I would like to thank Amanda Sowden and Ben Dyson for their comments on an earlier version of this article. Bill Prinzmetal very kindly redrew my attempt at his figure.

Correspondence may be addressed to: Philip T. Quinlan, Department of Psychology, University of York, York YO10 5DD United Kingdom. Electronic mail may be sent to ptq1@york.ac.uk.

An aim in writing this article was to describe the development of visual feature integration theory (FIT) since it was first put forward as a general framework for thinking about visual perception. Although it might be assumed that the theory is concerned only with the nature and deployment of visual attention, it is far more encompassing than this. FIT has influenced thinking on processes that range from the early stages of sensory encoding to higher order characteristics of attentional control. Indeed, it is no exaggeration to state that FIT was one of the most influential and important theories of visual information processing in the last quarter of the twentieth century. Despite this, and even though the theory evolved over the years, little has been written about the nature of this theory development. In tracing the history of the theory it is possible to see how thinking about many fundamental issues about attention and perception has also developed.

The preliminary sections of the article contain an attempt to provide a clear exposition of the early development of the theory. Basic experimental tasks are described, and it is shown how data from these tasks were used to support the theory. Because the intent is merely to describe the theory and the supporting data, a critical stance has not been adopted. As will become evident from the later sections in the article, the theory has been heavily criticized in almost all its respects, and inclusion of these criticisms alongside the passages of exposition may detract from conveying the original rationale and elegance of the
After setting out the theory and its supportive data, the discussion concentrates on the empirical evidence that stands in contrast to certain basic tenets of the theory. It is here that the major shortcomings of FIT are discussed. The important contradictory data are described, and it is shown how such evidence motivated changes to the theory. The intermediate sections conclude with a synopsis of the theory incorporating the many amendments that were, seemingly, demanded by the data.

Following the review of FIT and its development, the ensuing sections set out, in some detail, alternative accounts. These accounts predominantly have attempted to provide different explanations of performance in various—now well established—visual search tasks to those of FIT. Although such accounts have made important contributions to understanding visual search performance, none has the scope of FIT. The comprehensive nature of FIT is conveyed in the subsequent sections of the article that deal with alternative paradigms. Issues concerning the coding of features and their locations are considered next, and then the focus shifts to errors in perception relating to miscombinations of visual features. The article concludes with a summary of what basic facts have been uncovered because of FIT, and suggestions are made as to how outstanding issues may be resolved in the future.

**The Original Theory**

In its original conception, FIT (Treisman, Sykes, & Gelade, 1977; see also Treisman & Gelade, 1980) was based on the claim that visual perception may be characterized by two functionally independent and sequential processing stages. According to the theory, processing at the first stage is preattentive: All separable features are coded independently and in parallel using populations of feature detectors for such stimulus dimensions as color, size, and shape. Functionally, the assumption is that separate code the presence of particular features in the optic array through the use of corresponding sets of dedicated feature detectors. Feature registration is said to be preattentive and automatic and carried out in parallel across the entire optic array.

One impression created was of a scheme whereby the different feature maps comprised complete arrays of feature detectors such that there was one detector for each point in the visual array. Despite such an impression, the theory is reasonably vague about how the location of any given feature was to be specified at this stage of processing. Perhaps the clearest explanation was that provided by Treisman (1982), who claimed that the separate feature maps record the presence of stimulus features independently and that, within each map, a relational coding scheme is used to specify the positions of the features. Each map has its own independent coding system, and initially, there is no coordination of information across the different maps. A corollary of this is the notion of free-floating features (Treisman & Gelade, 1980). At the earliest stage of processing, because there is no coordination of information across the different feature maps, the positions of features from the same object need not correspond with one another in a coherent fashion. Thus, although the different maps are assumed to preserve the spatial relations present in the visual input, cross-referencing this positional information is assumed to require additional processes.

The basic idea is that, at the first stage, spatial information is not directly accessible in parallel across the different feature maps (Treisman, 1985, p. 165); rather, coordinating features across the different maps would typically require attention. For example, coordinating activation across the different maps could be
due to a serial scan of the different locations. Such a scan, operating in a more general coordinate system, would be able to recover both the identity and location of any feature in the visual array. Furthermore, it is hypothesized that identifying a feature would be a separate operation from locating it. The presence of a particular feature would be signaled by activation on its own particular feature map, but additional (attentional) operations would be called on to locate the feature in the optic array. As a consequence, it should be possible for an observer to signal the presence of a particular feature in the absence of being able to say where, in the visual array, the feature actually is.

Although this discussion suggests that each map contains a punctate representation of single features, additional claims are also made about primitive preattentive-grouping processes that operate at the level of the feature maps. For instance, texture segregation and figure-ground grouping are said to be preattentive, and hence, edges and boundary information would be made explicit at the feature map level. In other words, features within a given map could form into coherent clusters.

At the second stage in the model, cross-dimensional processing is assumed to take place. For example, a particular color cluster may be combined with a particular shape cluster to give rise to the percept of a particular colored shape. Indeed, the act of combining constituent features to form a conjunction of features is, in essence, the notion of feature integration central to the theory. In an early expression of the theory (Treisman, 1985, 1986a), such integration was assumed to take place with respect to a master map of locations (i.e., the more general coordinate system referred to before). The individual feature maps were said to be linked to this master map so that they could "activate" particular points on this map. Such points on the master map were then taken to reflect the presence of salient discontinuities in feature activity present on the corresponding feature maps. Activation on the master map was then said to represent such discontinuities without also representing what features these corresponded to in the input.

To recover an integrated percept of the input stimulus, an attentional serial scan of the master map would be needed. Through the application of focused attention, the actual feature information would be recovered via links between the master map and the separate feature maps. The notion of the serial application of focused attention is taken to be the operation of the "spotlight" of attention. Moreover, such scanning is due, in part, to a top-down operation in which attention could be directed to activated places on the master map, which, in turn, represented the points of salient feature activity (Treisman, 1985, p. 165). The percept of an integrated object is, therefore, in part due to the application of focused attention that operates to conjoin featural information expressed on the independent feature maps. In this way, the spotlight of attention provides the "glue" to stick features together (Quinlan & Humphreys, 1987, p. 455). Hence, the spotlight of attention is the feature-binding (integration) mechanism in the theory.

In actual fact, Treisman and Schmidt (1982) stipulated that there are three ways in which features may become conjoined. The first is through the application of focused attention (in the manner described above). The second is by predicting the presence of a particular object and by fitting the various constituent features into "predicted object frames" (Treisman & Schmidt, 1982, p. 111). For instance, the contextual constraints may be so great that particular feature combinations fit perfectly with current expectations (e.g., a London street location leads to expectations about both the size and the shape and color of a double-decker bus). Finally, features may be conjoined at random. Such conjunctions may arise because of the lack of focused attention or the lack of any particular expectations. Percepts so formed have come to be termed illusory conjunctions (Treisman & Schmidt, 1982) and, in the theory, arise because of the free-floating nature of features registered at the first stage of preattentive processing. At this initial stage, information across the feature maps is not framed within a single coordinate system, and consequently, mistakes in combining features may arise that are manifest as reports of illusory conjunctions. It is interesting, then, and perhaps surprising, that Treisman and Schmidt (1982) allowed for illusory conjunctions to arise "within the spotlight" (p. 112) of attention. Implicit here is the metaphor of a real spotlight in which the beam’s focus can vary in spatial extent. The attentional spotlight can be spread over a large or small area so as to fall on more than one or just one item, respectively. If several items fall within the spotlight, their constituent features may be interchanged. More generally, though, it is the features of objects outside the spotlight that are most likely to interchange. Also according to the theory, features within the spotlight of attention are unlikely to combine with features outside the spotlight of attention.
**Standard Visual Search**

The main empirical support for the overall notion of two stages of processing came initially from two versions of timed visual search tasks (Treisman, 1982; Treisman & Gelade, 1980; Treisman et al., 1977). On a given trial in such an experiment, participants are presented with a visual display comprising a number of to-be-searched elements. Participants are timed to make present-absent manual responses as to whether a predefined target was contained in the display. On a random half of the trials a single target would be present in the display alongside one or more nontarget elements. Although the physical size of the display remains constant over the experiment, display size is defined as the number of elements contained in the display. Display size varies over the trials because the number of to-be-searched elements changes in an unpredictable manner over the trials. The main data of interest are measures of reaction time (RT) and accuracy as a function of display size.

Within this general framework, two main tasks are defined: feature search and conjunction search. Description of one example helps clarify these tasks. In Experiment 1 of Treisman and Gelade (1980), the search elements were colored letters. In the feature search task the target features were the color blue and the letter (i.e., shape) S. The nontarget elements were green Xs and brown Ts. In fact there were four different tokens of two types of target; that is, two targets were defined by color (i.e., a blue X and a blue T) and two were defined by shape (i.e., a green S and a brown S). In the conjunction search the same nontargets were used, but the target was a single green T. In both search tasks the display contained either 1 to-be-searched element or 5, 15, or 30 such elements.

The data from Experiment 1 of Treisman and Gelade (1980) were discussed in terms of derived search functions—known as RT-display size search functions or, more simply, search functions—for the present and absent feature and conjunction trials. In such graphical representations, mean RTs for each display size across participants were plotted separately for each response type for the feature and conjunction search tasks. The first result to note was that RT was essentially unaffected by increases in display size for the feature-present displays. Such a result was taken as being strong empirical support for the notion of a parallel preattentive stage of processing in which distinctive features are registered automatically and therefore “pop out” from the display. Indeed, any such flat search function has come to be taken as the signature of featural pop-out (Wolfe, 1994). Performance on feature-absent trials, however, was not so clear cut in terms of supporting FIT because the search function did show a linear increase with display size. (From the perspective of FIT, it might have been expected that the absence of the critical target feature would have been picked up immediately after the failure of a parallel scan. Such issues as these have recently been thoroughly explored by Chun & Wolfe, 1996.) It is critical to note, though, that linear increases in RT with increases in display size were observed in the data for both conjunction-present and conjunction-absent trials. Moreover, the slope for the absent search function was approximately twice that for the present search function. In other words, the two functions gave rise to an approximate 2:1 ratio in their slope values.

Taken together, the linear increasing search functions and their 2:1 slope ratio were central to the argument about the operation of the feature-integration stage of processing. Treisman and Gelade (1980) took the linear increases in RT to be indicative of the serial scan by the spotlight of attention. The 2:1 slope ratio was taken to be a strong indication of the serial self-terminating nature of this scanning processing: Participants simply scanned each element in turn until (a) the target was found and a present response made or (b) the last element was inspected and found not to be a target, in which case an absent response could be made.

Clearly, the fit between visual search performance and theory appear to be almost exact. It is perhaps because of this, together with the comprehensive nature of the theory, that the work of Treisman and Gelade (1980) has had such a significant impact.

**Features and Early Notions Concerning the Nature of Feature Maps**

In an important sense, the coherence of FIT depends on being able to define visual features a priori. At the
time of the theory's inception, the criteria for conferring the status of "feature" on some visual attribute were taken from both neurophysiology and psychology. From neurophysiology was the evidence of special-purpose systems in different areas of the brain that appeared to be primarily dedicated to processing a single sort of visual attribute (see, e.g., Zeki, 1976). The list contained areas for processing orientation, color, spatial frequency, and movement. Psychological phenomena were also used to back up these claims. For instance, the empirical distinction between integral and separable characteristics, as advanced by Garner (1974), was also used to establish the functional independence between the associated processing systems. Evidence for separability (as gauged by the effective selection of one attribute while ignoring irrelevant variation of another) was particularly important in establishing independence of feature maps. Effortless texture segregation (Treisman, 1985) was also cited as being relevant to deciding about the preattentive pick-up of a particular featural characteristic; the immediate detection of a texture boundary, consistent with the change in a given attribute across the boundary, was used to inform about a corresponding putative featural change.

Other aspects of task performance were also brought to bear on defining a feature. For instance, Treisman and Gelade (1980) discussed the possibility that, according to FIT, participants should be able to detect a feature without also being able to identify it, if attention was otherwise distracted. Such phenomena could be used to operationalize the definition of features. In a similar vein, Treisman (1986b) later discussed the formation of illusory conjunctions as being a signature of features. The occurrence of illusory conjunctions fit well with the idea of an initial stage in which features are processed independently and that this takes place prior to a stage of feature integration. As she noted, "if no decomposition into properties or parts took place, it would be difficult to explain ... illusory conjunctions" (Treisman, 1986b, p. 30). Thus, to avoid circularity in the definition of perceptual features, Treisman (1986b) set out various lines of converging evidence that could be used to independently and collectively establish featurehood. Indeed, performance in the visual search task itself came to be used as a tool for providing converging evidence on defining visual features (Treisman, 1985). Whatever popped out in the feature search task joined the list of candidate primitive visual features. By 1990 the list included orientation, size, direction of motion, color, the presence of terminators, and closure (Cavanagh, Arquín, & Treisman, 1990). Caution is perhaps warranted though, because as later commentators have noted (Cheal & Lyon, 1992, p. 136), if any feature can be defined post hoc, the theory is, essentially, unfalsifiable (see Briand & Klein, 1989, for a similar point). Nevertheless, Wolfe (1998) concluded that some consensus had emerged because efficient visual search and texture segregation are now taken to be two jointly sufficient criteria that allow for the classification of certain visual attributes as basic visual features. Further discussion of perceptual features is deferred until later in the review because as the theory evolved so did opinion over how best to define features.

Nevertheless, some demonstrations of pop-out in the search task were taken to be both problematic for FIT and for the claim that pop-out could be taken as an indication of primitive visual features. For instance, Bravo and Blake (1990) designed displays comprising small, oriented line elements. The background elements were all oriented horizontally, and patches of vertical elements were then embedded in the background. The overall orientation of the individual patches varied such that on target-present trials one of the patches was set at a different orientation to the rest.
Figure 1 - Example of a target-present display used by *Perception, 19*, p. 517.

Figure 1 provides an example of the sort of search display used in the study. Search performance was assessed as a function of the number patches present (i.e., display size), and the standard parallel search function obtained when the data were plotted accordingly: RTs were both fast and unaffected by display size. Bravo and Blake went on to argue that this particular pattern of performance was difficult to square with the notion of simple primitive visual features because the pop-out reflected preattentive processing of perceptual (texture) groups. By their view, this put the notion of featural pop-out under considerable strain; however, as was noted above, the original theory did allow for some early preattentive grouping and texture segmentation at the level of the individual feature maps.

To be clear on this point, some of the earliest data related to search performance with displays containing salient groups of items—for instance, in *Treisman (1982)*, the to-be-searched elements were grouped into checkerboard configurations by shared color. Performance with the grouped and (normal) random displays was compared, and the results were taken to inform about the nature of the search process as a function of the search task. (In this sense the data reported by *Bravo & Blake, 1990*, are not critically at odds with the theory.) The random displays contained 1, 4, 14, or 36 items. In contrast, the grouped displays each contained 36 items, comprising a single group of items or 4, 9, or 18 subgroups of items. Under these conditions feature search performance was essentially unaffected by the grouping manipulation. Feature-present searches were fast, and search functions were flat for both grouped and random conditions. Grouping did, however, produce a slightly surprising effect for conjunction searches. The standard pattern was observed for random displays, but nonlinearly increasing search functions were observed for the grouped displays. Responses for both target-present and target-absent trials described negatively accelerating functions, with slower responses accruing on target-absent trials.

Such a pattern was taken to show that, with the grouped displays, participants were typically not searching every item in turn; rather, performance appeared to reflect two components: a serial scan across the different subgroups of elements followed by a parallel check within a group. The presence of a particular conjunction target was registered only when attention had been focused on its containing subgroup. Performance with the grouped displays was taken to reflect preattentive grouping taking place within, but not between, the different feature maps. The coordination of such grouped representations across the different feature maps came about only through the application of focused attention.

**Practice Effects**

Given that the distinction between feature and conjunction search performance was taken to map onto the operation of two distinct functional modules—an early preattentive stage and a later attentional stage—a concern was with the degree to which search performance could be modulated by practice. The rationale for such studies was to explore the degree to which the differences between feature and conjunction search performance truly reflected qualitative differences in processing at a functional level. According to FIT, even though practice might induce an overall shortening of RTs, the qualitative differences between feature and conjunction search performance should remain unchanged. *Treisman and Gelade (1980)* provided the first relevant data by having participants repeatedly carry out conjunction searches. The data were presented in terms of slope and intercept values of the individual present and absent search functions over the number of blocks of trials completed. The critical finding was that, whereas overall RTs did decrease with practice—
as revealed by a diminution in the size of the intercept values over time—the slope values hardly varied at all. Treisman and Gelade used this pattern to argue that there was no evidence to suggest a switch from serial to parallel processing over the course of practice on conjunction search. Therefore, conjunction search appeared to reflect operations defined with respect to hard-wired independent feature-processing systems that were, essentially, immutable.

