



# How feature integration theory integrated cognitive psychology, neurophysiology, and psychophysics

Árni Kristjánsson<sup>1,2</sup> · Howard Egeth<sup>3</sup>

© The Psychonomic Society, Inc. 2019

## Abstract

Anne Treisman's Feature Integration Theory (FIT) is a landmark in cognitive psychology and vision research. While many have discussed how Treisman's theory has fared since it was first proposed, it is less common to approach FIT from the other side in time: to examine what experimental findings, theoretical concepts, and ideas inspired it. The theory did not enter into a theoretical vacuum. Treisman's ideas were inspired by a large literature on a number of topics within visual psychophysics, cognitive psychology, and visual neurophysiology. Several key ideas developed contemporaneously within these fields that inspired FIT, and the theory involved an attempt at integrating them. Our aim here was to highlight the conceptual problems, experimental findings, and theoretical positions that Treisman was responding to with her theory and that the theory was intended to explain. We review a large number of findings from the decades preceding the proposal of feature integration theory showing how the theory integrated many ideas that developed in parallel within neurophysiology, visual psychophysics, and cognitive psychology. Our conclusion is that FIT made sense of many preceding findings, integrating them in an elegant way within a single theoretical account.

**Keywords** Selective attention · Visual search · Feature integration · Basic features · Vision

## Introduction

In 1988 the following was written: “[feature integration theory] is certainly too simple and also certainly wrong in some respects.” The author of these remarks, in that year's Bartlett lecture, was Anne Treisman (Treisman, 1988, p. 203). She was in many ways correct in her assessment, but the lecture also made clear how feature integration theory (FIT) had involved a great leap forward for research in visual perception and visual attention.

In science, it is quite common that someone has to be wrong for others to be right. Bold claims, and broad theoretical

perspectives are often held to severe scrutiny. Anne Treisman's Feature Integration Theory (FIT), proposed in its most famous form in Treisman and Gelade (1980), is a good example of this. Many authors have discussed how FIT has fared since it was originally proposed (e.g., Duncan & Humphreys, 1989; Kristjánsson, 2015; Nakayama & Martini, 2011; Quinlan, 2003; Wolfe, 1989, 1994). While most agree that FIT was an enormous step forward for our understanding of visual perception, visual attention, and visual cognition, and should be considered a milestone in these fields, many authors have also tested the predictions of the theory, and often its predictions have not quite held up. Various modifications of FIT or alternatives to its central claims have been proposed (e.g., Eckstein, 1998; Geisler, Perry, & Najemnik, 2006; Hulleman & Olivers, 2017; Palmer, Ames, & Lindsey, 1993; Rensink, 2001; Rozenholtz, 2017; Treisman, 2006; Wolfe & Horowitz, 2017). This is the way science should work. Testable hypotheses are generated from theoretical conceptions that are then either supported or not. FIT has been extensively scrutinized, sometimes supported, and sometimes corrected. And the theory is still remarkably influential, 42 years after its basic tenets were initially proposed in Treisman (1977) and Treisman, Sykes, and Gelade (1977).

---

✉ Árni Kristjánsson  
ak@hi.is

<sup>1</sup> Vision Sciences Laboratory, School of Health Sciences, University of Iceland, Reykjavik, Iceland

<sup>2</sup> School of Psychology, National Research University Higher, School of Economics, Moscow, Russian Federation

<sup>3</sup> Department of Psychological and Brain Sciences, John Hopkins University, Baltimore, MD, USA

While it is typical to refer to FIT as the background to current work, it is less common to approach FIT from the other direction in time – to examine what experimental findings, theoretical concepts, and ideas inspired it. What were the experimental results and theoretical positions that Treisman was responding to with FIT and trying to account for? In this short overview, we try to answer these questions.

