attention selects the region that contains them, *or* when they are spatially intermingled with others but share one or more highly distinctive features that are sufficiently segregated within their feature modules for selective inhibition of master-map locations to be effective.

*Feature Coding in Other Media.* The third set of findings that will elaborate my model of feature coding and object perception arise from a question that Patrick Cavanagh, Martin Arguin, and I have begun to explore (Cavanagh, Arguin, & Treisman, in preparation). It is related to Nakayama's results and also to a question that this paper so far has begged—the question of how to define a feature. My approach has been to treat this as an empirical question that might be answered by using converging operations—or rather by using a number of different operational tests and seeing whether they do in fact converge on the same candidate features (Treisman, 1986b). If there is an elementary alphabet of visual building blocks or primitives, they are likely to be identified early in visual processing, and without any complex analysis requiring attention. They may be "hard-wired" into the structure of the visual system, either innately or through early or prolonged experience. They are likely to play a role in segregating figure from ground, as a prerequisite for the later, more complex processing necessary to identify objects and events. We might expect them to reveal themselves in some or all of the following behavioural tests:

1. automatic and spatially parallel detection, as shown by the "pop-out" test in visual search;
2. easy and salient perceptual segregation of areas that contain them from areas that do not;
3. separable or modular analysis, shown by slower detection of disjunctive targets when these are defined on different dimensions and by the absence of interference from irrelevant variation in other modules;
4. interchangeability between objects when attention is divided or diverted, as shown by the occurence of illusory conjunctions;
5. the partial independence of correct identification from correct localization for stimuli defined by a single separable feature.

The experiments I have described showed that some features at least meet all five criteria. Examples include distinctive values on the dimensions of colour, line orientation, size, and curvature. We have also tested several more candidates, including closure, line ends or terminators, and angles (Treisman & Paterson, 1984; Treisman & Souther, 1985), and found converging evidence supporting the featurehood of the first two, but less evidence for the angles.

All the stimuli I have described so far were defined by the spatial distribution of luminance differences; Cavanagh (1987) has studied perceptual objects whose boundaries are defined in other media than that of luminance differences. He has explored the phenomenology of three-dimensional forms defined only by motion, texture, colour, or binocular disparity (Figure 16). He found that, although all these separate media allow the identification of form, not all give good depth and surface inferences from two-dimensional cues; colour and texture in particular did not.

The existence of separate media in which shapes can be defined raises another question in the context of the research I have described. Would the features of shape that function as primitives by the tests I've developed also do so when they are defined not by luminance but by texture, motion, colour, or disparity boundaries? If so, we might infer that a recursive extraction of the same vocabulary of spatial features is repeated within each dimensional module. In a typical experiment, we might create oriented texture bars that differ from their background only in being stationary while the background moves, or only in their binocular disparity relative to the background, or in the different spatial frequency of their random texture elements. In all other

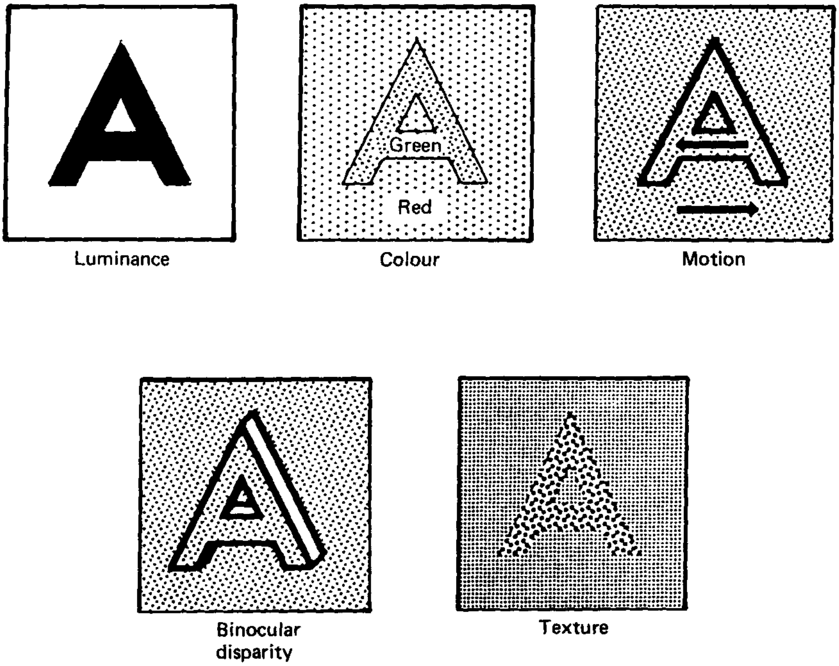


FIG. 16. Different media in which forms can be defined (from Cavanagh, 1987).

respects, these bars can be characterized by the same set of features as luminance bars; they will have lengths and widths, orientations, degrees of curvature, terminators, and so on.