Issues concerning practice at conjunction search were revisited by Steinman (1987), who examined performance from various conditions after approximately 10,000 trials. An apparently contrasting pattern of results to those described by Treisman and Gelade (1980) emerged. In Steinman’s experiments, the to-be-searched elements comprised small line segments, and manipulations were introduced to test target detection based on the pick-up of featural vernier offsets, stereoscopic disparity, line orientation, lateral separation, and the various combinations of these features. In all cases of conjunction search tested, performance started out as describing linearly increasing search functions, consistent with the original predictions of FIT. However, at the end of testing all RTs were fast, and search functions were flat. Hence the contrast with the data reported by Treisman and Gelade. Steinman made the interesting argument, though, that this qualitative shift in performance may have arisen because participants learned how to segregate displays on the basis of one of the features used to define the conjunction target. By narrowing down the search to elements possessing one salient feature, pop-out could then take place on the corresponding feature map. Such a strategy, however, only came about through becoming highly familiarized with the relevant features.

Very similar ideas can be found in the earlier article by Egeth, Virzi, and Garbart (1984). They put forward data consistent with the view that participants could effectively screen out half the nontargets in conjunction search displays on the basis of one particular feature and, in effect, confine search to the elements that possessed the other feature. For a recent and more thorough examination of this aspect of performance, see Kaptein, Theeuwes, and van der Heijden (1995).

A different approach to addressing the consequences of practice also relates to conjunction search. Treisman, Vieira, and Hayes (1992) reported a study in which participants were highly practiced at searching for a target composed of connected line segments against similar nontarget items. The issue was whether such training with such targets gave rise to the formation of an "integrated wholistic representation for these ... arbitrary conjunctions of lines" (Treisman et al., 1992, p. 345). This was tested by assessing performance with the practiced targets across a range of other tasks that were assumed to tap into participants’ particular memory representations. The basic issue was whether the initially arbitrary combination of features had become unitized in memory to serve, in a sense, as a new feature detector. However, this was not upheld by the data, and whatever had been acquired during practice failed to support generalization in the other tasks.

In summary, the data concerning practice effects in visual search are generally consistent with the distinction between independent feature and feature-binding stages. Whereas practice might induce particularly efficient search strategies to develop, there is little evidence to suggest that attentional processes invoked by conjunction processing can be circumvented through a process of task familiarization. It remains possible though that particular feature detectors may develop as a function of practice (see Rabbitt, 1967).

**Feature Discriminability**

Having developed the view that visual search performance could provide critical supportive evidence for FIT, Treisman and Gelade (1980) did go to some lengths to examine the boundary conditions on featural pop-out. They wished to rule out any simple account of the feature-conjunction search differences based on contingent differences in feature salience. To this end they were concerned with notions of feature discriminability and designed search conditions in which target-nontarget (T/NT) similarity was manipulated. (As I later make clear, different concerns about element discriminability have plagued the theory since its inception. Indeed, it has been interesting to follow how the theory has developed so as to take account of such concerns while also attempting to hold onto the qualitative difference between feature and conjunction processing. ) In the critical experiment, all of the to-be searched elements were ellipses that varied in
size—they were referred to as being **unidimensional**—and for ease of exposition, increasing size is denoted by the range of numbers 1-5 (see

![A. Stimulus Set:](image1)

![B. Intermediate Target Condition](image2)

![C. Disjunctive Target Condition](image3)

**Figure 2** - A: Examples of the sorts of search elements used by *Cognitive Psychology*, 12, A. M. Treisman and G. Gelade, "A Feature-Integration Theory of Attention," 1980, p. 110.

Figure 2 for further clarification). Two different conditions were run, and in both cases Shapes 2 and 4 were the nontargets. In the intermediate target condition the participants searched for Shape 3, and in the disjunctive target condition participants searched for either Shape 1 or Shape 5. In terms of similarity, therefore, searching for Shape 3 (the intermediate target) ought to have been the most difficult because it was equally similar to both of the nontargets. In contrast, searching for Shapes 1 (the small target) and 5 (the large target) should have been correspondingly easier because these items were most similar to only one of the nontargets—Shapes 2 and 4, respectively.

Data from the two conditions were broken down according to target type and response type (i.e., target present and target absent). There clearly was an effect of display size on search times in all cases, yet all of the search functions were nonlinear and negatively accelerating. This was taken to show that in neither condition did performance conform to the standard conjunction search pattern. **Treisman and Gelade's**
general conclusion was that, even though feature search might have varied in difficulty, qualitative differences still existed between feature search and conjunction search. Treisman and Gelade therefore asserted that the differences in search performance with feature and conjunction targets reflected something more than T/NT discriminability despite the fact that such a factor may well have affected search performance. In sum, support for the theory had been bolstered by (a) the pattern of differences between feature and conjunction searches, (b) the fact that such differences remained constant after extended practice, and (c) the fact that such differences could not be explained purely in terms of differences in target discriminability across the two search conditions.

**Search Asymmetries**

The modulation of feature search performance shown by Treisman and Gelade (1980) allowed to assert boldly that "the ease with which we can find a target in a display of distractors depends on their discriminability" (p. 285). Despite such assurances though, it was also clear that search performance could not be determined solely by interelement discriminability because of the repeated demonstrations of certain feature search asymmetries. It is important to discuss these findings in some detail because the discovery of search asymmetries resulted in important developments to FIT.

In demonstrations of search asymmetries, a set of to-be-searched elements are defined, and for separate conditions, two types of targets are also defined. In a target+ condition the target is given an additional feature not contained in any of the nontargets. In the target- condition a critical feature is removed from one (target) element, and this feature is retained in all other nontarget elements. Thus, for instance, a target+ condition would be one in which the target is a and the nontargets are s. In the target- condition the mapping is reversed such that now is the target and s are the nontargets. The critical feature in this example is the presence of the small line segment in the .

The central finding of such paradigms is that search is far easier in the target+ condition than in the target- condition; that is, there is a feature search asymmetry (see Treisman, 1985). In fact, Treisman and Souther (1985) showed that search was fast and the search function was flat when the target had a unique distinguishing feature, as with the Q in Os condition. Search appeared to be serial and self-terminating when the target was signaled by the absence of a particular feature, as with the O in Qs condition. Treisman and Souther took this pattern of data as additional support for the idea that an observer may be able to detect the presence of a primitive feature in the absence of being able to locate it. In contrast, search for the absence of feature demanded a serial scan of each of the elements. Such a scan was taken to reflect the operation of the spotlight of attention pointing to each element in turn so that the otherwise free-floating features may be bound together. The scan would be needed to determine whether a particular circle was conjoined with a particular line segment in the Q.

Central to this account is an argument about the nature of the different feature maps within a particular domain. For example, FIT claims that color processing depends on a map for each of the different hues. Any one of these constituent maps is activated by the corresponding feature in the input and, in turn, represents both the presence and amount of activity. Given this, it is accepted that a featural difference between a target and corresponding nontargets need not reflect a categorical distinction (e. g., blue vs. not blue). According to Treisman and Souther (1985), featural pick-up could be due to a preattentive mechanism that was sensitive to the amount of featural activity, that is, pooled activity, within a feature map and not just the presence of such activity. As before, though, search depends on a serial scan of the master map of locations. The attentional spotlight traverses this map such that, at a given position of the spotlight, all of the featural information could be recovered from the respective feature maps.

To explain the basic search asymmetries, FIT posits that the efficacy of preattentive pick-up is dependent on the discriminability of the target from the nontargets. By way of example, assume that a target registers more activity on a particular feature map than any given nontarget—that a given target registers five units of activity and each nontarget registers one unit. For small display sizes, the difference in activity produced for a target-present display as compared with a target-absent display would be large, with more activity being produced for the present than absent case. However, as display size increases (to more than eight items), proportionally less activity is produced by the present than absent displays. As a consequence, the target
would be more difficult to find in large displays than in small displays, and participants would be forced to parse the large displays into smaller subgroups of elements and consider each of these subgroups in turn. Such a strategy would explain increasing RTs for difficult feature searches as display size increases.

The notion of pooling activity across a particular feature map also can be extended to account for the slower searches for the conditions in which the target was defined relative to a missing feature—as in the target-searches. In such a case, the target would register less activity than a given nontarget; hence, participants would be forced to parse the displays into smaller subgroups than they would for the target+ searches to detect the target. By this view, the search asymmetry arises because of (a) differential amounts of feature activity produced by the targets and nontargets in the target+ and target- conditions, respectively, and (b) the concomitant change in search strategy used in the two conditions.

Further clarification of these ideas was later provided by Treisman and Gormican (1988), and again, the notion of pooling activity from a given feature map was discussed at length. The major extensions to the ideas concerned the manner in which such pooling might take place. In its original conception, the pooling of activity involved summing activity across the representation of the different display elements on the relevant feature map, with the proviso that any given target produces more activation than any corresponding nontarget. Treisman and Gormican appeared to favor an account based on averaging activation across a given feature map. As before, each point on the map reflects the activity of a given feature detector, but in the amended account, the notion of mutual inhibition between different feature detectors is used to explain the averaging process. Different features within a given map could now mutually inhibit one another, and as the number of active detectors increased, each would receive a corresponding increasing number of inhibitory inputs. Such an averaging process allows the search mechanism to distinguish displays containing a few intense points of feature activity from displays containing more points of lesser activity. The interesting claim, though, is that, whereas the feature maps capture the presence and average activity of pertinent features, the number and exact location of these features is recoverable only from the master map of locations. As in the original theory, locating features in the optic array is still assumed to be an effortful process. Clearly, though, the additional assumptions about the underlying architecture and contingent processes implicated by featural search are much more involved than those contained in the original theory.

In testing these ideas, Treisman and Gormican (1988) reported results from very many different types of visual search experiments, and search asymmetries were reported for a wide range of stimulus types. Discussion of only two examples suffices here. In the first, participants had to search for an easy and a difficult feature target in different conditions, respectively. All the nontargets were small line segments of a given length. There was a very noticeable difference in length between the target and nontargets in the easy condition but only a subtle difference in this length in the difficult condition. Two general conditions were tested for each of these targets: one in which the target was longer than the nontargets and one in which the target was shorter. For both the easy and difficult conditions, search asymmetries were apparent: Search was more efficient when the target was longer than the nontargets. Moreover, feature discriminability clearly had an effect because the slopes of the respective search functions differed across the easy and difficult conditions. Slopes were considerably steeper in the difficult condition than in the easy condition.

Such a pattern of findings had been predicted by Treisman and Gormican (1988) on the grounds that a feature target ought to be easier to detect if it produced relatively more activation on its associated map than did the nontargets. Treisman and Gormican’s account was slightly more subtle than this, though. Although long targets were easier to detect than short targets, this result was not, fundamentally, about the absolute length of an element per se, but rather the relative difference in length between the target and nontargets. This was shown most clearly in the matched conditions. In matched conditions, only the length of the target changed across the longer and shorter conditions (i.e., 10 mm and 5 mm, respectively)—the length of the nontarget lines was the same in both conditions (7.5 mm). Under these conditions the search asymmetry disappeared. The important conclusion was that feature search can be critically affected by the difference in similarity between targets and nontargets as gauged by the amount of activation produced on a given target feature map relative to that produced on a nontarget feature map. This is a very clear statement of how feature search may be critically dependent on the discriminability of the different types of to-be-searched elements.
Treisman and Gormican (1988) then carried forward this line of argument by attempting to test two possibilities regarding the representation of "standard" and "deviant" feature values. Either a standard feature was coded uniquely, with deviations from the standard generating reduced levels of activation, or deviations were coded positively, with the standard taking a default value. To help adjudicate between these possibilities, Treisman and Gormican ran two further search tasks. In the first condition participants had to search for a circle in the presence of ellipses (cf. Treisman & Gelade, 1980), and in the second, the T/NT mapping was reversed, with participants now searching for an ellipse in the presence of circles. A priori the circle was defined as the prototype and the ellipses were defined as deviations from this prototype. The central finding was of a search asymmetry for cases in which search for an ellipse target was easier than search for a circle target. In neither case, however, did the target pop out.

interpreted this pattern of results as supporting their claim that standard features "are coded as the absence of activity on the deviating dimensions ... [and] that they are positively coded on their own channels, with the proviso that the deviating stimuli also produce substantial activity in the prototype channel" (pp. 30-31). The use of the term channel here was undefined. Nevertheless, the circle-ellipses findings were explained in terms of ellipses activating both the circularity detectors and the elongation detectors, whereas the circle primarily activated the circularity detectors and produced hardly any activity in the elongation detectors.

Notable Amendments to the Theory

By 1988 a reasonably complicated view of feature search had emerged. Apart from the standard pop-out effect, which only occurs if the critical feature is unique to the target, other types of feature search could arise. For instance, if there is no categorical featural difference between the target and nontargets and the difference between target and nontargets is with respect to degree rather than kind, then search most likely would be serial. In contrast, if the target possesses more of the critical feature than the nontargets, then the corresponding search function would be flatter. Variation in feature search was explained (Treisman & Souther, 1985) with recourse to the notions of feature activity on feature maps and processes of averaging such activity to make a response. Treisman and Gormican (1988) conceived of the idea of averaging feature activity in terms of mutual inhibition spreading between feature detectors within a given feature map. Finally, the slope of a feature search function would also be flatter when the target is defined as a deviation from some prototype value than when the target conforms to a prototype. Again, some attempt at fitting these findings with FIT had been attempted (see Search Asymmetries section), but the post hoc arguments were far from compelling.

Perhaps the most significant modification to the theory arose because of the realization that both conjunction search and some feature searches appeared to depend on an attentional serial scan of the input (or rather, its internal representation). For difficult feature searches, the idea was that large displays would have to be parsed into smaller subgroups and then attention could be directed to each of these subgroups in turn—this encapsulates what Treisman and Gormican (1988) referred to as the group-scanning hypothesis. In this respect, Treisman and Gormican had begun to question the original fundamental distinction between preattentive and attentive processing. Given the admission that both feature and conjunction searches could be predicated on attentional processes, they now argued that differences in search performance reflected differences in the degree to which the attentional spotlight could be focused. Feature searches typically depended on attention being focused across the whole display, and activation pooling across a given feature map was critical in target detection. In contrast, conjunction searches depended on focusing the spotlight on subgroups of elements (and, at the limit, each one of the display elements) in turn. Variations in the size of the attentional spotlight accounted for variations in search performance between the two extremes of pop-out and serial, self-terminating search. Now, according to the amended theory, the size of the spread of the spotlight indexed the accuracy of localization of a given target and the accuracy with which the different features of a given conjunction would be bound.

Clearly, with the publication of the Treisman and Gormican (1988) article, many central properties of the original FIT had been revised (e. g., the notion of featural pick-up) or even discarded (e. g., the strict adherence to feature and conjunction searches mapping onto the operation of preattentive and attentive
processes, respectively). Perhaps the most contentious modification to the general framework, however, came with the publication of an article by Treisman (1988). Until this time, most would have been seduced into believing that there had been only one version of FIT in which the initial stage of featural processing involved the preattentive registration of feature activation on the separate feature maps as has been described here. Associated with this view was the idea that outputs from this initial stage fed into a master map of locations where it was possible to recover both the nature and location of composite objects by means of an attentional scan. However, according to Treisman (1988), there had actually been two versions of the theory present in the literature: Version 1, that is, the version as has just been described and a radically different version (Version 2), in which the initial registration of the visual input was with respect to a master map of locations in which "different dimensions are initially conjoined in a single representation before being separately analysed, dimension by dimension" (Treisman, 1988, p. 204; see also Wolfe & Bennett, 1997, for further support for such an idea). According to Treisman (1988), this latter framework had been presented in the much earlier article (Treisman, 1985), but in revisiting this work it is difficult to gain such an impression. Nevertheless, in Version 2 of the theory an early master map of locations is posited, and it is the outputs from this stage that propagate forward for analysis on separate feature maps. The view was of an early location-addressable but not content-addressable representation which was subsequently analyzed into separate feature maps (see Treisman, 1990a, p. 455).