## Overview and summary

Feature Integration Theory reflects how major developments within neuroscience and psychophysics on the one hand and cognitive psychology on the other came together. The so-called cognitive revolution occurred during the 1950s. Researchers belonging to this tradition tried to understand basic mental processes, such as decision making, memory, attention, and so on. For example, Neisser, with his visual scanning tasks (1963, 1964) tried to measure the “speed of thought.” Sperling (1960) performed groundbreaking experiments on visual memory, and Broadbent (1958) investigated selective attention. The cognitive revolution can be considered a backlash against the behaviorist tradition that was highly influential as an explanation of behavior during the first half of the 20th century. Discoveries made within cognitive psychology introduced the idea that an *internal* mechanism such as attention could be used to describe how stimuli are selected for processing. Treisman’s earliest work involved major contributions to this literature on the nature of the so-called attentional filter. Another key idea from cognitive psychology that made its way into FIT was that there were different processing stages (Sternberg, 1969) that played distinct roles in stimulus processing.

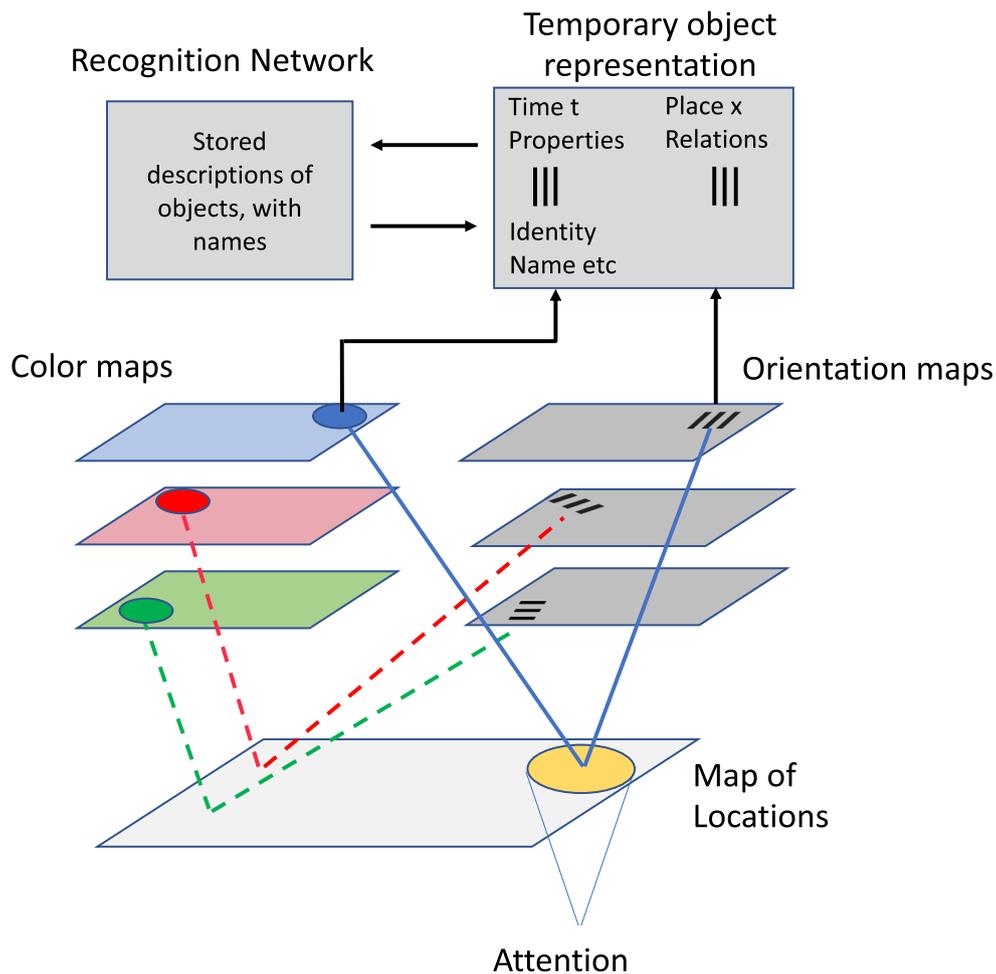
Revolutionary findings regarding the structure of the visual system had already started to emerge from single-cell neurophysiology. Such methods enabled direct investigation of brain processes that had been neglected under the behaviorist tradition. Firstly, it became clear that cells within the visual system had *receptive fields*, which corresponded to areas in the visual field that they were particularly responsive to, while they did not respond to other areas in the visual field (e.g., Hartline, 1940). These studies also revealed cells in the visual system that responded preferentially to certain characteristics of the visual input that fell within their receptive fields and led to the idea that the visual system has neurons tuned to specific features (sometimes called feature detectors). For example, certain cells in the primary visual cortex were shown to respond selectively to stimuli that were oriented in a specific way, moved in a particular direction, or had a certain color. Important discoveries were subsequently made about how these cells seem to involve a hierarchy where the processing becomes gradually more sophisticated (Hubel & Wiesel, 1962).

