Has Feature Integration Theory Come Unglued?
A Reply to Tsal

Kevin A. Briand and Raymond M. Klein
Dalhousie University, Halifax
Nova Scotia, Canada

Tsal (1989) presented a critical review of feature integration theory (FIT) and of a number of studies that have purported to show a direct relationship between focal attention and the perception of illusory conjunctions. Conceptual difficulties with the theory are highlighted, and Tsal concluded that the reviewed studies have not provided sufficient empirical support. We argue that neither Tsal's conceptual criticism nor his empirical ones are convincing enough to warrant rejection of FIT and that he has not suggested an obvious and unique theoretical alternative that inspires an appropriate empirical test. A problem for FIT that Tsal did not raise is discussed. That is, the visual primitives assumed by the theory are not clearly specified or related to known neurophysiological or psychophysical evidence concerning early vision. Nevertheless, we conclude that FIT remains a viable theoretical framework.

Illusory Problems with Feature Integration Theory

Tsal (1989) questions the "basic tenet of the theory which identifies feature perception and object perception with corresponding stages of preattentive and focal attention" (p. 394). The problem is simple: if attention can facilitate encoding of features, then a clear separation between preattentive and illusory conjunctions and argued that the lack of empirical support for FIT, combined with conceptual difficulties the model faces, renders its major claims unsupported. In this article we hope to demonstrate that none of his attacks on FIT can be sustained and that despite some ambiguity FIT remains a viable framework.

Treisman's feature integration theory (FIT) proposes that illusory conjunctions can occur if attention is diverted from a display of several figures (Treisman & Schmidt, 1982). Converging evidence from a wide variety of paradigms has supported the theory (e.g., Treisman, 1985; Treisman & Gelade, 1980; Treisman & Schmidt, 1982). Tsal (1989) critiqued four studies that purported to show a direct relationship between attention and illusory conjunctions and argued that the lack of empirical support for FIT is a viable framework.

If encoding of features is facilitated by attention, will this seriously impair FIT? Probably not, because encoding is a general term. There is a big difference between the registration of features by peripheral visual channels and the detection of those features, both of which can be considered encoding processes. Detection is used here in Posner's (1980) sense to indicate the ability to assign an arbitrary response to a given stimulus. Even if registration of features occurs preattentively, it is unlikely that the mapping of an arbitrary response to that feature is carried out preattentively. That is not to say that the facilitation of overt responses to features would be entirely at the response level, which implies a bias rather than a sensitivity effect. Instead, visual attention may have the effect of allowing an automatically registered (encoded) feature to have access to decision and response mechanisms, which would not involve criterion shifts. Tsal's (1989) use of the term perception (e.g., "The theory claims that feature perception can be performed preattentively, p. 395)" fails to distinguish these two aspects of encoding. If visual features can be registered preattentively but require attention to be used in some arbitrary task, this hardly seems incompatible with FIT.

It is also conceivable that preattentive processes provide only the knowledge that there is a discontinuity in a certain domain, without any characterization of the precise nature of that discontinuity. Sagiv and Julesz (1985) showed that observers could easily count the number of nondiagonal line segments in a background of diagonals but required focal attention to identify them as horizontal and vertical. Compatible findings have been reported by Pashler (1988), who found that detection of a discontinuity in one domain (e.g., shape) was interfered with by a simultaneous discontinuity in a different domain (e.g., color). Both of these results are compatible with the notion that featural differences are detected preattentively but that explicit identification of the nature of those differences requires attention. Although probably not clearly predicted by FIT, such a finding is still compatible

Footnote:
1 There is reason to question the conclusion that attentional effects on feature perception do involve a facilitation of encoding. At present, it is still an unresolved empirical issue whether orienting effects are a result of sensitivity changes or criterion shifts. Clearly, if the latter is true then demonstrating orienting effects with feature targets has no bearing on feature integration theory. Features may actually benefit from attention, but this case is not proven yet.
with the general notion that preattentive processing allows for only gross discrimination between regions on the basis of differences between features.