So far, we have tested the detection of targets defined by a unique orientation or by a unique size, using random texture bars or spots whose boundaries were created by relative motion, or by binocular disparity, or by differing in spatial frequency from their backgrounds. Performance in visual search with these stimuli shows very similar results to those obtained with shapes defined by luminance boundaries. For example, a tilted texture bar defined by relative motion pops out of a display of vertical texture bars with no effect of the number of vertical distractors, as if its orientation were coded automatically and in parallel, without focused attention. The conclusion seems to be that the vocabulary of primitive shape components can be extracted in the same parallel and automatic way at a number of different levels. It becomes a more general and abstract visual language than at first seemed likely. The oriented bar or grating detectors found by Hubel and Wiesel (1959) in V1 and V2 may be just one instantiation of a more general feature extraction process.

There is, however, a problem for the theory I originally proposed: in a sense, each shape feature that pops out in these new detection tests is actually a conjunction of its own defining property with the medium that carries it—for example orientation with motion, or length with disparity. Notice that these conjunctions differ in their logical structure from those tested by my earlier experiments, or by Nakayama. In the standard conjunction search experiments, the medium was always luminance; the target-defining conjunctions were of other properties that characterized the objects defined by luminance patches or edges, whether their colour, or their motion, or their disparity. In the displays that Cavanagh, Arguin, and I are now exploring, the objects are themselves created by variations in one property (the medium), and the crucial feature that differentiates targets from distractors is defined by another (we could call it the message). However, all the items share the same value in the medium—for example all are stationary against a moving background—and the target differs from the distractors in a different single property (the message), not in a conjunction of properties.

It may be helpful to differentiate a logical hierarchy of features (Figure 17), paralleling the hierarchy of levels of representation defined by Marr (1982). He characterized visual coding as moving from points to surfaces in a 2½D sketch and finally to 3D object descriptions. Similarly, we can distinguish the most basic properties—luminance and colour; these characterize points and areas of space. Discontinuities of luminance and colour can either directly form the boundaries of objects, or they can define the local elements of a texture medium in which a second class of properties defines the boundaries of objects. This second class of properties also characterizes

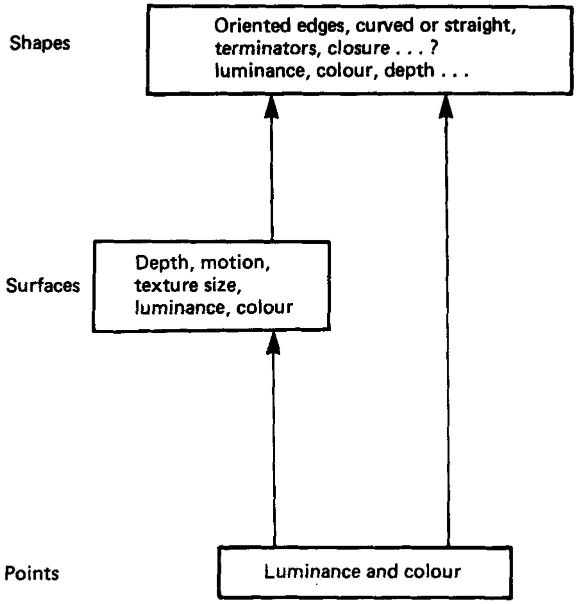

Shapes
Oriented edges, curved or straight, terminators, closure . . . ? luminance, colour, depth . . .
Surfaces
Depth, motion, texture size, luminance, colour
Points
Luminance and colour

FIG. 17. A feature hierarchy with features characterizing points in the frontal plane, surfaces in depth and motion, and the shapes of object boundaries.

points and patches, but now in terms of spatial and/or temporal variables rather than fundamental dimensions of light. It includes distance (carried by binocular disparity or monocular parallax), relative motion, and the spatial frequency or size of texture elements. Discontinuities in these properties can define the boundaries of objects in the same way as discontinuities in luminance or colour. Finally a third class of properties characterizes the shapes of boundaries (whether between different luminances, textures, depths, or relative motions). Edges have orientations, lengths, curvatures, angles, termination points, and possible higher order properties like closure, symmetry, convergence, containment (Treisman & Gormican, 1988). So we move from the dimensions defining points and areas, to those defining surfaces in depth, and their movements in time, and finally to those defining the shapes of two or three-dimensional objects.