In psychophysics, selective adaptation studies (Blakemore & Campbell, 1969) revealed how certain visual mechanisms are devoted to particular functions and supported the idea of specialized mechanisms for the detection of particular features in the environment. These discoveries across a variety of fields cried out for an encompassing theoretical framework, and we argue here that FIT was largely successful in filling this role.

FIT is at heart a very simple but elegant theory, which undoubtedly explains its influence and longevity (see Fig. 1). A key notion is that the perceptual process is hierarchical and that the visual scene is initially encoded on a finite number of feature dimensions. They are detected automatically and in parallel across the visual field. The input is first represented on a master map of locations that specifies where things are but not what they are (see Treisman & Souther, 1985; Treisman & Gormican, 1988). In particular it does not indicate which features occupy which locations. This information is then integrated into the “correct compounds” (Treisman, 1977) by focal attention, which “glues” them together (to use a phrase from Treisman and Gelade, 1980). Such integrated entities came to be called object-files (Treisman, Kahneman & Burkell, 1983). Only features that are registered in the same location, within the same attentional focus, can then be encoded as belonging to the same object. Importantly the capacity of this attention integration mechanism is limited. Focal attention is needed for recognition of items that are defined by a conjunction of properties as they do not stand out on a single feature and must be integrated with attention (Treisman, 1977). Interestingly, focal attention is also required for proper spatial localization even of featural information; in the absence of attention, features may be “free floating” and may combine incorrectly, resulting in illusory conjunctions. For example, a red square and a blue triangle might be perceived and reported as a red triangle and a blue square. Note, however, that although the theory applies to featural conjunctions, subsequent research showed that what have been called “propositional conjunctions” (Quinlan, 2003) such as the meanings of verbal stimuli may also be free floating when attention is taxed (e.g., Virzi & Egeth, 1984; see also Prinzmetal & Millis-Wright, 1984).

The theory has been particularly important for the light it has cast on the processes involved in visual search (see Fig. 2). Much of the data used to develop FIT comes from visual search studies, and it is unavoidable for us to spend considerable time in what follows on that topic, in particular ideas that inspired FIT.

For some searches in which a target item can be distinguished from nontarget items on the basis of a single feature (such as a pink target among brown and purple distractors, or an O target among N and T distractors), Treisman and Gelade found that the function relating search time to the number of elements in the display had a slope close to or equal to zero



**Fig. 1** The theoretical framework of the Feature Integration Theory (FIT). A particular stimulus creates activations in feature maps (for color and orientation in this case). Attention then binds features together in the master map of locations, but can only do so for a limited

amount of information in the display (other features are “free-floating”). The temporary object representation is then compared against stored object descriptions. (Based on Fig. 1 in Treisman & Gormican, 1988)

(Fig. 2A; see also Fig. 5). Such functions were thought to indicate parallel processing. However, when targets could be distinguished from nontargets only by a conjunction of features (e.g., a pink O target among nontargets consisting of green Os and pink Ts), they found that target-absent and target-present searches both have slopes of response time against set size that are substantially greater than zero (Fig. 2B; see also Fig. 5), with the former slope approximately twice that of the latter slope. Such functions are said to indicate serial processing where observers search through the display items one-by-one until the target is found (or, in a later version of the model, subjects may serially scan small groups of items; Treisman & Souther, 1985).