This account allows for different dimensions' being initially conjoined (in the master map) prior to analytic processes operating on the separate dimensions (via the different feature maps). According to Treisman (1988), the final level of perceptual coding—as a precursor to object recognition—is one in which the different input features are recombined into a form of temporary object representation known as an object file. It is at the level of the object file that conjoined features are compared against stored descriptions of real-world objects. In very general terms, therefore, the theory now posits (a) initial conjunction representations that are not "directly interpreted" (Treisman, 1988, p. 204), or alternatively, that are not open to conscious awareness (these are registered on the master map), (b) a stage in which the input is decomposed into its constituent features (on the respective feature maps), and finally, (c) a stage of recombination in which feature conjunctions are established through feature integration and are represented within the current object file. An implication here is that these later conjunction representations support conscious perception.

More radical though was the proposal that attentional spotlight operates on the early master map prior to feature analysis! Regardless of the evolution of such a framework it is important to ask why such a radical option (i.e., Version 2) was favored by Treisman (1988). The answer lies in the desire to incorporate findings that contradicted Version 1 of the theory. Discussion of two notable examples should suffice.

Houck and Hoffman (1986) The first example is the work of Houck and Hoffman (1986). Central to this study was the McCollough effect. The standard effect arises after viewing two alternating horizontal and vertical gratings of the same spatial frequency. Such a dynamic display is known here as a McCollough stimulus. Consider a case in which Grating 1 (the vertical grating) contains green and black columns and Grating 2 (the horizontal grating) contains red and black rows. After an adaptation phase in which observers are presented with the alternating gratings, they then view a stationary black-white grating of the same spatial frequency as the originals. Participants then report different impressions depending on whether they view an horizontal or a vertical test grating. In the horizontal case, the white rows now appear reddish (given that the participant has been adapted to green columns and red rows), and in the vertical case the white columns appear greenish (given that the participant has been adapted to red columns-green rows). The critical point about this aftereffect is that it is orientation-color specific—that is, it is defined relative to a particular orientation-color conjunction. Given Version 1 of FIT, such conjunctions can only systematically arise because of the application of attention during adaptation.

Houck and Hoffman (1986) therefore examined this aftereffect under various attentional conditions. Each condition comprised an adaptation phase and a test phase.
Figure 3 - Example of the sorts of display used by *Journal of Experimental Psychology: Human Perception and Performance*, 12, p. 188,

Figure 3 provides a schematic representation of the type of displays used in the experiment. During the adaptation phase, the central display element was a McCollough stimulus and was encompassed by other elements situated in the periphery. The element positions were equally spaced and can be numbered from 1 starting at the 12 o’clock position (see Figure 3). For displays of nine elements, the central element was encircled by eight other McCollough stimuli. For displays of five elements, only four peripheral McCollough stimuli were presented at Positions 2, 4, 6, and 8; the other positions remained empty (Figure 3 provides an example of a five-element display). Display set size was introduced as a between-groups variable, and a task variable was introduced in addition. Participants were asked to monitor for possible changes to patterns superimposed on the gratings.

In the central task the pattern was a grid of $3 \times 3$ dots, and the task was to respond if any dot was momentarily removed. In the peripheral task each superimposed pattern was a left bracket, and the task was to respond if one of these was replaced by a right bracket. Four such superimposed brackets were presented on the gratings, respectively, at Positions 2, 4, 6, and 8. These positions were known as the *peripheral-relevant positions*. Positions 1, 3, 5, and 7 were known as *peripheral-irrelevant positions* because no imperative stimulus occurred at these positions. Participants in different groups enacted either the central task, the peripheral task, or both tasks.

Houck and Hoffman (1986) assessed the strength of the aftereffects at all display positions during the test phase. Baseline performance was assessed by measuring the strength of the aftereffect for the peripheral-irrelevant positions for the displays containing five elements. Essentially no McCollough stimuli had been presented at these positions during adaptation, and indeed, no aftereffect was found for these positions. The critical findings were that (a) the strength of the aftereffects were the same for the center and peripheral-relevant positions and (b) these effects were unaffected by the manipulation of attention. That is, regardless of attentional instructions, the strength of the aftereffects were the same for all of the adaptation positions. This shows that, during the adaptation phase, particular orientation-color conjunctions had been
registered even in cases in which attention had been directed away from the actual locations of these conjunctions.

Such a conclusion therefore stands in complete contrast to a central tenet of Version 1 of FIT—namely, that conjunctions arise because of the earlier registration of their constituent features which are then bound together through the operation of attention. It was this finding that had a profound effect on the development of FIT because it prompted Treisman (1988) to abandon Version 1 of the theory and accept Version 2. Despite this though, both Houck and Hoffman (1986) and Treisman (1988) did speculate that the McCollough effect might reflect very early (retinal) processes that precede any of the processes alluded to in the model. Indeed, in a later article, Briand and Klein (1989) even went so far as to suggest that the data might be less of a threat to the theory because "there is no need to equate the channels carrying the spatial frequency and color information with those that act as inputs to the object-building process" (p. 405). Nevertheless, the Houck and Hoffman article clearly did have an important influence on the development of FIT simply because of the demonstration that certain conjunctions could arise in the absence of focused attention.

Nakayama and Silverman (1986) The second notable set of data that Treisman (1988) discussed, as far as motivating changes to Version 1 of the theory, related to demonstrations in visual search tasks of pop-out for what otherwise might be deemed conjunctions. A classic example is provided by the work of Nakayama and Silverman (1986). In that study, (just 2) participants searched through displays containing various colored elements comprising small random checkerboards. The elements could be stationary or move upward or downward across the display. Moreover, all the elements could be presented in one depth plane or divided between two depth planes. Various combinations of these presentation conditions were tested. For the motion and color conditions all elements were presented in a single depth plane and RTs were both fast and unaffected by display size. When participants also searched for a standard conjunction of color and direction of motion, a linearly increasing search function obtained.

Stereo depth was introduced in an additional two conditions. In the stereo-motion (SM) condition (in which conjunctions of stereo depth and motion were tested), half of the nontargets were presented in one depth plane and consistently moved in one direction. The other nontargets were presented in the other depth plane, and all consistently moved in the opposite direction. The target element moved in the direction of motion associated with the elements in the other depth plane. In the stereo-color (SC) condition (in which conjunctions of stereo depth and color were tested), the nontargets were split between different depth planes. Nontargets within a depth plane shared a common color, but nontargets across the two depths differed in color. The target was present in the color associated with the other depth plane.

The (perhaps) surprising result was that the search functions for both the SC and SM conditions were essentially flat—a result taken to be consistent with parallel search for the respective conjunctions. Nakayama and Silverman (1986) interpreted the data as contradicting a central tenet of FIT—namely, that conjunction targets in visual search can only be recovered through a serial scan of the display. This contrast between data and theory therefore motivated some rethinking of the original version of the theory (Treisman, 1988). Upon further inspection however, even though the searches for both participants were clearly unaffected by display size, the mean RTs in these conditions were reasonably slow. For the SC conditions the RTs were slightly less than 1 s, whereas for the SM conditions the RTs were over 1 s. It is therefore difficult to accept these data as providing compelling evidence of pop-out of the SM and SC conjunctions given that most examples of pop-out are indexed by much faster RTs (see, e. g., the range of cases considerably less than 1 s described by Cheal & Lyon, 1992). 7

Although the work of Nakayama and Silverman (1986) has been discussed in some detail and problems in the interpretation of results have been raised, the critical point is that Version 1 of FIT remains challenged by any data that reveal featural pop-out for what otherwise might be deemed conjunctions. Over the years many such examples have been described in the literature (see Dehaene, 1989; Humphreys, Quinlan, & Riddoch, 1989; McLeod, Driver, & Crisp, 1988; Nakayama, 1990; Steinman, 1987; Wolfe, Cave, & Franzel, 1989). Nevertheless, the work of Nakayama and Silverman had an important impact on the development of FIT.
Further issues arising over conjunction processing.

The work of Houck and Hoffman (1986) and that of Nakayama and Silverman (1986) provided early demonstrations of particularly easy conjunction search. Both of these demonstrations sat uncomfortably with FIT, and because of this, Treisman (1988) went on to provide different accounts as to why easy conjunction search was observed in the different studies. For instance, Treisman (1988) cited Houck and Hoffman's data as providing compelling evidence of early conjunctions of color and orientation. In Version 2 of the theory, the claim is that such conjunctions arise at a very early stage in processing prior to "analysis of separable features by specialized modules" (Treisman, 1988, p. 223); that is, the master map of locations is now situated prior to the independent feature maps. Some further discussion of relevant data and theoretical issues is warranted, however.

Early conjunctions.

Further evidence for the very early detection of particular color-orientation conjunctions has been bolstered by the recent data reported by Holcombe and Cavanagh (2001). In their research, participants viewed displays comprising two rapidly alternating displays. In the spatially superimposed condition, one semicircular patch of a colored grating was alternated with another grating of a different color and orientation than the first (e.g., red right-diagonal bars interchanged with green left-diagonal bars). The gratings appeared at the same screen location, and each stimulus comprised a different color-orientation conjunction. In contrast, in the spatially separate condition, each display comprised two semicircular patches arrayed one above the other. The top patch was filled with color, and the bottom patch contained an achromatic grating. Thus, a red patch was paired with (spatially separate) right-diagonal bars, and a green patch was paired with left-diagonal bars.

In these conditions, the rate of alternation of the stimuli varied, and the data of critical interest were the 75% threshold alternation rates at which participants could report both of the color-orientation pairs. The central finding was that thresholds were considerably lower in the spatially superimposed than the spatially separate condition. That is, at very high rates of alternation (i.e., when each frame was present for approximately only 14 ms), participants gained a much clearer impression of the color and the orientation of the bars than they did in the case of the spatially separate patches of colors and bars. (In similar tasks, with brightness and orientation as the defining dimensions, Holcombe & Cavanagh, 2001, found comparable results: Thresholds were much lower in the superimposed conditions than in the spatially separate conditions.) Holcombe and Cavanagh (2001) took these findings to be consistent with the idea that pairings of color and orientation (and separately, brightness and orientation) "are coded in combination explicitly by early stages" (p. 127), a view that is, essentially, the same as Treisman's (1988) idea of early conjunctions.

Different studies therefore collectively provide support for the notion of the early registration of color-orientation conjunctions. Moreover, D. J. Cohen (1997) also has argued that similar early conjunctions may exist for form and color. In his experiments, participants viewed very briefly presented and masked displays containing a matrix of colored shapes. In the form-alone condition all the elements shared the same color, but there was a unique target shape. In the color-alone condition the elements were all of the same shape, but the target was presented in a unique color. In the coincident condition there were homogeneous nontargets, but now the target was unique with respect to both its shape and color. In the disparate condition two targets were present: a distinctive color target together with a separate distinctive shape target. Finally, an homogeneous display was used on target-absent trials. On each trial participants simply had to respond as to whether a target was present. Measures of report accuracy were of main interest.

The most important result was that even at very brief durations (i.e., 33 ms) participants were more accurate on coincident trials—those in which a color-form conjunction was present in the display—than on disparate trials—those in which separate target values of color and form were present. Moreover, the responses from the coincident trials could not have been predicted from assuming independence of processing of color and form. Thus, D. J. Cohen (1997) concluded that "color and form are not perceptually independent at the very early stages of perceptual analysis" (p. 630). Such a result is very difficult to square with Treisman's (1988) initial ideas about independence of processing but is much more in keeping with the
later idea of early conjunctions.

**Feature inhibition.**

Treisman (1988) was at pains to point out that the notion of early conjunctions that had been used to explain Houck and Hoffman's (1986) data probably would not suffice to explain the easy detection of the particular SM and SC conjunctions studied by Nakayama and Silverman (1986). A much more thorough examination of these issues, though, was reported by Treisman and Sato (1990). They set out three plausible hypotheses as to why conjunction search might be so efficient. The first was the idea that special grouping processes might operate to segregate out the two different sorts of nontargets. Following this act of segregation a feature search could then be invoked for only one of these groups. The second was that the effect reflected the operation of special-purpose conjunction detectors located at some early point in the visual system—which is the early conjunction idea discussed before. Finally, the third was, essentially, the notion of feature inhibition—namely, that nontarget locations could be suppressed because of the presence of features that were inconsistent with the target.

A critical test was reported in Treisman and Sato's (1990) Experiment 2, in which they examined search performance across a range of features and associated conjunction conditions. Values of size, color, motion, and orientation were tested in feature search tasks, and the six possible conjunctions of these four features were tested in corresponding conjunction search tasks. For all feature searches the search functions were flat, and RTs were relatively fast. Conjunction search functions all showed linear increases with display size, but the slopes of the different functions varied considerably. It was this variation in the slope values that was critical in adjudicating between the three different alternative accounts of performance described above.

According to the conjunction-detector idea, neurophysiological data would suggest the existence of hard-wired analyzers for certain combinations of features such as orientation and motion (De Valois, Albrecht, & Thorell, 1982) but not for particular combinations of color and motion (Hubel & Livingstone, 1987). Performance with the two corresponding types of conjunctions did not, however, support such a dichotomy in Treisman and Sato's (1990) experiment. There was no obvious difference in performance (i.e., search rates) for conjunctions that might plausibly have detectors and those that might not. In this regard the data had failed to provide further support for the notion of early conjunction detectors (discussed before), even though the authors had advanced such an explanation to account for the results reported by Houck and Hoffman (1986).

Treisman and Sato's (1990) data also contradicted a prediction of the segregation hypothesis. According to this hypothesis, search would be facilitated by the presence of a particularly salient feature so that participants could screen out elements possessing this feature and limit search to the remaining elements. For instance, a salient color feature might operate as such a cue regardless of which other feature it was conjoined with. The results provided no support for this, however: There was no such systematic pattern across the various conjunction search conditions. No such subset of salient features was uncovered.

The most telling result was that conjunction slopes appeared to reflect an additive contribution of the rate at which each of the constituent features could be checked. The implication here was that the slope of the conjunction search function reflected, in part, a process of serially checking the constituent features of the target against each search element. Treisman and Sato (1990) went on to argue that the data were most consistent with the feature-inhibition idea. Registration of any nontarget feature, accordingly, resulted in suppression of activity in the master map of locations from the corresponding nontarget feature maps (cf. Treisman, 1988). In this way, the representation of nontargets may be jointly suppressed by inhibition stemming from more than one feature map onto the master map. Although such an account fit reasonably comfortably with the data, the consequences for the theory were critical in other respects. Now conjunction search could be characterized as a product of both feature inhibition and a serial scan of the master map of locations (Treisman, 1990a; Treisman & Sato, 1990). (Although Treisman & Sato, 1990, discussed the notion of feature inhibition in some depth, they did acknowledge that their ideas differed only at the level of very specific details to those made more explicit in the guided search model put forward by Wolfe et al., 1989.)
Overall, therefore, the early registration of conjunctions is allowed by Version 2 of FIT in that it brings forward the master map of locations prior to the stage of a feature analysis. The model embodies a master map of locations "that registers the locations of regions without giving access to the features that define them" (Treisman, 1999, p. 93). This implied mechanism can apparently register the position of a visual event in the absence of being able to determine the nature of that event—contrary to the predictions of Version 1 of FIT. In Version 2 of FIT an attentional scan is allowed to take place on the master map of locations, and such a scan can be directed either by salient external events—such as a cue in a visual cuing experiment—or by top-down inhibition from the feature maps. Thus, not only can feature detectors within a given map mutually inhibit one another, but these also may in turn inhibit locations on the master map. As in Version 1 of the theory, the purpose of the serial scan is to recover and combine all featural information from the relevant feature maps. However, the central point is that conjunction search could be facilitated either because of salient properties of the input and/or by featural inhibition of the nontargets.

**Attentional Effects in Feature Detection Tasks**

Various demonstrations of so-called easy conjunction search had been used by researchers to question the necessity for attentional processes in conjunction detection posited in Version 1 of FIT. Version 1 also struggles to accommodate findings suggesting a role for attention in simple feature tasks. Although Treisman (1992) generally cited her own data (i.e., the difficult feature searches reported by Treisman & Gormican, 1988) as suggesting a role for attention in some feature search tasks, other evidence did exist at the time to suggest some form of attentional involvement in featural processing. A central reference here is that of research performed by Prinzmetal, Presti, and Posner (1986).

In their first experiment, Prinzmetal, Presti, and Posner (1986) designed a cuing task in which, on each trial, participants were initially presented with a colored letter that defined the target for that trial. This was replaced by a fixation point, and there then followed a brief peripheral cue located at one of the four corners of the display area. On valid trials the "search" display was then presented at the cued location, and on invalid trials it occurred at one of the other three possible corner locations. Search displays contained four colored letters that were presented briefly and then masked. Participants had to report whether the target was present in the display. The critical data were the incidence and type of false alarm responses. Such responses were classified as either feature errors or conjunction errors. Generally speaking, participants made more false alarms (of both types) on invalid trials than on valid trials. In addition, they made considerably more conjunction errors than feature errors, and the size of the validity effect (the difference in error rates for the invalid and valid trials) was greater in the conjunction data than in the feature data. Although much might be made of this overall pattern of data, Prinzmetal, Presti, and Posner (1986) were most concerned to argue only that the manipulation of attention (i.e., the cuing manipulation) had influenced the incidence of feature errors. They took this to demonstrate that when attention was diverted from the target's location, the incidence of feature false alarms increased relative to when attention was focused at the target's location. Clearly, if featural processing was insulated from any attentional involvement, such a pattern should not have occurred.