What then of conjunctions and feature-integration theory? Perhaps attention is needed only to prevent illusory conjunctions *within* a class of properties—those that characterize a given set of objects defined by another medium—and not *between* the properties of the medium and the properties of the objects that it carries. This hypothesis provides an alternative account of the figure–ground experiment described earlier, in which colours were exchanged between two figures but not between a figure and its background.

The colour difference between the figure and the background (together with correlated differences in luminance) were the medium that defined the shape of the figure. The colours of the two figures, on the other hand, were not essential to differentiating them from each other.

Both these research projects are in their early stages. But if the preliminary results are confirmed, they will force some changes in the story I tell about the perception of features and objects. I hope that the possible directions I have indicated will prove fruitful when worked out in more detail—or that some more exciting alternative emerges to make sense of both the old and the new results.

## REFERENCES

Ballard, D. (1986). Cortical connections and parallel processing: Structure and function. *Behavioral and Brain Sciences, 9,* 67–120.

Barlow, H. B. (1986). Why have multiple cortical areas? *Vision Research, 26,* 81–90.

Briand, K. A. & Klein, R. M. (1987). Is Posner's "beam" the same as Treisman's "glue"? On the relation between visual orienting and feature integration theory. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 228–241.

Cavanagh, P. (1987). Reconstructing the third dimension: Interactions between color, texture, motion, binocular disparity and shape. *Computer Vision, Graphics and Image Processing, 37,* 171–195.

Cavanagh, P., Arguin, M., & Treisman, A., (in preparation). *Visual search in perceptual pathways.*

Cowey, A. (1979). Cortical maps and visual perception. The Grindley Memorial Lecture. *Quarterly Journal of Experimental Psychology, 31,* 1–17.

Cowey, A. (1981). Why are there so many visual areas? In F. O. Schmitt, F. G. Worden, G. Adelman, & S. G. Dennis (Eds.), *The organisation of the cerebral cortex.* Cambridge, Mass.: M.I.T. Press.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113,* 501–517.

Garner, W. R. (1970). The stimulus in information processing. *American Psychologist, 25,* 350–358.

Garner, W. R. (1974). *The processing of information and structure.* Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.

Harvey N. & Treisman, A. (1973). Switching attention between the ears to monitor tones. *Perception and Psychophysics, 14,* 51–59.

Houck, M. R. & Hoffman, J. E. (1986). Conjunction of color and form without attention: Evidence from an orientation-contingent color aftereffect. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 186–199.

Hubel, D. H. & Wiesel, T. N. (1959). Receptive fields of simple neurons in the cat's striate cortex. *Journal of Physiology, 148,* 576–591.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye movement. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 187–203). Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.

Kahneman, D. & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention.* New York: Academic Press.

Kahneman, D., Treisman, A., & Gibbs, B. (1983). Moving objects and spatial attention. 24th Annual Meeting of the Psychonomic Society, San Diego, California.

Kanwisher, N. (1987). Repetition blindness: Type recognition without token individuation. *Cognition, 27,* 117–143.

LaBerge, D. (1975). Acquisition of automatic processing in perceptual and associative learning. In P. M. A. Rabbitt & S. Dornič (Eds.), *Attention and performance V* (p. 52). London: Academic Press.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: W. H. Freeman.

Maunsell, J. H. R. & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience, 10,* 363–401.

McCollough, C. (1965). Color adaptation of edge-detectors in the human visual system. *Science, 149,* 1115–1116.

Milner, P. M. (1974). A model for visual shape recognition. *Psychological Review, 81,* 521–535.

Minsky, M. (1961). Steps towards artificial intelligence. *Proceedings of the Institute of Radio Engineers, 49,* 8–30.

Minsky, M. (1975). A framework for presenting knowledge. In P. H. Winston (Ed.), *The psychology of computer vision.* New York: McGraw-Hill.

Nakayama, K. & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature, 320,* 264–265.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.

Neisser, U. & Becklen, P. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology, 7,* 480–494.

Norman, D. A. (1986). Reflections on cognition and parallel distributed processing. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 2: Psychological and biological models.* Cambridge, Mass.: M.I.T. Press.

Posner, M. I. (1978). *Chronometric explorations of mind.* Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.

Posner, M. I. & Snyder, C. R. R. (1975). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornič (Eds.), *Attention and performance V.* London: Academic Press.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109,* 160–174.