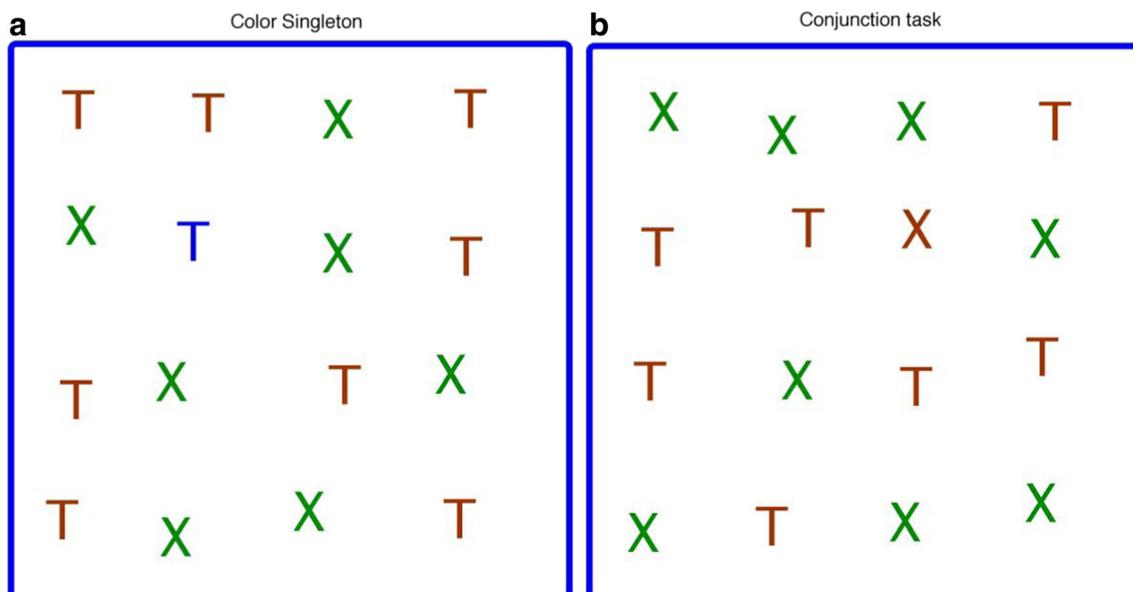
All this is well known in the literature on visual attention. Perhaps less well known today is that there is a large, highly interesting literature from the decades that preceded FIT that provides information about where Treisman’s ideas came from and puts them in context, highlighting what the theory was intended to respond to. Attentional feature integration

was a very elegant concept that made good sense in light of the existing literature at the time.

The above is the short version of the story, but the story is more complicated than this conveys and in what follows we try to cover it in more detail. Given that many parallel developments occurred it is hard to tell a linear story and the choice for presentation order will always be somewhat arbitrary.

### The attentional filter

With the cognitive revolution there was an explosion of research into how we select relevant information from the large amount of information in our perceptual environment. The main concept was that there is more information available in the environment than we can process (Cherry, 1953; Miller, Heise & Lichten, 1951; Webster & Thompson, 1953). The way we deal with this is by selecting certain stimuli over others, and this became known as *selective attention*. An early



**Fig. 2** Different visual search types, similar to those investigated by Treisman. Performance on tasks of this sort formed one of the main backbones of Feature Integration Theory (FIT). **(A)** The target (blue T) is a color singleton and is easily distinguished from the other items on the screen (the brown and green distractors). **(B)** Conjunction search where

the target (brown X) differs from the distractors by its unique *conjunction* of letter identity and color. Detecting the target in A only requires preattentive processing, which is spatially parallel, while detecting the target in B requires attention to integrate its color and form which, according to FIT, results in an item-by-item search

and very influential account of selectivity based on the notion of an attentional filter was developed by Donald Broadbent in his book *Perception and Communication* (Broadbent, 1958). The inspiration for Broadbent's studies came from a problem faced by radio operators during World War II where many operators could be simultaneously speaking on the same channel. Broadbent proposed that selection occurred early in the perceptual process so, for example, if items in the auditory environment or in the visual field were not attended they would not be perceived – they would not get through the filter. For example, early dichotic listening studies where different information was presented to each ear, and participants were to “shadow” one of the sources of information, seemed to indicate that only the attended information made it through.