Tsal's (1989) second criticism of FIT is that conjunctions of features can occur preattentively. In actual fact, FIT claims that attention is only one of the means by which features are conjoined. In the absence of attention, various default procedures, real world knowledge and so forth, are used to conjoin features, precisely because we are biased to perceive objects and not dissociated features. The demonstration that features can conjoin independently of attention is thus not incompatible with FIT. It of course strains the plausibility of the model if so many examples of this are demonstrated that the actual role of attention in conjoining features turns out to be minimal. The assumption that knowledge and experience affect feature integration ought not to be used as an escape hatch in the face of numerous counterexamples to the assumed role of attention. At the present time, however, there is no plethora of such counterexamples, and the burden of proof appears to be left with those who question whether attention plays a major role in conjoining features.

A further conceptual issue Tsal (1989) raises is that of exactly how attention "glues" features into objects. The three alternatives he discussed share some partial overlap with those proposed in various writings of Treisman, but we disagree on which is more similar to FIT or has most empirical support. His first alternative, that attention takes features that have already been localized and joins them together by focusing on a specific location, is not similar to FIT, which assumes that features are free-floating (i.e., "their individual locations are not directly accessible," Treisman & Gelade, 1980, p. 126). Indeed, there may be no need for attention if features are already localized; FIT's point is that they are not. This point is made even more succinctly in recent formulations (Treisman & Gormican, 1988). Tsal's third proposed role for attention (localizing features) seems most similar to that advocated by FIT. Thus, in addition to possibly facilitating the identification of features, attention also localizes them, which may be partly responsible for the integration of features into objects. Is this the metaphorical glue that FIT has in mind?

FIT clearly specifies that features are preattentively free-floating. It has been less clear in specifying whether attention can also facilitate identification of these. As indicated earlier, even if this is true it seems to be irrelevant for the theory. Taken on its own, FIT represents a strong theory that summarizes a large body of data and has generated many predictions that have been successfully tested. We think it is best viewed as a valuable framework that embodies characteristics that any theory of object and pattern recognition will have to deal with. At present it is still evolving (see Treisman & Gormican, 1988), and the fine-tuning that seems to be an ongoing process will likely only strengthen it. It seems unlikely that the popularity of FIT will decrease in the near future unless sufficiently damaging empirical evidence is found that mandates that a new theory or framework be adopted. So far, such evidence is lacking.

Before continuing, we note that Tsal (1989) made a general criticism of studies assessing feature and conjunction perception, suggesting that they test only whether attention facilitates these processes and not whether they can occur preattentively. Supposedly the proper research strategy to answer this question is to assess whether, say, feature perception can occur without attention. Tsal's criticisms of certain paradigms begged the question of whether any paradigm could meet the stringent criterion he advocated. Strictly speaking, he was correct in stating that orienting paradigms do not assess preattentive processing. No one that we are aware of has ever suggested that targets occurring on invalid trials in a cuing paradigm are really unattended. Attention may end up focusing on the target after a delay and do a less efficient job, but it is wrong to assume that if attention is misdirected (spatially), it plays no role in processing an uncued target. But if it is true that making an arbitrary response to targets requires and benefits from selective attention, then one ought not to be surprised if even preattentively registered features are more quickly responded to when preceded by valid cues. Orienting paradigms are valuable not because they show that only conjunctions benefit from focal attention, although this may sometimes be the case (see Treisman, 1985). It is the relative magnitude of cuing effects with feature and conjunction targets that is important. Orienting paradigms represent an invaluable way of comparing the role of visual attention in the perception of feature and conjunction targets.

Illusory Problems With Empirical Evidence Supporting Feature Integration Theory

Tsal's (1989) empirical criticisms involve a number of ad hoc assumptions. For his rejection of FIT to be sustained, each of these has to be true, or the strength of his arguments diminishes accordingly. The studies he critiqued, carried out by separate laboratories with different procedures and parameters, have shown remarkable agreement. These findings are also compelling because they conform to results from other paradigms used to support FIT. To reject converging evidence of this type, either a fundamental weakness of FIT must be demonstrated, or some strong (and unique) theoretical alternative must be proposed in its stead. Tsal has not succeeded in either of these, and his strategy appears to be to find any weakness he can with the various studies, whether or not the assumptions required to accomplish this hang together in any plausible manner.