More recently, the idea that attention might be implicated in feature detection gained some further support from the work of Joseph, Chun, and Nakayama (1997). They constructed a dual-task variant of visual feature search in which the display elements were small, oriented gratings (Gabor patches). Half the displays contained an odd man out (i.e., the target), which possessed a distinctive orientation, against a background of nontargets, which were of a common but different orientation. In the single-task condition, performance provided evidence of parallel search, with RTs being essentially unaffected by display size. However, a quite different pattern emerged when this task was combined with a second task in the dual-task condition. In both the single- and dual-task conditions the to-be-searched elements were positioned in the periphery on the circumference of a virtual circle. For the secondary task in the dual-task condition, participants monitored a rapidly changing stream of colored letters presented at the center of the display. Participants were instructed to report the letter presented in white. This time the critical manipulation was the lag between presentation of the target letter and the presentation of the search display. The data clearly revealed a dual-task decrement: Search accuracy showed a dramatic effect such that participants were roughly 60% accurate when the presentation of the target letter and search display coincided. Accuracy then improved, in an approximate linear fashion, as the lag between the target letter and the peripheral
The presence of the dual-task decrement was taken by Joseph et al. (1997) as an indication of clear attentional constraints on featural processing. Having participants concentrate some of their attention on the central letter detection task clearly interfered with their ability to detect the peripheral feature target. In contrast, the strict notion of preattentive feature processing would have been consistent with feature detection performance being unaffected by the inclusion of the secondary task. Joseph et al. took this to be a dramatic demonstration of how attention is implicated in feature detection and argued that such a finding was quite at odds with the notion of preattentive forms of perception.

More recently, though, these conclusions have been undermined somewhat by the work of Braun (1998). In very similar tasks, Braun demonstrated that performance was critically dependent on how practiced participants were. That is, he was able to replicate the findings of Joseph et al. (1997), but only with novice observers. For trained and expert observers—those highly familiar with the task and the stimuli—there was no discernible dual-task decrement. Featural pop-out occurred under both single- and dual-task conditions. According to Braun (1998), task familiarization resulted in a tuning of the perceptual system in giving rise to "a direct route from preattentive processing to perceptual report" (p. 425).

Clearly, therefore, caution is warranted against attributing too much importance to some of the demonstrations of attentional effects in feature detection tasks. However, there are now also clear examples that featural detection is an effortful process. For further discussion of these issues, see the interchange between Folk, Remington, and Johnston (1992, 1993) and Yantis (1993).

**Summary of Extensions to the Theory—FIT Version 2**

By the early 1990s the theory could be summarized in all important respects thus (Treisman, 1991, 1992):

- Featural pop-out arises because of the presence of a distinctive feature value signaled on a particular feature map.

- When nontargets and targets share feature values (as in the case of conjunction search), the attentional focus must be narrowed so that a small number of elements (possibly one) can be checked sequentially. Difficult feature searches may also demand a narrowing of the focus of attention, and hence a serial search function may be produced.

- Attentional selection operates at the level of a master map of locations. Accessing a location on this map results in the immediate recovery of all the elements' attributes from the corresponding feature maps.

- Attentional selection is defined relative to an attentional window—the spotlight of attention is replaced by an attentional window—which operates as a selective filter on the master map of locations and the size of the window is variable (under top-down control). Typically, in visual search, participants operate in a divided-attention mode in which attention is distributed across the whole display. The detection of a given target feature then depends on the narrowing of attention to the location of this feature on the master map of locations.

- Rapid conjunction search may occur because of feature inhibition that operates on the feature maps that are critical to the task. Active feature detectors mutually inhibit one another, causing nontarget feature activity to be suppressed. In addition, the activity on the various feature maps may result in the selective inhibition of particular locations on the master map and hence speed search. The locations of nontargets represented on the master map are inhibited, thus allowing the serial scan to then operate over the remaining locations. Nevertheless, it is maintained that the need to bind features in establishing the presence of a particular combination of features contributes to the difficulty of performance in the conjunction search task. Such feature binding is an attention-demanding process. However, whereas identifying conjunctions typically does require attentional feature binding, there may be cases in particular search tasks in which such attentional processes can be circumvented. For example, performance may
reflect the operation of very early hard-wired detectors for particular conjunctions.

In terms of theory development, the initial strict one-to-one mapping relating feature search to the operation of preattentive processing and conjunction search to the operation of attentive mechanisms had been discarded. Such a simple classification scheme fails to do justice to the wealth of data that had been amassed in the visual search literature since the earliest description of the theory. Indeed, the strict dichotomy between feature and conjunction search was apparently laid to rest by Wolfe (1998) in his demonstration that there does not exist a strict mapping between type of search (i.e., serial vs. parallel) and the feature-conjunction distinction. (Upon further analysis, though, the notion that there exists a type distinction between fast and slow searches does have some support; see Haslam, Porter, & Rothschild, 2002).

Attentional Engagement Theory (AET) Versus FIT

Regardless of the other modifications to the theory that have been introduced over the years, it is still assumed that there is a fundamental distinction between featural detection and featural binding. That is, single features are typically easily detected, but conjunctions are typically recovered only through an additional (attentional) process of feature binding. It is this claim that remains at the heart of FIT (see Treisman, 1999). Nevertheless, it was this claim that came under direct attack from Duncan and Humphreys (1989). According to them, differences in feature and conjunction search performance should not be taken as indicators of an underlying type distinction between feature and conjunction processing as postulated in FIT. In contrast, Duncan and Humphreys (1989) argued that two factors are central in determining search performance. By their AET account, search difficulty is (essentially) directly related to the similarity of the target to the nontargets and (essentially) inversely related to the similarity of the nontargets to one another. T/NT similarity is taken as an index of how firmly the target groups together with the nontargets. The thinking here is that target detection should be easier when the target element stands out from a group than when it constitutes just one element in a group. Nontarget-nontarget (NT/NT) similarity provides a corresponding index of how firmly the nontargets group together. All other things being equal, target detection should be easier in cases in which the nontargets cohere into a single homogeneous group than in cases in which there is heterogeneity across the nontargets.

Clearly, the combinations of these two factors—that is, T/NT and NT/NT similarity—give rise to an indefinite number of search situations, and this range of possibilities is captured in AET by the notion of a search surface. Each point on this surface stands for a different and particular combination of T/NT and NT/NT similarities. The height of the surface (above some undefined baseline level) could be taken as an index of the ease of search as revealed by the slope of the corresponding search function. Targets are most likely to stand out (i.e., pop out) in cases of low T/NT similarity and high NT/NT similarity.

In extending these ideas, Duncan and Humphreys (1989) also discussed the notion of interalternative similarity with respect to the relations between the various possible targets and nontargets that might appear in a given block of trials (see also Duncan & Humphreys, 1992). Accordingly, performance can be critically affected, not only by the current similarity relations that exist in a given search display but also by the relations that exist within the set of possible stimuli. Such a possibility had not been entertained within FIT, and although ad hoc assumptions could be made to bolster the account, there is nothing in the theory to predict such an outcome.

Critical here is a visual search experiment reported by Pashler (1987b). Possible target letters were C and E, and two sorts of nontargets were defined. Nonconfusable nontargets were X and N, as neither of these was similar to either target. Confusable nontargets were defined as G, in being similar to C, and F, in being similar to E. The critical comparisons involved performance between different blocks of trials in which the nonconfusable and confusable nontargets were presented separately. The aim was to examine the effect of the presence of nontargets that were confusable with one target on performance with the other target. For instance, would the presence of Fs interfere with detection of a target C?

Each display contained six to-be-searched letters, but performance was classified according to the number of confusable nontargets that were present in the display: Displays contained either zero, two, or four
confusable nontargets. Overall, there was an effect of T/NT similarity on performance because RTs increased as the number of confusable nontargets present in the search display increased. Most important though, RTs were slowed regardless of whether the search display actually contained a target that was confusable with the currently displayed nontargets. Thus, there was an RT cost even in the case in which the target C was embedded in an array containing instances of the F nontarget. The mere fact that a target could occur among similar nontargets slowed performance.

**AET in Detail**

In overview, AET (Duncan & Humphreys, 1989, 1992) is divided into three components: (a) a parallel stage of perceptual description eventuating in a "hierarchically structured representation of the input across the visual array" (Duncan & Humphreys, 1989, p. 445), (b) an act of selection in which any input that matches with the current **attentional template** is allowed to proceed for further processing, and (c) the entry of this information into a visual short-term store where further analysis takes place. According to this theory, an initial parallel stage of perceptual segregation and analysis is followed by a limited-capacity stage that is characterized by the properties of a visual short-term memory. It is only the elements that enter the second stage that have a chance of controlling or determining future action. Central to this second stage is the notion of an attentional template, which represents significant possible stimuli (e.g., the targets in a visual search task). By this view, the probability that a given element enters the short-term store is indexed by the degree of similarity it shares with the current attentional template. To-be-searched elements are weighted according to how well they match the current attentional template, and an additional form of weighting known as **weight linkage** is invoked to explain interelement grouping in a display. According to AET, elements are grouped together by means of weight linkage and may be rejected en masse as nontargets because of "spreading suppression" along these common connections.

The most important criticism raised by Duncan and Humphreys (1989) was that the functional difference between featural and conjunction processing could be eliminated in favor of consideration of the different notions of interelement similarity described. By their view, search performance varies as a consequence of differences in interelement similarity across the different display types and not because of a type distinction between feature and conjunction processing. The argument was that difficulties in conjunction search arise, not because of the operation of additional feature-binding mechanisms, but (a) because the target shares one of its features with half of the nontargets and its other feature with the remaining half of the nontargets and (b) because the two types of nontargets are maximally dissimilar from each other. In other words, in conjunction search displays, T/NT similarity is relatively high (and therefore the target is likely to group with the nontargets) and NT/NT similarity is relatively low (and therefore the two types of nontargets are unlikely to group together). According to AET, the presence of both of these characteristics result in difficult search conditions. In contrast, in standard feature search displays, the target stands out because it possesses a unique feature (i.e., not one shared with any other element), and it shares its other feature with only half of the nontargets (e.g., given nontarget green Xs and brown Ts, the target blue T has a unique color, but it shares its shape with the nontarget Ts). In the terminology of AET, both NT/NT similarity and T/NT similarity are relatively low.

Duncan and Humphreys (1989) reported data from various visual search tasks that generally supported their ideas about how important interitem similarity is on performance. It is, nevertheless, interesting to note that in all of the cited cases, the to-be-searched elements were taken from the form domain: Conjunctions were defined as combinations of line and letter segments. In this respect, the data reflect featural combinations within a particular featural domain rather than across independent featural domains (cf. Humphreys et al., 1989). As a consequence, the data do not necessarily speak to the issue of how different featural types are combined during object perception (see Treisman, 1993, for more on this point).

The publication of AET in 1989 eventuated in an extended debate between Duncan and Humphreys, on the one hand, and Treisman, on the other. Although some of the points made are conveyed here, much of the detail is not. However, to appreciate the basic issues, several of the experiments need to be discussed in some detail.

Treisman (1991) reported a series of experiments aimed at addressing the counterclaims of
Duncan and Humphreys (1989), but only the second experiment is be considered here. Discussion also is limited to the so-called 2-D feature and conjunction search tasks. For these tasks, the to-be-searched elements were oriented colored bars, and for both the 2-D feature and conjunction conditions, the target was defined as the same blue vertical bar. In the conjunction search condition the blue vertical target was paired with pink vertical bars and blue left-diagonal bars; in the 2-D feature search condition the blue vertical target was paired with violet right-diagonal bars and turquoise left-diagonal bars.

Figure 4 - Examples of the types of target-present displays used by Journal of Experimental Psychology: Human Perception and Performance, 17, p. 656,

Figure 4 provides a schematic illustration of the types of displays used.

The 2-D feature search condition is notably different from previous examples because the target shares neither of its features with any of the nontarget elements. Previously, the target always shared one of its features with half of the nontargets. Here, Treisman's (1991) intention was to generate displays in which NT/NT similarity was the same as that in the conjunction condition. For each of the feature and conjunction search displays, two types of nontargets were defined so that the types differed according to both color and orientation; that is, they shared no features. The overarching rationale, though, was to try to equate both NT/NT and T/NT similarities across
the feature and conjunction search conditions. As an empirical check, therefore, two types of calibration experiments were run prior to the main search tasks, with the primary objective being to generate displays in which the interelement similarities were controlled for but in which the feature-conjunction distinction still obtained. Given this state of affairs, AET predicted no difference in performance across the feature and conjunction conditions, whereas FIT predicted the standard feature-conjunction difference.

The first kind of calibration tests were known as target-distractor similarity tests and involved tests of T/NT similarity. First, the salience of the target relative to each of the nontargets was assessed under standard visual search conditions. On different trials, the different types of nontargets (distractors) were used, although only one type of nontarget was present in the display. In this regard, the displays were homogeneous. Within a block of trials, a target element was present at random on half of the trials. Given such conditions, the results (predictably) revealed featural pop-out and, it is important to note, search performance did not vary across the different T/NT pairings. That is, the target was equally discriminable from each type of nontarget; Treisman (1991) argued that, therefore, T/NT similarity had been equated prior to running the main feature and conjunction search conditions.

In the second type of calibration experiment NT/NT similarity was tested in so-called distractor-distractor tests (Treisman, 1991). Participants had to make a certain response if they detected the presence of an odd element from an otherwise homogeneous background, and they were to make an alternative response if the elements were all the same. In these displays, the to-be-searched elements were taken from the nontargets defined for the main search tasks. Pink vertical bars were paired with blue left-diagonal bars in one case, and violet right-diagonal bars were paired with turquoise left-diagonal bars in the other. The trials were intermixed in a block so that participants never knew in advance which element would act as the odd man out. The data revealed that performance was the same regardless of which of the elements acted as the odd man out. The corresponding search functions were flat, and RTs were relatively fast. Treisman (1991) interpreted this pattern as demonstrating that NT/NT similarity had been equated prior to running the feature and conjunction main searches. As a consequence, the calibration studies had been successful in providing empirical support for the claim that both T/NT and NT/NT similarities were equated prior to the main search conditions.

The results from the main search tasks showed the following pattern: All of the search functions revealed that RTs linearly increased with display size, with the effect being more pronounced for absent responses than for present responses. However, for Treisman (1991), the most revealing result was that "the conjunction condition gave mean slopes that were more than three times steeper than the feature condition" (p. 668). In this regard there were striking differences in performance across the conjunction and 2-D feature searches that could not have been predicted solely from consideration of the types of interitem similarities as defined by AET. (For a further examination of the 2-D feature search condition, see Eckstein, Thomas, Palmer, & Shimozaki, 2000.) Indeed, Treisman (1991), in discussion, focused on the qualitative differences in search performance across the feature and conjunction conditions and was at pains to point out the "special source of difficulty in search for a conjunction target" (p. 659), a source that, according to her, reflects the use of attention to integrate the individual features.

Replies and Further Debate

In replying to Treisman (1991), Duncan and Humphreys (1992) took issue with both the reasoning and experimental approach adopted. For instance, they claimed that, although performance in the second type of calibration experiments—the distractor-distractor tests—provided a measure of spreading suppression among the nontargets, it failed to address the degree to which the nontargets compete with the target to match the current search template. In their original article, Duncan and Humphreys (1989) had invoked two ideas concerning access to visual short-term memory, namely (a) an item's similarity to the current attentional template and (b) spreading suppression between search elements. In AET, each search element is conceived as having an associated weight that reflects the degree to which it may be a likely target. Such weights index the degree with which the element competes with others to enter the short-term store. A given element's weight can be lowered through the operation of spreading suppression; this typically happens if the element groups with similar nontargets in the display. Duncan and Humphreys (1992) went on to claim that Treisman's (1991) experiments had, essentially, only addressed strength of grouping and
not item-template matching.