However, Moray (1959) showed that certain information in the unattended channel sometimes did make it through, such as the participant's name – the famous “cocktail-party” effect. This meant, in other words, that the filter was far from perfect – it was, in fact, quite leaky. Treisman played an important part in the development of this early work on auditory selective attention. For example, using implicit rather than explicit testing methods, Treisman showed that the meaning of unattended words could be registered without them actually reaching conscious perception (Treisman, 1960; Treisman, Squire, & Green, 1974; see also Lewis, 1970; Mackay, 1973). The findings indicating that the filter was leaky accorded well with an important concept proposed by Treisman (1960, 1969). In her view, deep processing of unattended stimuli was the exception

rather than the rule. But she also proposed that the irrelevant information was “attenuated” rather than completely filtered out, which in essence involved a more flexible version of Broadbent's filter theory.

It is very interesting to note that many of the important aspects of FIT were foreshadowed in this earlier stream of work by Treisman. For example, Treisman, Sykes, and Gelade (1977) pointed to her earlier work (Treisman, 1970) as well as that of Day (1967) on dichotic listening, which showed fusion responses in which phonemes from each ear were combined into a single speech item (e.g., “back” and “lack” were heard as “black”). Most tellingly, she considered, “...how the outputs of analyzers are recombined and in particular how they are correctly related to a common source or to different sources. For example, how does one know that it is the “H” that is large and red while the “G” is small and black and not some other combination? One suggestion is that the sensory inputs are labeled by the results of some early stage of analysis, for example, with their spatial location or their time of occurrence, and retain their label throughout analysis” (Treisman, 1969, p. 283–284.) In this paper, which is largely about switching attention between the ears, she was led to speculate that a simple detection response could be triggered by the output of a single analyzer, while conscious experience requires combining the results of several analyzers and is slower and more attention demanding. The seeds of the later discoveries regarding visual attention had clearly been planted.

## Attentional set and spatial attention

Having established that capacity is limited, a natural next question is to wonder just how that capacity is allocated. Does the allocation depend on what your goals may be at a given moment? The dichotic listening studies seemed to suggest that this was the case, but what was the nature of this goal-based selection? Was it the information that was particularly important to the observer or was the filter less selective than this? (Perhaps very salient stimuli could break through the filter even if they were irrelevant to the task at hand.) The concept of how our goals determine what directs our attention became known as attentional set (Gibson, 1941; Mowrer, 1938; Posner et al., 1978), and there has indeed been a lot of research into how our goals influence what is perceived. One of the early questions that was investigated is the extent to which our attention could be directed to specific features or locations in the visual field.

Many studies during the 1960s and 1970s addressed these basic questions and it quickly became clear that attention could be applied to certain stimulus properties that were dependent on participants' goals. Much earlier, researchers had demonstrated that when complex stimuli were briefly presented, report accuracy for stimulus attributes that participants had been instructed to observe was greater than for incidental attributes (e.g., Külpe, 1904; Yokoyama [cited in Boring, 1924]). Mowrer, Rayman, and Bliss (1940) compared performance where observers knew whether an upcoming stimulus would be visual or auditory to when they did not know this, finding that performance became better if observers knew the upcoming modality, strongly arguing for the importance of attentional set focused on the modality of the stimulus. Later, in a similarly motivated experiment (Egeth, 1977, Exp. 6), observers had to judge whether sets of visual stimuli were homogeneous or not. Homogeneous displays consisted of small black discs. Heterogeneous displays were the same except that one disc was replaced with either a large black disc or a small red disc. The results showed that when the single odd item in the set was variable across trials, performance was much worse than when it was constant within a block of trials. These studies imply that participants can be "tuned" to certain stimulus dimensions. Another example is a study by Sekuler and Ball (1977), who had observers judge which of two intervals contained moving dots (the other interval was blank). The luminance of the dots was adjusted until subjects were about 75% correct when the movement was in the same direction on every trial. Under that same luminance condition performance was near chance (55% correct) when direction changed from trial to trial. This suggests that set for a particular direction can improve performance.

Attentional set can also be spatial. Eriksen and Hoffman (1972; see also Colegate, Hoffman, & Eriksen, 1973) investigated spatial selection of particular locations with cues

appearing briefly before target stimuli were presented. They presented a target letter among distractor letters on an imaginary circle, with a cue appearing close to the letter 1–150 ms before the targets appeared. They found that when the cue appeared before target appearance this speeded responses by about 40 ms (they controlled for general alerting effects by also using a noninformative precue that had little speeding effect). These spatial cues considerably improved visual performance at the cued sites, showing how particular locations can be selected by attention.