Treisman and Schmidt (1982)

One example of the ad hoc assumptions Tsal (1989) made can be seen in his critique of Treisman and Schmidt (1982). One of Tsal's points is that the critical Experiment 5 used greatly reduced exposure duration. If this affects feature identification more so than feature location and if very few illusory conjunctions are based on feature errors (Tsal's assumptions), then the number of feature errors will increase relative to conjunction errors. This largely eliminates the difference between these error types, the exact result found in Experiment 5. However, it is difficult to accept this interpretation given the lack of any evidence to support these assumptions.

In his critique Tsal (1989) indicated that memory processes may explain illusory conjunctions. It would be surprising if
memory processes did not influence illusory conjunctions, but this does not imply that only memorial processes influence them. Furthermore, the memory problem he hypothesizes involves a failure to retain combinations of unrelated features. Is there any independent evidence that individual features are easier to remember than objects? Logically, it seems that the opposite ought to be the case, particularly given the primacy of objects in Treisman’s approach (e.g., Kahneman & Treisman, 1984). Finally, any encoded representation (at whatever level, such as feature, etc.) can be considered to be held in memory to the extent that it continues to exist after the event that caused it. But if that representation (i.e., memory) is not available for conscious report until attention has been directed to it, why call it memorial as opposed to perceptual? There seems to be a gray area here, but in our opinion encoded representations that have not yet become (or that could not have become) conscious can as easily be described as perceptual phenomena as memorial.

As for Tsal’s (1989) dismissal of Treisman’s (1985) cuing study, it is claimed that a longer duration for conjunction targets gave more time for attention to move to the cued location, resulting in higher d’.

Tsal (1989) rejected Experiment 3 from this study on the basis of two “suggestive findings,” (p. 398) but his arguments here are not compelling either. One is that the “data manipulations” (p. 398) used in Experiment 3 reduced illusory conjunctions, which ought not to be the case according to FIT. These manipulations involved the use of more saturated colors and an increase in the spacing between items. Tsal’s interpretation of this is unjustified, because the spacing manipulation used by Prinzmetal et al. in this experiment could easily affect the number of illusory conjunctions simply because features were further apart and hence were less likely to conjoin incorrectly. In any event, task differences between Experiments 2 and 3 make direct comparison of error rates hazardous if not impossible. Tsal’s final point, that attention does not eliminate illusory conjunctions, is wrong because there is no reason for attention to do this.

Tsal (1989) pointed out the presence of a possible confound in our study, namely, that the difference between RPB and RPQ conditions was not simply feature versus conjunction search. It may also be considered a difference between search for one versus two features. (This criticism can also be raised about Treisman, 1983.) The seriousness of this potential problem depends on the existence of data showing that searching for two features is slower than for one. To our knowledge, very little data exist on this point, and certainly Tsal does not mention any. Treisman has at times used disjunctive search tasks, in which subjects indicated the presence of either Feature A or Feature B, to make more appropriate comparisons to conjunction search, which explicitly requires two features to be searched for. Disjunctive displays lead to parallel search, whereas conjunction search does not and presumably requires serial allocation of attention to each item. On the other hand, there is some evidence suggesting that feature differences, but not identities, are computed preattentively (see Sagi & Julesz, 1985; Pashler, 1988). If explicit identification of multiple features takes longer than identification of one feature, this is clearly a potential problem for interpreting our results. However, there is little evidence that supports this contention, so adopting this explanation is premature. In particular, it is doubtful whether searching for two features is sufficiently more difficult than is searching for one for this to explain the magnitude of the effects we observed.

Tsal (1989) tried to show that data from one of our own experiments (Experiment 1) provides support for his alternative explanation. In this experiment the stimuli were mixed, so the only logical strategy for subjects to engage in when they had the opportunity was to search for a conjunction of two particular features. They ought to have adopted the same strategy for both PQ and PB stimuli, and Tsal assumed this led to identical performance. This is in fact what we obtained, but it is unclear to us why PQ and PB must lead to equal latencies under this strategy. That is, even if the same strategy is adopted for these two stimuli, they will not be processed exactly the same way. Although a subject’s set may be iden-
tical in both cases, the actual processing steps that must be engaged in for the two stimuli are not likely identical. Whatever conceivable mechanism one uses to accomplish this task, Tsal's explanation requires the assumption that detecting the presence of two critical features requires exactly the same amount of time as does detecting the presence of one and the absence of the other. It is by no means obvious that this must be true.