To address these concerns, Duncan and Humphreys (1992) examined performance in two new feature searches and a standard conjunction search. Identical targets were used in both feature and conjunction searches, and both types of search displays were intermixed within the same block of trials. Participants never knew in advance the nature of the next search display, although the target element was constant throughout. The search elements were small, colored rectangular patches of gratings defined relative to the dimensions of color, orientation, size, and spatial frequency. In the feature search tasks, there were four different display types, and across these display types the target contained a unique color, orientation, size, and, spatial frequency, respectively. In the condition in which the target's color was unique, no other to-be-searched element possessed this color even though every nontarget shared two of its other features with the target. For the conjunction searches, there never was a unique feature contained in the target—each of its features was shared by at least one other nontarget. However, compared against each type of nontarget, the target possessed a distinctive pair of features. On an important note, though, across the feature and conjunction searches, each nontarget shared two features with the current attentional template that defined the target element.

Duncan and Humphreys (1992) also added control displays that contained homogeneous nontargets that differed from the target by only one feature. Applying the notion of spreading suppression, Duncan and Humphreys (1992) predicted that performance would be most efficient in the control condition. In the associated displays, high NT/NT similarity would result in considerable suppression of the nontargets because these constituted a salient group of elements. Duncan and Humphreys (1992) also predicted equivalent levels of performance in their feature and conjunction tasks. In contrast, they argued that FIT predicted that performance in the control and feature conditions should be roughly equivalent and should also be considerably better than in the conjunction condition. The results of the experiment, in conforming to the predicted pattern, were taken by Duncan and Humphreys (1992) as providing more support for AET than FIT.

In reply, Treisman (1992) disputed the relevance of the data to her earlier (Treisman, 1991) account and also provided further evidence against the notion of spreading suppression. According to Treisman (1992), the poor performance in the feature search task defined by Duncan and Humphreys (1992) may have reflected possible search strategies adopted by the participants. Given that the different display types were intermixed, participants may have defaulted to a serial scan of the elements whenever the displays contained more than one type of nontarget. In this respect, performance with the feature search displays may have reflected the serial application of attention. Alternatively, participants may have used a simple feature-counting strategy. Participants may have engaged in the process of checking whether the target feature was present, once, more than once, or not at all. Feature-present responses depended on detecting a single occurrence of the feature; all other displays demanded a feature-absent response. Treisman (1992) argued that both the conjunction and the counting strategies were effortful and hence depended on attentional allocation. It is therefore understandable why Duncan and Humphreys's (1992) feature search task was so difficult.

Finally, Treisman (1992) reported an experiment intended to test the spreading suppression idea directly. The basic rationale was to design a new search task in which new nontargets were included in the standard conjunction search displays that would selectively suppress the other nontargets but would not affect target processing. For instance, assume that the target is a blue vertical bar and that in the standard conjunction search, the nontargets are blue diagonal bars and pink vertical bars. Treisman's (1992) argument would mean that, according to AET, introducing pink diagonal bars as new nontargets ought to help suppress the other two type of nontargets and hence make search particularly easy. In contrast to this prediction, though, there was no evidence of featural pop-out in the new search condition. Moreover, performance was predictable from a serial scan of just those nontargets that shared a feature with the target. Treisman (1992) therefore concluded that the notion of spreading suppression discussed by Duncan and Humphreys (1992) had little empirical support.

In the absence of any further reply from Duncan and Humphreys, the debate appears to have been concluded on this note. Treisman's (1992) position is bolstered by the repeated observation of the special difficulties that participants exhibit with conjunction searches when all other factors have (apparently) been
controlled for. Such difficulties in conjunction search have been taken as empirical evidence for the
attention-demanding process of feature binding. Nevertheless, Duncan and Humphreys (1992) convincingly
demonstrated that no account of visual search performance is complete unless some consideration is given
to interitem similarity of both the current and the possible to-be-searched elements. Their evidence is less
convincing in showing that the difficulties exhibited with conjunction search can be adequately explained
merely by recourse to such notions of similarity. Other data, however, do bear on these issues, and it is to
these that discussion now turns.

**Signal Detection Theory (SDT) Accounts of Visual Search**

Perhaps the critical difference between FIT and AET has been put most succinctly by Palmer, Ames, and
Lindsey (1993). According to Palmer et al., differences in feature and conjunction search performance are
typically explained by FIT in terms of qualitative differences in attentional factors, whereas AET claims such
differences arise as a consequence of quantitative changes due to sensory and decision processes. In
attempting to adjudicate between these alternatives, Palmer et al. carried out a series of experiments that
used variants of standard feature search. Accuracy, rather than latency, was the critical dependent
measure, and the basic task used a two-interval, forced-choice discrimination procedure rather than timed
visual search. In each interval a search display was presented, and in the first experiment, each to-be-
searched element was a short line segment. The two displays were presented briefly, one after the other,
and a target was always present in one of the displays. Participants had to indicate in which interval (i. e.,
display) the target had appeared. In a target-absent display, all the line segments were the same length. In
a target-present display, one of the segments was longer than all the others. Two manipulated factors were
central, namely (a) display set size, that is, the number of line segments present in a display (both of the
displays on a trial contained the same number of segments), and (b) the length of the target element.

The timings used on a given trial were fixed throughout—the displays were presented for only 100 ms so as
to avoid any possible effects of eye movements; however, across trials, the physical length of the target
was varied. The intention was to estimate the 75% correct threshold (target) length for each of the display
set sizes tested and to plot these threshold measures as a function of display set size. The data clearly
revealed a display set-size effect because thresholds increased monotonically as display set size
increased. To interpret this display set-size effect, though, Palmer et al. (1993) carried out a second
experiment.

The second experiment involved the same choice task, but only displays containing eight elements were
used. These elements were arrayed around a central fixation point, and on each trial a pre-cue was
presented prior to onset of the first display. The pre-cue contained a centrally presented indicator that
demarked possible target positions, and across trials, the pre-cue indicated either two, four, or all eight
possible target locations. Palmer et al. (1993) motivated this experiment by defining a relevant set-size
manipulation on the grounds that any contingent effect would indicate the influence of attentional rather than
sensory factors. In this case, relevant set size was defined with respect to the number of pre-cued
locations. The sensory factors were the same throughout because the same sorts of search displays were
used; that is, there were always eight elements present. Attentional factors varied because the relevant set
size changed as a consequence of the cue. Thus, the prediction was that a relevant set-size effect should
obtain whereby thresholds would increase as the size of the relevant set increased. At stake though was
whether the display-size effect in the first experiment and the relevant set-size effect would be qualitatively
different. Any such difference could then be taken to show that nonattentional factors were playing a role in
the original display-size effect. In contrast, no such difference would suggest a common attentional
mechanism was responsible for both effects. The results clearly showed no detectable difference across
the two types of set-size effects. Therefore, Palmer et al. confidently ruled out sensory factors in
interpreting the display-size effects in their first experiment. The same attentional mechanism was
assumed to underlie performance in both tasks. Indeed, Palmer et al. went on to argue that the attentional
effect reflected properties, not of a perceptual stage of stimulus encoding, but of a postperceptual decision
stage.

Central to the account is a view of processing as characterized according to the assumptions of SDT (D. M.
Green & Swets, 1966). By this view, the derivation of an internal representation of a stimulus (i. e., the
percept) is prone to noise such as that produced by the firing of neurons (Tolhurst, Movshon, & Dean, 1982). In other words, the percept is only an approximate indicator of the input stimulus. Given this, there is a degree of uncertainty associated with the input which must be taken into account in arriving at a decision over its identity. According to the SDT framework, because any display element can be confused with the target, increasing the number of nontarget elements produces a proportionate increase in the probability of making a false alarm. In other words, performance suffers as the number of to-be-searched elements increases because the probability of mistaking a nontarget for a target increases accordingly. Such ideas form the basis of the SDT explanation of display set-size effects.

In discussing such accounts of performance, Palmer et al. (1993) pointed out that attention operates to alter the signal-to-noise ratio such that T/NT discrimination is improved at the cued locations. Prior to discussing their own findings, Palmer et al. stated that such an attentional effect can operate at a perceptual stage of processing, at a decisional stage, or at both stages (see also Luck, Hillyard, Mouloua, & Hawkins, 1996). Within the SDT framework, each display is seen to give rise to activation on some continuum that specifies the critical perceptual dimension that defines the target's identity. Over trials the level of activation produced by the displays can be characterized by a normal distribution. As in standard SDT, the displays containing a target produce, on average, an activation distribution (the signal-plus-noise distribution) that is higher on the continuum than displays containing no target (the noise-alone distribution). Given that these distributions overlap, however, there will be trials in which a target-absent display takes on a value that is greater than that for some target-present displays. Therefore, participants must establish a criterion-activation value to establish a decision boundary. By this view, the main factors that influence performance are (a) the position that this criterion occupies relative to the means of the two (target-present and target-absent) distributions and (b) the distance between the two distributions. The greater the separation distance, the easier the decision for a given display.

The central insight in the account is that, depending on the composition of the search displays, the distance between the target-present and target-absent distributions changes. Roughly speaking, this distance diminishes as the number of to-be-searched elements increases, and at the largest display sizes, the distributions can overlap considerably (see Eckstein, 1998; Eckstein et al., 2000, for clear illustrations of this point). Most important, using formal techniques, Palmer et al. (1993) were able to quantify the changes in distance between the target-present and target-absent display distributions and use these to estimate changes in threshold as a function of the display size. Such estimates were then used to fit the data from the search tasks.

Palmer et al. (1993) concluded that the data were best explained by a model that assumed that participants tended to base their responses according to a decision integration hypothesis and not according to a perceptual coding hypothesis. By the decision integration hypothesis explanation, the effects of display size arise at a decision stage of processing and relate, essentially, to increases in uncertainty over the identity of any given search element as the size of the search display increases. Contrary to the perceptual coding hypothesis, there was no need to posit any change in the perceptual representation of the display elements as a consequence of increases in the number of to-be-searched elements. Implicit in the perceptual coding hypothesis is the notion that some perceptual operations are capacity limited. It is because of these processing constraints that the quality of the percept for any display element suffers as the display set size increases. As Palmer (1994) noted, the failure to support the perceptual encoding hypothesis also throws doubt on the claims of capacity limitations that in FIT are assumed to be, in part, responsible for difficult feature searches (cf. Treisman & Gormican, 1988).

According to the decision integration hypothesis, the decrement in performance with increases in display size is due to the operation of integrating the information produced from an increasing number of noisy sources so as to arrive at a decision. The next to-be-searched element gives rise to an additional noisy source of input, which then needs to be taken into account (i.e., integrated) in the process of arriving at a decision as to whether a target is actually present. (Although the work by Palmer et al., 1993, is relatively recent, Pashler, 1987b, cited the much earlier work of Gardner, 1972, as being an early discussion of, essentially, the decision integration hypothesis.) By this logic, Palmer et al. (1993) interpreted their data as being more in line with the AET (decision) account of search than the FIT (perceptual) account.

More recent extensions to the SDT account of search performance have been used to address further
issues concerning conjunction search. For instance, Eckstein (1998) discussed extensions to the model whereby consideration is given to cases in which the to-be-searched elements are defined relative to two featural dimensions. Here the idea is that each element now produces information along two dimensions, both of which give rise to noisy sources of information. Simply put, by this account the reason for the relative difficulty of conjunction search versus feature search is now explained with recourse to the increase in noise produced when two target feature dimensions need to be considered rather than one.

Some caution is warranted here, though, because of the design used by Eckstein (1998). The search elements were ellipses that varied in both orientation and brightness. Eckstein ran separate feature search conditions in which participants searched for either a distinctive orientation or a distinctive brightness. It is crucial to note, though, that orientation and brightness were tested in separate blocks of trials. In the conjunction search condition the target was a unique combination of orientation and brightness. In this respect, the contrast in performance across the feature and conjunction search conditions may have simply reflected differences in having to search for one versus two target features in the respective feature and conjunction conditions (see Quinlan & Humphreys, 1987, for a more thorough discussion of these issues). In the original Treisman and Gelade (1980) searches, the number of target features (i.e., one color and one shape) were equated across the feature and conjunction conditions, and because of this their experiment did not fall foul of the same confound. Despite these reservations, however, when Eckstein (1998) compared his SDT account of conjunction search performance with a simple serial-scanning mechanism of the type discussed in FIT, the SDT account provided the best explanation (see also Eckstein et al., 2000).

In assessing the significance of the SDT work it must be borne in mind that the data under consideration are measures of accuracy rather than latency; thus, whether SDT accounts can provide adequate fits with the wealth of RT data from visual search tasks remains to be seen. Indeed, the experimental conditions are very different across the two sorts of search task used. In the standard visual search case, displays typically remain on until a response is made, whereas in the SDT experiments discussed, the displays are briefly presented and the display elements are masked. It is therefore pertinent to question the degree to which the two sorts of paradigms tap into the same visual processes. Such caveats aside though, the critical points are that the SDT work brings into question central assumptions that FIT makes about visual search performance. Namely, the work casts some doubt on the necessity of positing a serial-scanning mechanism that is characterized as a limited-capacity spotlight of attention that governs feature binding. More interestingly, the work holds the promise of explaining differences in feature and conjunction search performance without recourse to positing qualitative differences in processing for the two types of target. On these grounds, the SDT accounts of search performance provide quite different and more parsimonious explanations to those put forward by FIT: Critically, they contain no mention of feature-binding mechanisms.

**Alternative Paradigms**

Since its inception, FIT has undergone considerable modification in order to address contrasting patterns of data. For instance, the idea that there is an initial featural processing stage characterized solely by preattentive operations followed by a conjunction-processing stage characterized solely by attentional operations has been reappraised. What remains, however, is the basic belief in the qualitative difference between feature detection and feature binding. The notion of feature integration remains central to the account, and the argument remains that feature binding is a real problem solved by the perceptual system (Treisman, 1999). In the current article, alternative accounts to FIT have been considered, and although these differ at the level of detail, they all attempt to explain feature and conjunction search in terms of the operation of same set of processes. None of the alternatives assumes a special role for feature binding. In this regard, it is useful to consider other aspects of task performance that are assumed to reflect on (putative) processes of feature binding. Alternative paradigms to standard visual search therefore are considered now, but what is of note is that all of these were motivated by additional hypotheses advanced in the original version of FIT. In particular, I discuss two further aspects of the performance that address the theory's predictions about (a) the coding of features and locations and (b) the occurrence of illusory conjunction. Each has generated separate lines of research even though they both, fundamentally, concern the same issues about whether, where, and when feature binding takes place.
Features and Locations

Although the development of FIT has been based on many well-articulated statements about underlying representations and processes, it is fairly difficult to try to provide a very clear statement about what the theory has said about location coding. This is particularly true of Version 1 of the theory. As M. Green (1991) pointed out, the claim that features are coded relationally within a particular map does seem to be at odds with the notion that they are also free floating. It is, however, only fair to point out that Treisman and Schmidt (1982) revisited the original notion of free-floating features on the grounds that this idea did not gel with the evidence of preattentive texture segregation: If texture segregation is to take place, then the positions of edges must be registered within a given feature map. For them, the more critical point was that although there may well be some form of spatial organization within a given map, the act of coordinating information across the different feature maps depends on the operation of the serial scanning (attentional) mechanism.

Nevertheless, one contentious claim of FIT that became the object of much research effort, was about how it should be possible to identify a feature without also necessarily being able to locate it. In retrospect, it is reasonably surprising that such a claim was made given the then-extant literature on iconic (i.e., very short-term visual) memory. It had been well established by Sperling (1960) that participants could effectively use a cued location to report items from briefly presented masked displays. Hence, the data did suggest that at the very earliest stages of processing, features were encoded together with their locations.

Nevertheless, the claim was made, and one early study that addressed this issue was reported by Nissen (1985). In Nissen's (1985) first experiment, participants viewed brief (and masked) displays containing four colored shapes. Each to-be-searched element was defined relative to one of four possible values from each of three defining dimensions, namely, location, color, and shape. Following the offset of the stimulus, participants were provided with either a location name or color name, and they had to report the color of the item at the named location or the location of the item in the named color, respectively. These were all of the significant events in the no-foreknowledge condition. By comparison, in the foreknowledge conditions, participants were provided with an additional verbal cue prior to each trial. The cue informed them whether they were to be quizzed about location or color. Report accuracy was the measure of interest, and the data revealed no effects of task instructions—there was no overall benefit shown in the data for the foreknowledge condition relative to that for the no-foreknowledge condition. More important though, there was no difference in accuracy levels for the two kinds of perceptual report. Participants were as accurate in reporting the color of the stimuli (given the location cue) as they were in reporting the location of the stimuli (given the color cue).