Egeth (1977) discussed early studies of spatial uncertainty and argued that spatial regularities improve performance just as featural regularities did. Similarly, Posner, Nissen, and Ogden (1978) had subjects fixate at the center of a display, and a spot of light appeared to the left or right of the fixation point. They found that response times to the luminance onset were faster when the target appeared in the location that was more likely to contain the target. These findings on attentional cueing argued that attention could be directed to particular locations in the visual field, the locations that were most useful for performing the task at hand.

But not all findings were in agreement. Shiffrin, McKay, and Shaffer (1976) presented dots in a  $7 \times 7$  matrix and had subjects indicate if one dot was missing from an otherwise filled matrix. Performance was just as good when subjects had to monitor all 49 locations as when they knew which location would contain the missing dot (if it was indeed missing). This argued against the efficacy of attentional set. Other findings indicated, however, that this strong conclusion of Shiffrin et al. (1976) that attention was spatially non-selective was untenable (Bashinski & Bacharach, 1980; Carr & Bacharach, 1976; Keren & Skelton, 1976; Posner et al., 1978; van der Heijden & Eerland, 1973). Additionally, Bashinski and Bacharach (1980) showed importantly that cueing attention away from a location hinders performance at that location.

Overall, these findings on *selective spatial attention* showed that observers could select certain locations, or some particular part of the visual input, again echoing the findings from selective listening studies. These findings were highly important as spatial location plays a major role in FIT. In the master map of locations the beam of attention integrates the features that are attended – gluing them together (Treisman, 1977; Treisman & Gelade, 1980).

## The role of processing stages

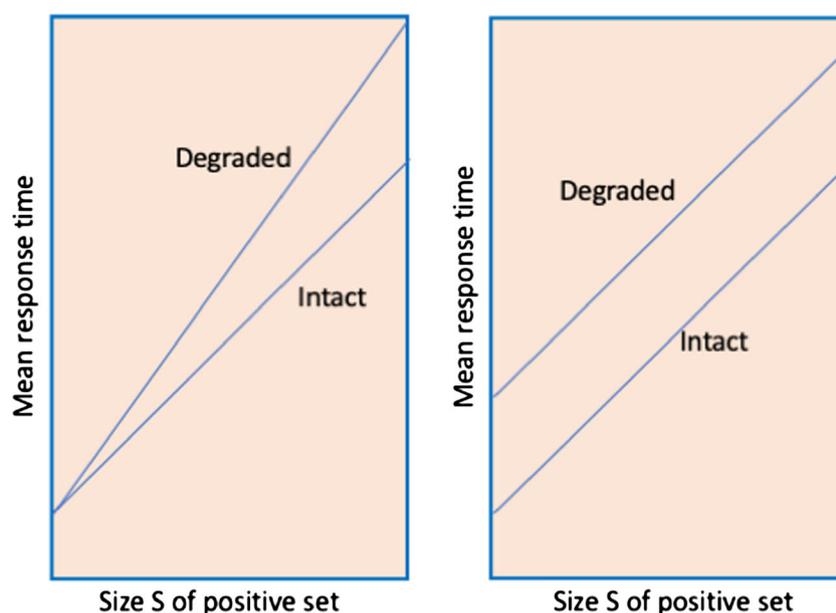
It is a longstanding assumption that there are a number of "stages" underlying behavioral responses. Such stages might include things such as initial stimulus processing, comparison of the stimulus to representations of previous stimuli held in memory, decision making, and response execution. Saul