Prinzmetal, W., Presti, D. E., & Posner, M. I. (1986). Does attention affect visual feature integration? *Journal of Experimental Psychology: Human Perception and Performance, 12,* 361–370.

Rock, I. & Gutman, D. (1981). Effect of inattention on form perception. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 272–285.

Shiffrin, R. M. (1976). Capacity limitations in information processing, attention and memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (vol. 4). Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.

Ternus, J. (1926). Experimentelle Untersuchung über phänomenale Identität. *Psychologische Forschung, 7,* 81–186. Abstracted and translated in W. D. Ellis, (Ed.), *A sourcebook of Gestalt psychology.* New York: Humanities Press, 1960.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology, 37A,* 571–590.

Treisman, A. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology, 12,* 242–248.

Treisman, A. (1969). Strategies and models of selective attention. *Psychological Review, 76,* 282–299.

Treisman, A. (1979). The psychological reality of level of processing. In L. S. Cermak and F. I. M. Craik (Eds.), *Levels of processing in human memory.* Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and objects. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 194–214.

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing, 31*, 156–177.

Treisman, A. (1986a). Features and objects in visual processing. *Scientific American, 254*, 114–124.

Treisman, A. (1986b). Properties, parts and objects. In K. Boff, L. Kauffman, & J. Thomas (Eds.), *Handbook of perception and human performance*. New York: Wiley.

Treisman, A. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97–136.

Treisman, A. & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review 95*, 15–48.

Treisman, A. & Kahneman, D. (1983). The accumulation of information within object files. 24th Annual Meeting of the Psychonomic Society, San Diego, California.

Treisman, A., Kahneman, D., & Burkell, J. (1983). Perceptual objects and the cost of filtering. *Perception and psychophysics, 33*, 527–532.

Treisman, A. & Paterson, R. (1984). Emergent features, attention and object perception. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 12–31.

Treisman, A. & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14*, 107–141.

Treisman, A. & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General, 114*, 285–310.

Treisman, A., Sykes, M., & Gelade, G. (1977). Selective attention and stimulus integration. In S. Dornič (Ed.), *Attention and performance VI*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press.

Van Essen, D. C. & Maunsell, J. H. R. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neuroscience, 6*, 370–375.

Wolfe, J. M., Franzel, S. L., & Cave, K. R. (1988). Parallel visual search for conjunctions of color and form. *Journal of the Optical Society of America, 4*, 95.

Zeki, S. M. (1981). The mapping of visual functions in the cerebral cortex. In Y. Katsuki, R. Norgren, & M. Sato (Eds.), *Brain mechanisms of sensation*. New York: Wiley.

*Manuscript received 16 November 1987*

# APPENDIX 1

Since this lecture was given, Briand and Klein (1987) have also published a paper comparing the effect of spatial cues on feature and on conjunction identification. They distinguish the effects of "exogenous" and "endogenous" attention cues, equivalent to the "pull" and "push" cues described by Jonides (1981). Exogenous, "pull" cues are peripheral cues close to the target's future location, that automatically attract attention. Endogenous, "push" cues are central cues (such as an arrow pointing left or right) that inform the subject of the future location of the target without themselves sharing it. They presumably require deliberate movement of attention under voluntary control. Briand and Klein found that the feature and the conjunction conditions differed only when the exogenous, peripheral cue was used. They suggested therefore that only exogenously controlled attention is involved in feature integration. This is a very interesting conclusion. However, their data may not support it unequivocally. The stimuli were letter pairs presented to the left or right of fixation; the subject was to decide whether the pair presented on any given trial included the target "R" or not. In the conjunction condition, the distractors were "P" and "Q", which include the parts of an "R"; in the feature condition, they were "P" and "B", which lack the diagonal line of the "R". The problem in interpreting the results arises from the fact that the difference between endogenous and exogenous cues was confined to the feature condition with "PB" distractors rather than to the conjunction condition with "PQ" distractors. The "PB" stimuli showed very little effect of the exogenous cue, whereas with an endogenous cue the costs were significantly higher. The result would be consistent with the hypothesis that on "PB" trials, focused attention was unnecessary for detecting the target "R", and, if voluntarily focused in the wrong place rather than divided across the display, could be actually harmful. Subjects nevertheless attempted voluntarily to direct their attention in response to the endogenous cue, producing costs on invalid trials. Exogenous cues, on the other hand, could attract attention without costs on invalid feature trials, as attention is not needed for the detection of separable features and the exogenous cues are assumed to induce the attention shift automatically and without effort (Posner, 1978). Both cues affected performance equally in the "PQ" conjunction condition, as they should if focused attention is necessary for detecting conjunction targets.