In her second experiment, Nissen (1985) examined participants' ability to report conjunctions of attributes when cued accordingly. In the location-cued condition, participants were to report the color and the shape of the item at a particular location. In the color-cued condition, they were to report the location and the shape of the item in a particular color. The data revealed clear differences in performance across these two new conditions. From her analysis, the data showed strict independence in participants' reports of color and shape when cued by location. However, when cued by color, participants' reports of shape and location were not independent. Participants' accuracy in reporting shape appeared to be dependent on the accuracy of reporting the target's location. Moreover, the data in the color-cued condition were predictable from estimates of participants' abilities to locate the target color and judge the shape at that location. As Bundesen (1991) stated, these results were taken to show that "cross-referencing between color and shape was mediated by location" (p. 87).

Overall, therefore, two basic assumptions of FIT appear to have been supported by Nissen (1985): (a) There was evidence of independent analyzers for color and shape, and (b) the coordination of information across these separate analyzers appeared to have taken place via some general spatial coordinate system. Nevertheless, both of these claims have been the cause of some further controversy. The claim of independence of processing of color and shape was disputed by Monheit and Johnston (1994; see later discussion). Debate has also ranged over the second conclusion. For instance, Bundesen (1991) recast Nissen's findings as showing that the data were also in line with the idea that the same principles of selection operate for location, color, and shape, given the additional assumption the location information is
resolved extremely fast. – Van der Velde and van der Heijden (1993) countered this argument and took issue with Bundesen's (1991) interpretation, but in the absence of critical empirical tests it is difficult to reach a clear conclusion. Nevertheless, as Bundesen (1993) stated, his basic idea (Bundesen, 1991) can be tested by making the location judgments particularly difficult because, as he argued, if his interpretation of Nissen's data is correct, the relationship (i.e., apparent dependence of shape report on the accuracy of location report) ought to break down when the location judgments are made more difficult. Given that this experimental manipulation has yet to be carried out, the issues remain unresolved.

A different thread in the literature took some of the findings reported by Treisman and Gelade (1980) as the starting point. In one of their experiments, participants were required to report both a target's location and its identity from briefly presented and masked displays. Displays contained two rows of six letters, and targets could occur only in the middle four letter positions. Nontargets were pink Os and blue Xs, and the target was either an X or an O colored orange (a distinctive color) or an H (a distinctive shape) colored pink or blue. In this respect, the target always shared its shape or color with one of the types of nontarget. The basic finding was that, on the trials in which participants inaccurately reported the target's location, they were, nevertheless, able to report its identity at levels greater than chance. Such a result clearly supports the FIT prediction that participants should be able to identify a feature in the absence of being able to locate it. Moreover, Treisman and Gelade took these results as supporting the idea of free-floating features as well as the claim that the recovery of location was an attention-demanding process. In a later article, however, Johnston and Pashler (1990) were much more cautious in their appraisal and argued that the experiment may well have suffered from two important problems: (a) a negative-information problem and (b) a location-reporting problem.

The negative-information problem relates to a form of sophisticated guessing that participants may have engaged in. Johnston and Pashler (1990) argued that, over trials, participants may well have been more confident in reporting one of the target types (e.g., the target color) than the other target type (e.g., the letter H) on the grounds that the target color was simply more detectible than the target shape. On trials in which participants missed the target completely, they may well have guessed the less salient target type because had the more salient target color been present they would have detected it. Given this strategy, participants' reports of the target identity would be inflated by guesses. Such a strategy would not, however, produce a concomitant benefit in reports of the target's location. On the other hand, the location-reporting problem goes some way to explain why participants' location reports were so inaccurate; the idea is that the target always shared its shape or color with one of the types of nontarget. The general thrust of the argument, therefore, was that both identity and location may well have been encoded perceptually, but particular difficulties arose in reporting the location of the target.

Having described both the negative-information and the location-reporting problems, Johnston and Pashler (1990) designed their first experiment to obviate both. To avoid the location-reporting problem, Johnston and Pashler designed their displays according to a hollow square configuration. Each corner and each midpoint of the side of the (virtual) square contained a letter. In addition, the masking display contained a masking character for each of the possible target positions. The aim was to provide, via the mask, a general coordinate system that would facilitate participants' location reports. In addressing the negative-reporting problem, Johnston and Pashler undertook pilot experiments to equate the levels of difficulty in identifying the separate color and shape target features. They also included trials (in the main experiment) on which no target was present so that participants' actual guessing rates for the different features could be estimated. Participants were allowed to make target-absent responses.

Overall, the data were clear in showing that participants were correct in reporting both target location and identity on a proportion (i.e., .524) of trials; that is, there appeared to be close binding of location and identity. In addition however, the data also revealed that participants knew the target's identity but not its location on approximately only 10% of trials. On these grounds, Johnston and Pashler (1990) argued that the data revealed rather tight binding of location and identity—although there was also some evidence showing some independence of reporting identity and location on a small proportion of trials. Participants were able to report a target's identity in the absence of being able to locate it on a small percentage of trials, and similarly, on a small percentage of trials, participants' location responses were accurate even when they were unable to identify the target. Having ruled out both the negative-information and location-
reporting problems, they took their data to be more compelling than those originally reported by Treisman and Gelade (1980). Indeed, they argued that their data suggested tight location and identity binding even when performance had been tested under single-feature search conditions. Such a conclusion is clearly at odds with the earlier assumptions of FIT, especially those concerning the free-floating nature of features. 10

More recent studies have extended these findings to encompass possible processing dependencies between other visual features. For instance, in a series of carefully controlled experiments, Monheit and Johnston (1994) examined participants' perceptual reports of briefly presented and masked displays that contained colored letters. The results of these experiments appeared to reveal quite profound dependencies in participants' reports of color and form. Monheit and Johnston, however, were concerned to point out that such dependencies did not necessarily reflect a process of feature integration; rather, they construed attention as modulating the ease or difficulty of perception (similar ideas had been discussed previously by Isenberg, Nissen, & Marchak, 1990). That participants were able to report both attributes with equal efficacy might merely indicate that both attributes shared the same attentional state rather than that the attributes had been conjoined into an integrated representation. Such an argument therefore leaves open the possibility that even though there are apparent interdependencies in participants' reports of distinguishable visual features, this need not reflect any processes of feature binding. Cast in this way, though, the claim is tantamount to admitting that the data have no bearing on the issues of whether, where, and when feature integration takes place.

Although the data reported by Monheit and Johnston (1994) were apparently clear cut, they nevertheless provoked controversy, because van der Velde and van der Heijden (1997) imposed a quite different interpretation. They preferred to argue that the dependence of color and form reports may have been an artifact caused by random guessing (see also Brouwer & van der Heijden, 1996). Much of this debate hinged on different assumptions concerning how best to treat the data. Van der Velde and van der Heijden (1997) offered a framework in which the overall matrix of responses is composed of two component matrices: one in which at least one of the features is perceived correctly and another in which performance is based on guessing. In working through this method, they demonstrated that both matrices show independence between color and form reports even though the undecomposed matrix shows statistical dependence. In reply, however, Johnston, Ruthruff, and Monheit (1997) argued that the approach advocated by van der Velde and van der Heijden (1997) was flawed and showed that, in working through the mathematics, "independent feature perception combined with independent guessing must produce independence in the observable data" (p. 1816). The debate (at the time of the writing of this article) concluded on this reasonably unsatisfactory note. However, it is possible that some further progress might be made if alternative approaches to data analysis were to be adopted (see later discussion of multinomial models in the Modeling illusory conjunctions section).

In summary, it seems that the more recent data—data collected after 1980—have cast some considerable doubt on (a) the notion of free-floating features and (b) the claim that, in general, features can be identified but not located. There is also much contradictory evidence over issues concerning the degree to which the processing of different visual features is independent. According to Version 1 of FIT (at least), the notion is that, at its earliest stages, visual perception is characterized by the operation of independent feature analyzers. A problem facing the perceptual system, therefore, is to bring about, in a coherent fashion, cross-dimensional feature integration. Various later reports fundamentally challenged this characterization in arguing for (a) the initial close binding of features and (b) a special role for location information in, essentially, providing the fixed coordinate system about which different featural information is integrated. According to Pashler (1998), "the simplest interpretation (of the relevant data) is that the selection by location mediates selection by ... other precategorical attributes" (p. 99; see also Tsai & Lavie, 1988, 1993, for supporting evidence). This is not to argue that attentional selection cannot be governed by characteristics other than spatial locale (such as objecthood; see Duncan, 1984) but merely that selection by location seems to pervasively influence visual processing (Moore & Egeth, 1998). Indeed, more recently, Bundesen (1999), in describing his own theory of visual attention (Bundesen, 1990), has accepted a special role for location in vision. According to him, there is a sense in which attention can be spatially focused at a particular region of the visual field with the consequence that elements outside the region of focused attention can be effectively ignored.
Other issues regarding the dependence versus independence of featural processing have been picked up and picked over in the literature on illusory conjunctions, and it is to these that the discussion now turns.

**Illusory Conjunctions**

In the original version of FIT, a corollary of the notion of free-floating features involves the occurrence of *illusory conjunctions*, that is, the incorrect bindings of features currently present in a given visual array. This is not to argue that, by necessity, reports of illusory conjunctions reflect the free-floating nature of features but merely that such demonstrations sit comfortably with such an assumption. Moreover, the occurrence of illusory conjunctions adds credibility to the claim that attention is fundamentally implicated in feature binding. The basic premise of this notion, taken from FIT (Treisman & Gelade, 1980), is that incorrect feature combinations are perceived whenever attention is diverted from the target object or is overloaded in some other way. As a direct test of this claim, Treisman and Schmidt (1982) designed a type of dual task in which, essentially, the primary task was intended as a means of diverting attention away from the secondary task. In their first experiment, participants were presented with brief visual displays that were immediately followed by a mask. The displays each contained a single row of characters in which the two outer characters were black digits and the three central characters were colored letters. Participants were asked to first try to report the identity of the digits (in the primary task); they then had to try to report the colors and identities of the letters (in the secondary task).

The central findings relate to the accuracy of report of the colors and letters: Participants were far more likely to incorrectly report combinations of letters and colors contained in the display than they were to report one attribute from the display bound with another that had not been present. That is, so-called *conjuction errors* were far more prevalent than so-called *feature errors*. The experiment had, apparently, demonstrated the occurrence of illusory conjunctions in line with the predictions of FIT.

The critical point Treisman and Schmidt (1982) took away from this initial demonstration was that it was perfectly in line with FIT insofar as that (a) each feature in the visual array is coded as an independent entity and (b) when attention is diverted, features are likely to combine at random. In addition, Treisman and Schmidt failed to find any distance effects in their data because the reported feature migrations did not appear to be sensitive to the distance between the display elements that contained the migrating features. Illusory conjunctions of adjacent features were as prevalent as were migrations between features that were spatially distant. In this regard, free-floating features were assumed to interchange without constraint (Treisman & Schmidt, 1982, p. 139).

More recent work, however, has cast some doubt on some of these conclusions. For example, in an interesting variant on the basic task used by Treisman and Schmidt (1982), Tsal, Meiran, and Lavie (1994) varied the colors of both the flanking digits and the central letters. The aim was to assess whether illusory conjunctions were differentially affected by whether the items were assigned high attentional priority (i. e., the digits in the primary task) or low attentional priority (i. e., the letters in the secondary task). The data, though, revealed that illusory conjunctions occurred regardless of the attentional priority assigned to the items. Tsal et al. (1994) therefore argued that this evidence contradicted the assertion contained in FIT that "illusory conjunctions ought to occur within and outside the attended subset, but not between the attended and unattended items" (p. 350).

In a slightly later article, Navon and Ehrlich (1995) took issue with the basic finding that the incidence of conjunction errors was significantly greater than that of feature errors and, essentially, dismissed it on the grounds that conjunction errors are just much more likely to occur by chance than are feature errors. As they stated, "any accident during perception would probably be reflected in a conjunction error rather than a feature error" (Navon & Ehrlich, 1995, p. 61). Nevertheless, they did argue that a critical prediction of FIT is that the difference in the rates of conjunction and feature errors should increase as attention is diverted from the target stimulus. To examine this prediction further, Navon and Ehrlich designed a variant of a task originally described by Treisman and Schmidt (1982). As before, displays contained five characters: two flanking digits and three central, colored letters. In the single-task conditions, participants were told to ignore the digits. In the dual-task conditions, participants were told to identify the digits for later report. In both conditions, though, a probe letter appeared after the display offset, and participants had to indicate...
whether the probe matched any of the displayed letters. Displays were presented briefly and masked. Three types of probe letter were used: (a) an identical probe, in which the probe matched exactly a displayed letter, (b) a feature probe containing one displayed feature and a feature not present in the display, and (c) a conjunction probe, in which features from two displayed letters were combined.

The experiment was much more involved than has been conveyed by the foregoing brief description and contained several further manipulations. First, performance was tested at short and long exposure durations to see whether "sensory factors" were critically important. Second, the order of report was manipulated so that in the dual-task conditions participants either had to report the digits first or the letters first, or no order of report was imposed. Performance was also tested as a function of the delay between the onset of the central letters and the flanking digits. In the standard case, the onset of the digits and the letters was simultaneous; however, Navon and Ehrlich (1995) also tested cases in which the digits followed the letters.

To summarize the results greatly, the experiment found that conjunction errors were much more frequent than feature errors but that this difference was not modulated by any of the experimental factors. In addition, the difference between conjunction and feature errors was the same regardless of the deployment of attention (i.e., across the dual and single tasks). The only systematic effect of order of report occurred in the dual task because participants were overall more inaccurate in their responses when they had to report the digits first. Navon and Ehrlich (1995) used this particular pattern to argue for the rather late (i.e., postperceptual) locus of the errors in report. Moreover, they took their overall pattern of results to be quite damaging to FIT, noting that the failure to show that conjunction errors were no more likely to arise because of the absence of attention than were feature errors contradicted a basic assumption of the theory.

**Illusory conjunctions and their constraints.**

The demonstration that participants could correctly identify features that they were unable to locate was taken by Treisman and Gelade (1980) as clear support for the free-floating nature of constituent features. Although Treisman and Schimdt (1982) had backed away from such a notion in arguing for some form of within-map locational coding, several studies took the original idea of free-floating features as the basis for studying illusory conjunction formation. In particular, Prinzmetal and colleagues began to collect data suggesting that the formation of illusory conjunctions is highly constrained by various physical and perceptual factors. For instance, Prinzmetal (1981) showed that features contained within a single perceptual unit were more likely to recombine than were features from separate perceptual units.

In Prinzmetal's (1981) first experiment, the displays contained either two columns or two rows of four circles, and these were presented briefly and masked.
Figure 5 provides examples of the types of displays used. A perceptual unit was defined as a column or a row, respectively, in accordance with the gestalt principle of proximity. In each display two of the circles contained additional internal features. On target-present trials one of the circles contained explicit horizontal and vertical diameters in the form of a plus; the plus was defined as the target. Another circle also contained either a horizontal or vertical diameter made explicit. On target-absent trials two separate circles contained a single explicit diameter only. In the critical trials, though, one circle contained an horizontal diameter and another circle contained a vertical diameter. Under these conditions, a misperception of the display would most likely result in the impression of a circle containing a plus. Indeed, the data revealed that reports of illusory pluses were more likely to occur when the horizontal and vertical lines were contained within the same row or column than if each row or column contained a diametric line.

Nevertheless, other studies have established more provocative constraints concerning cross-dimensional feature integration. For instance, the idea that spatial proximity might play an important role in the incidence of illusory conjunctions was examined further by Cohen and Ivry (1989). To address this issue, they devised a dual-task paradigm in which participants were presented with brief, masked displays containing colored letters. On a given trial the initial fixation point was replaced by a centrally presented digit together with two peripherally presented colored letters. Eight possible letter positions were used. One pair of positions were above the fixation and flanked the vertical midline by less than 1°. A corresponding pair of positions fell at an equivalent distance below fixation. A left pair of positions flanked the horizontal midline of the display, as did the right pair of positions. Whenever both of the letters in a display occurred together at one of these pairs of positions then the display formed part of the adjacent condition. The far condition demarked those displays in which the letters were distributed across different position pairs. Participants were instructed to first report the identity of the central digit and then report the identities and colors of the peripheral letters. It was ensured that the displays were presented so briefly that participants could not have executed a saccade to a peripheral letter position—hence the assumption that the focus of attention was held at fixation. On these grounds, Cohen and Ivry claimed that any reports of illusory conjunctions would have arisen from items outside the focus of attention.