Sternberg (1969) introduced a method of revealing such stages that he called an extension of Donders's method (Donders, 1868/1969). If we assume the existence of sequentially organized stages, and that stage times are independent of one another, then total reaction time should be the sum of the component stage times. If, further, each stage is responsible for a different aspect of processing, then it may be selectively influenced by a different set of experimental factors. Since the stages are independent, factors that affect different stages will have additive effects on reaction time, while factors that affect the same stage will have interactive effects on reaction time. An example of the application of this logic comes from Sternberg's early studies of character recognition. Sternberg (1966) presented subjects with a single alphanumeric character and they were to indicate whether or not this character was a member of a previously memorized target set. The size of the target set was varied systematically. The results showed that reaction time for both target-present and target-absent trials increased linearly at a rate of about 40 ms for each additional item in the target set. This result was interpreted as indicating a serial, exhaustive scan of the items in the memory set. In a 1967 paper, Sternberg masked the visually presented alphanumeric character with a superimposed noise pattern, comparing search performance with noise to performance without noise (Sternberg, 1967). Figure 3 shows theoretical predictions for the outcome of this experiment if stimulus degradation and size of the target set affect the same stage (left panel), and if they affect separate stages (right panel).

Sternberg's data were almost exactly as shown in the right panel; visual noise had an effect on the overall search times but not the rate of processing as measured by the processing time for each target (i.e., the effect was on the

intercept but not on the slope of the function relating reaction time to the size of the target set held in memory). He concluded that there are at least two separate operations involved in the classification of a character. The first encodes the visual stimulus, and the second compares the encoded representation of the character to the memory representations of the items in the target set.

Although there have been criticisms of the additive-factors approach (e.g., McClelland, 1979; Taylor, 1976), and although Sternberg (1966, 1967) studied memory search, not visual search per se, the logic of this basic method has since been widely applied in visual search studies where it is assumed that processes that are independent of attention have an additive effect on response times (affect intercepts rather than slopes) while the slopes measure the speed of search. Search slopes have therefore been assumed to reflect whether attention is involved in a given task. If the slopes are around zero (flat search slopes) and the response times are thus unchanged with increased set size, the common assumption is that the target can be detected preattentively. If, on the other hand, search times increase with set-size, this is taken to mean that attention moves serially around the array during search. This is assumed to be because the targets do not stand out on a potential feature map. Importantly, if an overall change in mean is observed, it is not thought to reflect any influence of attention if the slopes are not affected. This concept became highly influential and has been dominant in the visual search literature ever since (see discussion in Kristjánsson, 2015, 2016; Wolfe, 2016). The concept found its way into FIT since slopes of response times versus set size that were substantially larger than 0 were considered indicative of serial processing.



**Fig. 3** Two possibilities for the effect of degrading the test-stimulus by visual noise. The figure on the left shows interacting effects of degradation and memory-set size, which would suggest that those two

factors affect a common stage. The figure on the right shows an additive effect of degradation and memory set size, which suggests the two factors affect two different processing stages (based on Fig. 2 in Sternberg, 1967)

There are some well-known caveats concerning this interpretation. While flat search functions are plausibly interpreted as indicating parallel processing, linearly increasing functions do not necessarily indicate serial processing as it is possible to construct parallel models that can produce such data (e.g., Atkinson, Holmgren, & Juola, 1969; Eckstein, 1998; Palmer, 1995; Townsend, 1971, 1972; but see also Townsend, 1990). We return to this point later.

## Perceptual features – neurophysiological studies

FIT was strongly influenced by revolutionary findings within neurophysiology from the 1950s, 1960s, and 1970s. Extracellular physiological studies of individual neuronal firing patterns had revealed that cells in the visual system had *receptive fields*, corresponding to particular areas in the visual field that they were responsive to, but not to other areas in the visual field (e.g., Hartline, 1940). In the cat, Stefan Kuffler (1953) discovered retinal cells that responded in a certain way depending on the input and where in their receptive fields the stimuli appeared (see also Barlow, 1953; Lettvin, Maturana, & McCulloch, 1959). This concept of receptive fields with specific response patterns was important as it entailed that certain regions within the visual field were analyzed by dedicated neurons and that specific neurons were responsible for the analysis of particular features.