Tsal's (1989) proposed explanation for our results has more serious problems. It cannot address the dissociation between the results of peripheral and central cuing. Tsal made the mistake (as other readers of our article have; see Treisman, 1988, Appendix) of assuming that similar sized orienting effects are expected with endogenous and exogenous cues. Ignoring our own (Briand & Klein, 1987) data, other published data that directly compare central and peripheral cues (e.g., Jonides, 1981) give absolutely no reason to expect this to be true. Various factors may influence the relative magnitudes of orienting effects with these two types of cues. For example, there are certainly differences in cue effectiveness, SOA typically used, possible masking effects as well as the occurrence of inhibition of return. The motivational states of subjects may also differ for situations in which they must orient attention to a location rather than those in which they feel that their attention will be pulled around regardless of their intentions. Even if the attentional mechanism oriented by the various cues is identical, these factors alone suggest that identical orienting effects are not necessarily to be expected. We have also suggested that the spotlight metaphor might obscure possible functional differences between orienting processes engendered by endogenous and exogenous cues. In contrast, there is converging evidence that endogenous and exogenous orienting actually have different effects (Klein, Briand, & Hansen, 1987; Posner & Cohen, 1984). We emphasize that chance similarity in measured performance with these two cues may lead to the premature conclusion that identical processes underly that performance. In suggesting that these types of cues may have different effects, we have tried not to go too far beyond available data. Our suggestion may be incorrect, but his matter ought to be resolved empirically. Tsal's dismissal of this possibility is based on nothing more than a general claim that peripheral cues are more powerful. This is probably correct in a rough sense, but detailed explication of how this contributes to our pattern of results is lacking.

The Real Problem With Feature Integration Theory

There is at least one problem for FIT that Tsal does not touch on: There is no independent definition of what constitutes a feature. In practice the preattentive processes that provide features are equated with early neural channels in the visual system. Normally, the assumption is that orientation, motion, color, and form are carried by such channels. But beyond this (in particular within a given dimension), what are the relevant divisions that must be made to classify a given stimulus? This is not at all obvious, and it is not likely that phenomenologically observable qualities can serve as a useful index of activity in these early visual channels. These neural mechanisms can be shown to respond selectively to certain visual properties, but the actual properties that are encoded by the visual system are only indirectly related to the observed phenomenological properties of objects. For example, there may be cells in the visual cortex that are most responsive to an edge of a certain orientation. But it may be incorrect to treat this cell as an edge-detector (Marr, 1982). Rather, it can act as input into a higher level neural system that extracts some more abstract property at the level of the primal or 2½ D sketch. This in turn is insufficient to act as input to pattern recognition mechanisms. In other words, what we subjectively experience as a featural difference between any two patterns may have nothing to do with the representation of attributes of those patterns during early visual processing.

It follows from this that what seems like a trivial surface alteration in stimuli to us may have drastic consequences on how the underlying neural substrate processes a pattern. Conversely, what seems like a major change may leave the underlying neural processes largely unaffected. Stimuli may be very different phenomenologically and yet be represented very similarly in at least some visual channels. This is demonstrated by the finding that a single, small dot can effectively mask a large, concentric sinusoidal grating (see Weisstein & Harris, 1980) yet leave a horizontal or diagonal grating unaffected. Phenomenologically, a dot does not seem to be very similar to a large sinusoidal grating, yet if one thinks in terms of spatial frequency channels at different levels of resolution, Weisstein and Harris showed that these two are actually very similar in terms of their representation by Fourier power spectra. Of course, it is not clear that Fourier analysis into separable spatial frequency components is what is accomplished by the preattentive processes FIT proposes. It ought to be obvious, however, that neural representation of the attributes of a pattern can be radically different from that arrived at through introspection of phenomenological experience.