A. Cohen and Ivry (1989) went on to argue that, according to the free-floating features idea (Treisman & Gelade, 1980), featural interchanges should, essentially, occur independently of any spatial relations in the display. In contrast to this, conjunction errors were significantly more prevalent in the adjacent condition than in the far condition. That is, featural miscombinations were much more likely to occur between adjacent letters than between more distant letters. A. Cohen and Ivry therefore took this result to show that location information—albeit it in a rather inexact form—must have been retained with the features for such an effect to be observed (for a more thorough theoretical approach to these issues, see the work of Prinzmetal & Keysar, 1989).

Generally speaking, the evidence that had emerged subsequent to the Treisman and Schmidt (1982) article did reveal clear perceptual and physical constraints on the occurrence of illusory conjunctions, and consequently, these findings provided no support for the notion of free-floating features. On a more positive note though, Hazeltine, Prinzmetal, and Elliott (1997) began to use these constraints in a bid to ask more detailed questions about the putative nature of feature binding. In their experiments, participants were asked to report the location of a green letter in a string of five other-colored letters. Displays were briefly presented and masked. Participants also had to report whether the green letter was an O. Participants made location responses by a point and click operation on the computer’s mouse. In this respect, quite specific location responses were recorded. There were three types of trials: (a) target-present trials, (b) both-present trials, in which both green and O were present but in separate items, and (c) color-only trials, in which no O was present. On every trial the target color was present. Of the most interest was the performance on the both-present trials because it was predicted that these trials were most likely to induce illusory conjunctions. Indeed, this prediction was upheld by the data. Of equal concern though were the analyses of the location responses. These revealed that illusory conjunctions were most likely to be reported midway between the location of the color feature and the location of the shape feature. Moreover, the location responses
indicated that the position of the perceived illusory conjunction was determined by means of some aggregating procedure based on sampling both of the constituent features on each trial.

The basic point made by these data is that features are not combined at random. Rather, illusory conjunctions are spatially constrained by the actual locations of the individual features in the visual array. Moreover, the evidence suggests that the formation of illusory conjunctions arose because of online interactions between the processing of the constituent target features.

It is only fair to acknowledge that some of the demonstrations of spatial constraints on illusory conjunction formation can be accommodated by FIT on the understanding that the notion of free-floating features is clarified (Treisman, 1993, p. 17). Some of the demonstrations are consistent with the notion that features may be free floating but only before the attentional focus is narrowed to a particular locale. On this reading, FIT does predict distance effects through the process of “zooming in” of attention together with the idea of preattentive grouping. For example, given that such preattentive grouping gives rise to the sorts of perceptual units discussed by Prinzmetal (1981), there are the beginnings of an account of the fact that illusory conjunctions are more prevalent within perceptual units than between perceptual units.

In summary, the overall picture that has emerged is that illusory conjunctions are highly constrained by both perceptual and physical factors. The findings have challenged both the assumption of free-floating features and the contingent prediction that features recombine at random. Also, the notion the such illusions arise solely because of the lack of focused attention has had little support.

**Postperceptual illusory conjunctions.**

According to FIT, illusory conjunctions arise because of incorrect combinations of visual features at a perceptual-attentional stage which occurs well in advance of any response or decision stages. Nevertheless, as is clear from the study by Navon and Ehrlich (1995), this claim had produced some controversy, and there was additional evidence to suggest how such errors might arise because of postperceptual factors. For instance, Virzi and Eggeth (1984) had drawn, on an earlier occasion, an explicit distinction between perceptual conjunctions and propositional conjunctions. Whereas perceptual conjunctions are assumed to arise as a consequence of the operation of early perceptual coding processes, propositional conjunctions are assumed to arise because of higher level factors. This is an important distinction because it points to the possibility that similar sort of errors of reports may arise because of quite different reasons.

In relation to propositional conjunctions, Virzi and Eggeth (1984) cited the evidence provided by Treisman and Schmidt (1982, Experiment IV), which revealed quite dramatic miscombinations of features. For instance, in a display containing an outline of a red circle and a filled blue square, participants might report a filled, red square. In examining these issues, Virzi and Eggeth designed displays containing words printed in various colors. In their critical experiment (Experiment 2), a display contained a pair of flanking digits and a central column of three words. Two of the words were adjectives and the third was a color name. The displays were presented briefly (i.e., 200 ms), and participants first had to make a same-different judgment on the digits. Next they were to try to report both the words and their associated colors. For example, the display might contain BIG printed in red, BLUE printed in brown, and WIDE printed in black. Three sorts of error of report were critical: (a) word-to-ink (WI+) errors, in which participants mistakenly would report the color word as being an ink color (e.g., reporting WIDE in blue), (b) ink-to-word (IW+) errors—these were errors when the reported color name corresponded to another ink color in the array (e.g., reporting RED in brown), and (c) perceptual conjunctions—these were the standard illusory conjunctions—in which a word from the array was reported as being in the color of one of the other words (e.g., reporting BIG in black). The central finding was that all three types of errors occurred significantly greater than chance. Taking the WI+ and IW+ errors as indicative of propositional errors, Virzi and Eggeth argued that featural miscombinations might arise because of the operation of factors different from perceptual encoding; such miscombinations could arise because of faults in the operations concerning the semantic content of the input. Given this, they cautioned against the conclusion that so-called perceptual conjunctions arise solely as a consequence of the operation of perceptual processes.
The notion of propositional conjunctions was extended in the work reported by Prinzmetal and Millis-Wright (1984). They took the notion of a perceptual unit to a different level, designing a task in which strings of colored letters were briefly presented, at random, above and below fixation. In a critical experiment, participants were informed at the beginning of each trial of a color, and they then had to report whether the letter P was presented in this color in the display. The stimuli were taken from the set of 3 three-letter words—PIE, SPY, and MAP—and 3 three-letter nonwords—PLF, BPT, and NVP. Three sorts of trials were central: (a) those in which the target letter appeared in the cued color, (b) those in which the cued color was associated with another letter in the display, and (c) those in which the cued color was absent from the display. The important comparisons were made across performance with the words and nonwords for cases in which the cued letter was associated with another nontarget letter. On these types of trials, the finding was that participants were twice as likely to err when words occurred than when nonwords occurred. That is, the incidence of illusory conjunctions was significantly higher for words than for nonwords. Prinzmetal and Millis-Wright went on to argue that this finding was consistent with the idea that feature miscombinations were more likely to occur within a single perceptual unit than across different perceptual units, the central assumption being that whereas words correspond to single perceptual units, nonwords do not.

Consideration of this later body of work (Navon & Ehrlich, 1995; Prinzmetal & Millis-Wright, 1984; Virzi & Egeth, 1984) allows for the conclusion that that illusory conjunctions may arise for a number of quite different reasons. The distinction between perceptual and propositional conjunctions has been useful in highlighting the fact that some of the errors of perceptual report may arise for reasons other than failures at the level of perceptual encoding. Higher order factors can clearly influence performance, particularly when familiar stimuli (e.g., words) are used (see also Prinzmetal, Treiman, & Rho, 1986).

**Modeling illusory conjunctions.**

In an important attempt to derive a formal account of illusory conjunctions, Ashby et al. (1996) discussed a mathematical method for analyzing perceptual report data. Moreover, they discussed a location uncertainty theory, which provides an alternative account of illusory conjunction formation to that provided by FIT. According to the location uncertainty theory, there is a degree of positional imprecision associated with the coding of an object's constituent features (cf. A. Cohen & Ivry, 1989). For simplicity, take the case of a display containing a pair of colored letters, a red (target) A and a green (nontarget) S. The theory assumes that color and identity are coded separately but that there is imprecision in the codings of the positions of the displayed colors and identities. Moreover, the uncertainty associated with the position of the identities is independent of that associated with the colors. Central is the idea that the probability of a correct report of the target's color and identity is a function of the probability that the perceived distance between the target's identity and its color is less than the perceived distance between the target's identity and the nontarget's color. If one makes the additional assumption that there is variability associated with perceived locations of the constituent features (over trials), then the probability of a correct binding depends only "on the standardized distance between the mean perceived locations of the target and the nontarget letters" (Ashby et al., 1996, p. 168). By this account, the incidence of illusory conjunctions should decrease as the distance between the letters increases (see A. Cohen & Ivry, 1989), and it should also increase as the variance associated with the locational coding increases. In testing these ideas however, Ashby et al. acknowledged that the observed rate of conjunction errors may in fact arise through guessing.

To take account of such guessing, Ashby et al. (1996) went on to discuss various formal (multinomial) models of perceptual report (see Riefer & Batchelder, 1988), testing various assumptions about different kinds of guessing and their associated incidence rates. Ashby et al. compared three general types of models, and, using theoretically neutral terminology, they referred to feature miscombinations (i.e., illusory conjunctions) as conjugation responses. They initially used the models to account for conjunction errors arising in their own experiments. Displays contained a pair of inner, colored letters (i.e., a target and a nontarget letter) flanked by dollar signs. These display characters were briefly presented at random at one of four possible peripheral locations. Participants had to report the color and identity of the target letter, which was either an X or a T. As a direct test of the location uncertainty model, the spacing between the inner letters was varied systematically over trials.
The three types of models were known as the null, the location uncertainty, and the random binding models, respectively. In simple terms, the null model contains the postulate that feature-binding errors do not actually occur and that conjunction responses arise merely because of guessing. In contrast, both the location uncertainty model and the random binding model claim that feature-binding errors do occur. The incidence of such errors is predicted by the location uncertainty model in the manner as described at the beginning of this section. However, the random binding model is an attempt to formalize the claims of FIT. According to this model, on some proportion of trials features are bound at random, but, more important, such errors are essentially unconstrained by the physical and/or perceptual relations that are associated with the display.

In greatly summarizing their results, Ashby et al. (1996) reported that the worst fit to the data was by various types of null models and that both the location uncertainty and random binding models provided better fits. Moreover, the best fit of all was provided by the location uncertainty model because it was the only model to predict distance effects in the data. (Although the detailed original model, as described above, failed to provide the best account of the distance effects found in the data, a slight variant of the model did considerably better.) The critical point however, is that Ashby et al. interpreted the data by claiming that illusory conjunctions are real and do arise for reasons other than guessing. They also stressed that the multinomial modeling techniques overcame problems in previous attempts to partial out guesses from real illusory conjunctions (see, e. g., Treisman & Schmidt, 1982).

Although the account of conjunction errors provided by Ashby et al. (1996) appeared to settle many issues in the literature on illusory conjunctions, a dissenting point of view more recently was provided in the work of Donk (1999), and a new debate has taken place (Donk, 2001; Prinzmetal, Diedrichsen, & Ivry, 2001). The central point made by Donk (1999) was that conjunction errors can arise merely because participants confuse the nontarget letter as the target. Indeed, Donk (1999) went so far as to suggest that even the distance effects reported by A. Cohen and Ivry (1989) arose as a result of the differing probabilities of confusing the target and nontarget letters. More important, Donk (1999) argued that illusory conjunctions qua miscombinations of constituent features do not occur because the perceptual system never makes such binding errors.

Under attack in this debate were the central assumptions of FIT regarding the initial independence of the processing of different types of features and the subsequent recombination of featural information at a later feature-integration stage of processing. Donk's (1999) central claims were that (a) there is little evidence of featural processing independence and (b) illusory conjunctions may be due, in large measure, to confusions in assigning identities to displayed items (cf. Johnston & Pashler, 1990). Her basic position, though, was that miscombinations of features do not occur as a consequence of malfunctions of the perceptual system.

Prinzmetal et al. (2001) clearly disagreed, and the ensuing debate appeared to cement differences in the interpretation of the data; no rapprochement has been reached. However, other evidence can be marshaled against the view that perceptual illusory conjunctions never occur. There is good evidence to suggest that features of objects can interchange across time. Such evidence comes from studies using the rapid serial visual presentation (RSVP) paradigm. A typical experiment is that reported by Keele, Cohen, Ivry, Liotti, and Yee (1988), in which participants monitored a rapidly presented sequence of 15 characters. Fourteen of the items were letters, and the target item was a digit. The characters were presented in black on a colored background, and the color of the background was different for each item. Participants were required to report the color of the background of the digit. Evidence for the occurrence of illusory conjunctions came from the finding that, in participant's reports, there were color transposition errors. That is, on a small percentage (i. e., approximately 5%) of trials, participants reported the color associated with a letter two positions away from the target in the sequence, and on a larger percentage (i. e., approximately 15%) of trials, the colors reported came from adjacent letters. Such a pattern of performance has been replicated in many other studies with other types of stimuli (see, e. g., Botella, Garcia, & Barriopedo, 1992). The most relevant point is that such illusory conjunctions do occur in a predictable fashion. Moreover, given that such illusory conjunctions occur across time, it perhaps would be surprising to also discover that they were never to occur across space, contra the assertions of Donk (1999, 2001).

Performance in other nonsearch tasks.

http://bll.epnet.com/citation.asp?tb=0&_ug=sid+1CB31825%2DBA3B%2D34B7%2DA96%2D1A8E1B3D9%40sessionmgr3+b000&l=en&cxt=0&bLibsid=fv+1...
Before bringing this discussion to a close, it is important to consider the degree to which FIT explains performance in tasks different from visual search. It might be argued that some of the work on illusory conjunctions has included tasks that appear to be very different from standard visual search. Yet, in cases in which a target is embedded among other nontargets and the position of the target is unpredictable (e.g., Prinzmetal, 1981), it is clear that some form of search is necessary. It is therefore useful to consider other paradigms that do not rely on search. Although there are not many such cases, there are a few notable examples, and three are summarized here.

The first example is that provided in a recent article by Woods, Alain, and Ogawa (1998). In this experiment, participants were presented with RSVP displays containing a single stream of stimuli at fixation. Each frame contained an oriented colored bar, and participants merely had to press a key upon detecting a predefined target. In separate conditions, participants monitored for (a) a particular orientation, (b) a particular color, or (c) a particular orientation-color conjunction. Analysis of the RTs revealed that overall, participants were slower in responding in the conjunction condition than in either of the two feature conditions, yet on a more detailed analysis, this RT cost was not consistently observed across all participants in all conditions. RTs on the orientation and color trials were examined separately, and the feature identified with the longer responses was identified as the slowest feature. When comparisons were drawn between RTs on conjunction trials and those on the slowest-feature trials, in some cases, the cost in the conjunction data changed to a benefit.

Woods et al. (1998) discussed the relevance of these findings to FIT in terms of the supposition that at least some of the cost in performance with the conjunction targets in visual search reflects the spatial movement of the attentional spotlight. In their experiments, however, no such spatial orienting was necessary because every display element occurred at the same screen position. In this regard, performance in their conjunction-detection task might be taken to be irrelevant to the theory (for further discussion of this point, see Cortese, Bernstein, & Alain, 1999). What is, nevertheless, interesting is the fact that in some cases responses were shorter on some conjunction trials than on the corresponding feature trials. This particular pattern was taken by Woods et al. (1998) as favoring an account in which the processing of different features interacts at an early stage of analysis, contra the FIT notion of early featural independence. However, given that Woods et al. (1998) examined performance with color-orientation conjunctions, the data are equally consistent with the FIT idea of such early conjunction detectors (see previous discussion of the work of Holcombe & Cavanagh, 2001; Houck & Hoffman, 1986).

A different example of an alternative experimental paradigm to visual search is that described by Lavie (1997). Participants viewed displays containing three colored shapes: a central target and two flanking nontargets. Although displays contained more than one element, the position of the shapes was fixed over trials; thus, support for the notion of performance being based on search is somewhat strained. Participants were tested under go/no-go conditions; A response was only required on trials in which either a purple cross or a green circle was present as the target. Performance was assessed as a function of the types of flanking elements that were contained in the displays.
Figure 6 provides a schematic representation of the sorts of displays used.

In conjunctive displays, one of the flanking elements was one of the target types. In the compatible case (Figure 6A), this element matched the central target, and in the incompatible case (Figure 6B), the element did not match the central target. In the disjunctive displays, both features that defined the target were present but in the separate flanks. In the compatible case (Figure 6C), one flanker shared a common color with the target, and the other flanker shared a common shape. In the incompatible case (Figure 6D), one flanker contained the color of the alternative target, and the other flanker contained its shape. Two single-feature conditions were also included: In the compatible shape case (Figure 6E), the target shared its shape with both flankers, and in the incompatible shape case (Figure 6F), both flankers contained the shape of the alternative target. Similar contingencies obtained in the corresponding color conditions (Figure 6G and Figure 6H, respectively). A final neutral condition was included in which both of the target features were unique.