So how does one know when two given patterns are similar or different on a given dimension, in terms of what actually happens neurally? For now, only gross attributes like the types mentioned seem to be safe bets, but within these, finding features that are appropriately different from one another can be a highly unsophisticated procedure. We simply have no objective basis for stating that two shapes are composed of the same or different features, other than by confirming whether these features produce perceptual effects such as texture segregation. Of course one can assume the veracity of FIT and use it to catalog features, but the fact remains that this list of features will always be in danger of growing out of control so long as no notice is taken of exactly how certain attributes are represented in early visual channels (see Weisstein & Harris, 1980, for a related argument).

Julesz's (1981) texton theory is conceptually almost identical to FIT, but it was developed out of an interest in early visual processes and the neural mechanisms underlying them. Julesz started with the goal of generating textures of stimuli that met certain mathematical constraints (see Julesz, 1981, 1984). The hope was that a precise mathematical definition of patterns would dictate that some are discriminable and
others are not (i.e., textures that differed in only third or higher order statistics ought to be preattentively indistinguishable). The discovery of textures that satisfy these contraints but are in fact distinguishable leads to the conclusion that these textures must differ in either the identity or density of textons (i.e., features) they contain. It is unclear whether this approach has any advantages over FIT in terms of the rigor with which features are defined. On the one hand the mathematiclal contraints used to infer the existence of particular textons rule out a number of possible confounds that may exist in studies used to support FIT (for example, discrimination on the basis of nth order statistics). However, textons themselves do not appear to be defined mathematically but reduce to such entities as blobs, terminators, or line crossings. Though derived from a perceptual approach rather than a cognitive one, this list of primitive features still does not conform to existing neurophysiological evidence. There are no reports of line-crossing detectors, for example. In comparison to the hunt-and-peck strategy dictated by FIT to identify features, Julesz' approach seems more systematic and rigorous on the surface. But unless it can be more demonstrably related to known or proposed neural processing mechanisms, it will remain unclear whether this is a unique approach or simply a version of FIT with a different genesis.

Ignorance of what is actually happening in neural channels also applies to the perception of conjunctions of features. It is not entirely true to say that things like color and form are processed independently. Because there are obviously visual channels that are responsive to particular combinations of form and color (hence the McCollough effect), there must be some neural channels that respond most to a specific combination of these. Does this mean that color and form are not independent attributes? Logic and phenomenology (and the literature!) indicate that they are. Well, even if they are ultimately separable, might not conjunctions be encoded preattentively (Houck & Hoffman, 1986), as Tsal (1989) believes? Although visual channels responsive to particular combinations of form and color do not seem compatible with FIT, as stated earlier, it is incorrect to treat the cells responsive to such combinations, or the channels they compose, as involving feature detection. They may simply feed into higher level systems that then treat these attributes separately. Tsal's citation of Houck and Hoffman (1986) thus remains unconvincing, 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.

Conclusions

Does the bulk of evidence Tsal (1989) has introduced argue against the main tenet of FIT, that preattentive–attentive processing is closely linked to feature–conjunction perception? Converging evidence from a number of paradigms (target search, texture segregation, etc.) has provided strong evidence for the model. If one decides to ignore these other paradigms in favor of more direct manipulations of attention (see Briand & Klein, 1987), however, does Tsal's argument have any more legitimacy? On theoretical grounds we believe the case can be made that there is nothing fatal to FIT in either (a) evidence that features can be facilitated by attention or (b) evidence (sparse as it may be) that features can be conjoined without attention. FIT may have other problems, particularly the lack of a precise definition of what a feature is. In fact, we don't really know what features we are looking for, and speculation on this matter often ignores known facts about the nature of early visual processing obtained from neurophysiology and psychophysics. This does not mean that FIT is wrong, but only that precise specification of the basic features it assumes will be a nontrivial enterprise. Finally, Tsal's empirical critique does not present a strong challenge to the three studies reviewed (or FIT in general), nor has he proposed a distinct, well developed alternative. Converging evidence and parsimony suggests that FIT remains a viable model to explain how features and their conjunctions (objects) are processed.

References


Received November 1, 1988
Accepted November 3, 1988