The critical predictions concerned comparisons between the two two-feature conditions (i.e., the conjunctive and disjunctive conditions) and the single-feature conditions. Lavie (1997) took the effects to reflect the influence that the nature of the flanking elements had on responding to the target (in the same way that response competition effects have been interpreted in the flanking studies of B. A. Eriksen & Eriksen, 1974, and C. W. Eriksen & Eriksen, 1979). It was the differences in RTs to the compatible and incompatible displays defined for each condition (see Figure 6) that were germane. That is, assuming that information from the flanks did influence performance, Lavie predicted that RTs would be shorter in cases in which the flanking information and the target information was compatible than in cases in which it was incompatible. In addition, Lavie predicted that the compatibility effects should be larger in the two-feature conditions than in the single-feature conditions. This was hypothesized on the basis of the assumption that information from both flanking elements should influence performance, all other things being equal. However, the most important prediction was that the size of the compatibility effect should be equivalent across the conjunctive and disjunctive conditions.

Overall, all of these predictions were born out by the data. The critical finding was that large and equivalent compatibility effects were observed in the 2 two-feature conditions. This pattern is perfectly consistent with the FIT hypothesis that only separate features should be processed for unattended objects; that is, the size of the compatibility effect was the same regardless of whether the target-flanking features were themselves conjoined in a single unattended object or whether they were distributed across two different unattended objects. These data support the FIT hypothesis that the constituent features of unattended objects are processed separately and independently. The theory asserts that feature integration typically occurs only
for attended objects.

The final example of relevant data derived from alternatives to visual search are those recently reported by Pashler and colleagues (Huang & Pashler 2002; Morales & Pashler, 1999). In these studies, the basic task was speeded symmetry detection, in which on each trial a patterned display containing colored elements was presented. Participants simply had to respond as to whether the overall pattern was symmetrical with respect to the vertical midline of the display. In the experiments reported by Morales and Pashler (1999), it was found that RTs increased dramatically as the number of different colors (hues) were presented in the display. This was taken to argue against the notion that the spatial relations of all of the hues were checked in parallel. On the contrary, it appeared that participants adopted a serial mode of operation in which each hue was considered in turn.

Huang and Pashler (2002) extended this pattern of results in experiments that tested other basic features by examining performance with displays containing elements defined by size, orientation, and spatial frequency, respectively. Color was also retested. For all of these features the same basic effect obtained: It was found that the time to judge the symmetry of the display scaled as a function of the number of different feature values that were present. For Huang and Pashler, this basic result was generally in keeping with the FIT idea that the feature-binding mechanisms are effortful and are limited in capacity. In their tasks, the data revealed quite a strict constraint in showing that "the visual system cannot base a symmetry judgment on the simultaneous binding of different feature values to their respective locations" (Huang & Pashler, 2002, p. 1430). In this regard, therefore, the data provide converging evidence for the FIT idea that feature integration is an attention-demanding process.

In broad terms, therefore, it can be seen that FIT does have implications for attempting to explain performance in a range of tasks other than visual search. Moreover, although auditory processing is outside the scope of the present review, there is a discernible and recent trend in considering the FIT perspective as being an appropriate framework for thinking about problems in audition (Dyson & Quinlan, 2003; Hall, Pastore, Acker, & Huang, 2000; Mondor, Zatorre, & Terrio, 1998; Woods, Alain, Diaz, Rhodes, & Ogawa, 2001; Woods et al., 1998). This is another testament to the generalizability of FIT.

**Conclusion**

The influence of FIT has been felt throughout the field of visual cognition. The theory touches on many aspects in vision—from the earliest stages of sensory encoding to the nature of internal representations implicated in object recognition. Thus, although some might have taken FIT to be predominantly a theory of performance in visual search tasks, this is far too narrow a view. It has not, therefore, been my intention to provide a comprehensive review of theories of visual search. There are several notable examples of such theories (see, e. g., Bundesen, 1990, 1991; Grossberg, Mingolla, & Ross, 1994; Müller, Humphreys, & Donnelly, 1994; Wolfe, 1994; Wolfe et al., 1989), and I have mentioned some of these in passing. FIT, however, is considerably more than a theory of visual search, and even though most of the relevant data are taken from search tasks, the theory has much wider implications. Since its initial description, the theory has evolved, and some of the changes that have been introduced have substantially altered the FIT framework for thinking about visual information processing. The nature of this evolution has been traced in detail in the body of this article, and competing ideas have been described and discussed. It is, nevertheless, useful to provide a list summarizing the central basic points that have emerged from the review.

- The early alignment of featural detection with preattentive processing and featural binding with attentional processing can no longer be sustained (cf. Treisman, 1993). Not only are there many examples of conjunction formation in the absence of attention, but there are also several clear demonstrations of attention being implicated in featural detection. From its very promising beginnings, FIT actually has made little headway in clarifying the relationships between structural constraints of the perceptual system and constraints concerning the deployment of attention. This is not to argue that the distinction between feature detection and feature binding is not useful but merely to note that how such a distinction relates to the deployment of attention remains a matter of some debate.
• Several points concerning the independence of processing of different sorts of visual features contained in FIT are not without controversy. The theory is set firmly on assumptions concerning the earliest stages of processing, in which visual input is analyzed by means of independent feature detectors. Later stages operate to combine this information so as to generate a coherent percept. In addressing these ideas, the features of color and shape have been most studied, yet conflicting patterns of data have been reported. Perhaps this is not so surprising, for it may well be that different tasks tap into different stages of processing and their corresponding levels of representation—but this is not the whole picture. From the studies that have addressed the issue of independence of feature analysis directly, there has been little consensus as to how best to treat the data. Until such problems with data analysis are resolved, it is unlikely that there will be consensus over interpretation. Only a handful of studies have applied newer statistical modeling techniques in the context of FIT (e.g., the application of multinomial models in the work of Ashby et al., 1996, and of Donk, 1999), yet it seems clear that such techniques hold the promise of allowing substantial progress in the future.

Nonetheless, concern has been expressed over whether dependence in perceptual report of two or more features necessarily implies that some process of feature binding has taken place (Isenberg et al., 1990; Monheit & Johnston, 1994). The associated implication casts some doubt over whether issues concerning feature binding can be resolved by appealing solely to some form of behavioral marker. Further consideration of this point is included below.

• There is, however, some consensus over a special role for location in visual information processing. The whole notion of concentrating spatial attention has, typically, been discussed repeatedly in terms of an ability to selectively filter out irrelevant information from parts of the visual field that fall outside the current focus of attention. Contingent claims have been made about how spatial aspects of the visual array provide a general coordinate system about which feature integration takes place. This is perhaps not so surprising when the nature of the visual system is taken into account. At the earliest stages of encoding, information is couched in terms of retinal coordinates, and this coding system is preserved throughout the visual system in a variety of retinotopic maps (see Cowey, 1985). The assumption that different brain regions deal with different visual aspects of the same object implies that there must be some form of coordination of this information to achieve a coherent view of the world. How such coordination takes place, though, remains one of the most important unresolved issues in perceptual psychology.

• The early notion of free-floating features has had little support. There is some evidence to suggest that the positions of different features may be coded with varying degrees of fidelity (Hazeltine et al., 1997), but the idea that different features range freely without constraint has almost no support. Moreover, the idea that free-floating features may recombine without constraint has also been contradicted repeatedly by the data. Although some have argued that such featural miscombinations never occur as a result of some failure of the perceptual system, such a claim has little backing in the literature. Such perceptual errors are relatively rare, but the weight of the evidence suggests that they do occur between objects that are temporally and/or spatially contiguous. Nevertheless, the issue of whether such errors arise purely as a consequence of the misdirection of attention has been called into question on numerous occasions. Although it was initially postulated that such errors are closely linked with the deployment of spatial attention, supportive data are lacking. No clear picture has emerged over possible relations between the occurrence of such perceptual errors and attentional allocation. Moreover, it is important to remember that such errors in perceptual report can arise for reasons other than failures during early perceptual stages of processing. Higher order factors can demonstrably influence performance, especially when the stimuli are familiar, such as when they are common words. It is a mistake, though, to suggest that all such errors are attributable to higher order (nonperceptual) factors or arise purely as a consequence of guessing.

One alternative possibility that has had little consideration in the FIT literature arose in the literature on iconic memory. It might be that all visual features are coded initially by location (in a sense, this is trivially true given that the earliest sensory data are derived in terms of the retinal coordinate system) but that information about the different features decays over different time courses (see Coltheart, 1980, for a much more thorough discussion of these sorts of ideas). Such ideas stand in contrast to Version 1 of FIT but not Version 2.

• With respect to visual search; performance in such search tasks is clearly influenced by the degree of
similarity defined across the different members of both the current and the possible to-be-searched sets of elements. However, there is also good evidence to suggest that over and above these factors, there are particular difficulties in detecting conjunction targets. Whether such difficulties reflect the operation of feature-binding mechanisms is moot. Alternative accounts suggest that the difficulties in conjunction search may reflect the operation of postperceptual mechanisms but, in continuing to defend FIT, Treisman (1999) was adamant that the difficulties in conjunction processing reflect the attention-demanding processes of feature binding. Future work must attempt to adjudicate between these alternatives.

- Perhaps the most important issue relates to feature binding. For this reason, this article concludes with discussion of the one aspect of the theory that remains germane—namely, that cross-dimensional feature binding is a problem that is solved by the perceptual system and that this process is, typically, an attention-demanding process.

As Navon (1990) clearly stated, "the common property of processes of integration ... is that they serve to establish links between previously separate pieces of information" (p. 453). The notion of feature integration is therefore only understandable on the grounds that separately processed features are subsequently bound together. Strictly speaking, therefore, the critical test for this account is to show that, initially, visual features X and Y are processed independently and that, at a later stage, both X and Y are bound together through some form of attention-demanding process of integration. What the present review contains, in part, is reference to a variety of studies that have attempted to garner empirical evidence for the operation of such binding processes. In nearly every case, however, there has been dissent about whether attention invariably plays a critical role and, in more extreme cases, whether the data reflect, in any way, feature-binding operations. On these grounds, it is reasonably sobering to countenance the thought that the properties of such binding processes may have no unequivocal behavioral marker. In discussing particular claims about binding predicated on temporal synchrony of firings of encoding units (Singer & Gray, 1995), Hummel and Biederman (1992) have remarked "that behavioral tests are inadequate in principle to falsify the temporal binding hypothesis. Rather, questions about the neural mechanisms of binding, by definition, reside in the domain of neuroscience" (p. 508). If a similar conclusion is to be drawn about visual feature binding in general, it seems that the debates need to shift from psychology to neuroscience. It is, therefore, perhaps not surprising to note that evidence of such a shift has recently begun to appear in the literature (see the collected papers in Humphreys, Duncan, & Treisman, 1999).

For instance, it is interesting to note that Treisman (1999), in her most recent review, concentrated primarily on the performance of the patient R. M. who had bilateral damage to the parietal cortex (see Friedman-Hill, Robertson, & Treisman, 1995; Robertson, Treisman, Friedman-Hill, & Grabowecky, 1997, for more thorough discussion). The claim was that R. M. exhibited particular difficulties in tasks that demanded conjunction processing. Although some might wish to use this data to argue for some strict form of localization of function, the evidence appears to provide independent validation of the basic distinction in FIT between feature detection and feature binding. Indeed, further support for the view that the parietal cortex plays some role in conjunction processing comes from a variety of other sources (A. Cohen & Rafal, 1991; Petersen, Corbetta, Miezin, & Shulman, 1994; Shafritz, Gore, & Marois, 2002). Indeed, converging evidence for such a view was provided by Ashbridge, Walsh, and Cowey (1997) in a study involving the performance of participants who did not have brain damage. In that study, participants carried out various types of visual tasks while undergoing transcranial magnetic stimulation. Such studies have shown that noninvasive stimulation of the parietal cortex can interfere with performance, particularly on conjunction search tasks.

Other brain regions also have been implicated as providing specialized conjunction-processing modules in studies on samples of patients with laterialized brain damage (see, e. g., Arquín, Cavanagh, & Joanette, 1994; Arquín, Joanette, & Cavanagh, 1993; Ashbridge, Cowey, & Wade, 1999), but localization of function is only of secondary interest when the promise of such work is considered. Evidence from such studies of the brain provide quite compelling demonstrations for the distinction between feature detection and feature binding. Indeed, it seems that many of the debates that have sprung from consideration of FIT may be settled only when more fine-grained analyses of brain function are carried out. Although FIT is fundamentally a psychological account of visual information processing, it may be that its validity can be adequately judged only once sufficient consideration has been given to the results of neuroscientific investigations. Of course, this is not to deny the existence of the growing literature on a putative neural mechanism underlying attentional control in general (see the integrative review by Posner & Raichle, 1994).
and the relevant articles in Humphreys et al., 1999) but merely to suggest that some of the outstanding issues concerning FIT may be resolved only once appropriate neuroscientific data have been collected.

Footnotes

1
See, however, the later interchange between Navon (1990) and Treisman (1990b) for a more thorough discussion of how best to interpret the search functions.

2
The most highly practiced participants completed 1,664 trials in the conjunction condition.

3
Although there are many notoriously difficult issues that surround the notion of similarity (Goodman, 1951, pp. 111-113), vision scientists have, typically, taken a pragmatic approach and have defined the term operationally relative to various indices of task performance (see Attentional Engagement Theory (AET) Versus FIT section).

4
As an anonymous reviewer pointed out, these searches resemble those tested in experiments concerning linear separability effects in searching for targets defined in terms of position in color space (see, e.g., Bauer, Jolicœur, & Cowan, 1999; D'Zumra, 1991). Although this is true, it does not impinge on the conclusion that performance in the intermediate target condition did not conform to the standard conjunction search pattern.

5
By an alternative reading of the data, Bundesen (1990) stated that this particular pattern of results simply shows that effects of target visibility (which was high in the original longer condition) can be annulled by a decrease in T/NT discriminability (which was low in the original shorter condition).

6
A related and earlier idea can be found in Pashler (1987a). A central finding from his visual search study was that, for displays with up to eight to-be-searched elements, there was no difference in the slopes for the feature and conjunction search functions. On these grounds, Pashler (1987a) argued that, more generally, search proceeds through one subgroup of (eight or fewer) elements in parallel at a time, then switches to another subgroup, and so on, until the display has been exhausted (for similar ideas, see also Bundesen, 1990, 1999).

7
I thank John Findlay for pointing this out.

8
In a typical cuing task, a directional pre-cue precedes the target display. The understanding is that the cue may be effective in directing the participant's spatial attention to a particular display location (see Posner, 1980, for a review).

The claim that location information is resolved extremely quickly has some support from the studies of Sagi and Julesz (1985a, 1985b). They contrasted performance in two tasks. In the counting task participants had to report the number of targets (horizontal or vertical bars) in briefly presented and masked displays containing diagonal bar nontargets. In the discrimination task participants had to report whether all of the targets were the same or whether there was an odd man out. Whereas there was little effect of display size on the counting task, there was a large corresponding effect in the discrimination task. On the grounds that the counting task depended merely on recovering the presence of a target and not its identity, Sagi and Julesz (1985a, 1985b) interpreted these data as showing that information regarding where a target is can be computed on the basis of a fast parallel process (see Atkinson & Braddick, 1989, for further discussion of this point).

Given such a contrary view, Treisman (1993) did examine these claims further and took issue with Johnston and Pashler (1990) over how they had scored the location responses. According to her, Johnston and Pashler accepted any response as being correct if the reported location fell in the same side of the display as the target. As Treisman (1993, pp. 16-17) noted, this method of scoring apparently failed to rule out a strategy based on the recovery of reasonably crude location information. Participants may simply have focused attention at random on one half of the display and noted whether the target was present within that half. As a matter of detail, though, Johnston and Pashler counted as correct any location response that was adjacent to that of the target such that adjacency was gauged not only vertically but also horizontally and diagonally. It is therefore true that some false reports of location were scored as correct, but the scoring method did not completely compromise the findings in the manner alluded to by Treisman (1993).

It is perhaps only fair to acknowledge that even though the free-floating nature of features has been repeatedly discussed in the literature—and was set out in fairly unambiguous terms by Treisman and Schmidt (1982, p. 139)—such a notion no longer describes a central tenet of FIT (see Treisman, 1993).

A more detailed discussion of this particular issue can be found in the debate between Briand and Klein, on the one hand, and Tsal, on the other (Briand & Klein, 1987, 1989; Tsal, 1989a, 1989b).

Indeed, Chow (1986) has presented evidence to suggest that the typical mode of processing (in iconic memory tasks) is described by "select-then-identify," wherein the process of selection is governed by location information.
References:


Psychophysics, 52, 113-138.


Journal of Neuroscience, 7, 3378-3415.


94. Quinlan, P. T. & Humphreys, G. W. (1987). Visual search for targets defined by combinations of color,