Other studies revealed retinal neurons that were tuned to specific directions of motion (Barlow & Levick, 1965) and neurons at the retinal and geniculate levels responsive to particular colors or color combinations, so-called color opponent cells (De Valois, Abramov, & Jacobs, 1966). These discoveries within the retina had counterparts in neurons in the cerebral cortices devoted to motion analysis (Dubner & Zeki, 1971) and color processing (Gouras, 1974; Jameson & Hurvich, 1968; Zeki, 1973), and Michael (1978) discovered double-opponent color cells in visual cortical areas whose processing was even more intricate than that of the opponent cells. Additionally, Barlow, Blakemore, and Pettigrew (1967) found cells in the visual cortex that were sensitive to particular binocular disparity levels, presumably playing a role in depth perception.

Hubel and Wiesel (1959) discovered that cells in the visual cortex responded preferentially to stimuli oriented in a particular way. Their subsequent studies then revealed that stimuli were analyzed hierarchically, at various levels of complexity (see Hubel & Wiesel, 1962). This culminated in the idea that there are specialized units within the primary visual cortex (V1) that are each devoted to the analysis of a certain part of the visual field in terms of basic characteristics such as orientation, size, and color (so-called hypercolumns; Levay, Hubel, & Wiesel, 1975), which could serve functions such as edge-detection across the visual field (Marr, 1976).

Enroth-Cugell and Robson (1966) then showed that retinal cells were particularly sensitive to sinusoidal brightness variation, and the results of Graham and Nachmias (1971) and Albrecht, DeValois, and Thorell, (1980) then raised the possibility that a receptive field in the visual cortex might look something like *Gabor patches*, which involve a convolution of a Gaussian envelope and a sinusoidal grating. This, in turn, raised the intriguing possibility that these cells might perform the equivalent of a Fourier analysis of the image.

The important point is that these neurophysiological findings suggested that the visual system contained specialized detectors for certain key visual attributes of the environment, or *feature detectors* tuned to different basic features. This also implied that stimuli are initially analyzed along functionally separable dimensions, (De Valois & De Valois, 1975; Garner, 1974; Shepard, 1964). Among other things, this accorded well with models of visual recognition that assumed that perception involved assembling features into objects, such as in Selfridge's *Pandemonium* model of visual recognition (Selfridge, 1959; Selfridge & Neisser, 1960).

Based on these findings, Horace Barlow (Barlow, 1972) proposed a highly influential account of perception – the *single-neuron doctrine* where the basic claim was that understanding perception involved understanding the nature of the processing of these single units and their interaction. An obvious functional or neural mechanism for integrating the information had, however, not yet surfaced. Consider, for example, stimuli consisting of a combination of color and shape, say a red circle and a green square. Although these attributes are analyzed by separate areas of the brain, for accurate perception to occur there must, however, be a way to represent which color goes with which shape. This is one aspect of what is known as the *binding problem* (e.g., Rentzperis, Nikolaev, Kiper, & van Leeuwen, 2015; Treisman, 1996). One solution would be to have specific cells responsive to each of the possible combinations of feature values that might be presented. Although there are certainly cells that are responsive to more than one feature dimension (e.g., Gegenfurtner, Kiper, & Fenstemaker, 1996; Pack, Born, & Livingstone, 2003), this is implausible as a general solution to the problem because there is a combinatorial explosion of possibilities. Note, however, that FIT provides a possible solution to this problem, since spatial attention co-localizes the relevant color and shape. There are other proposed solutions, of course, (see, e.g., von der Malsburg, 1981), but FIT provides an appealingly straightforward approach.

An important parallel development along with the neurophysiological findings involved psychophysical studies indicating that there were neurons tuned to particular stimulus characteristics. Blakemore and Campbell (1969) found that when people adapt selectively to a grating of specific orientation and spatial frequency, this affected sensitivity precisely at this spatial frequency and orientation, but not at other spatial

frequencies or orientations. Blakemore, Nachmias, and Sutton (1970) and Blakemore and Nachmias (1971) reported similar adaptation effects on the sensitivity to high-frequency square-wave gratings. These findings argued for specialized mechanisms responsive to particular features, in harmony with the neurophysiological findings. Indeed, the selective adaptation paradigm was seen as providing psychophysicists with a tool somewhat akin to the microelectrode in its ability to reveal the workings of the nervous system (Mollon, 1974).

### Perceptual features – psychological studies

At the same time as these neurophysiological discoveries were being made, parallel psychological investigations into the nature of perceptual features and dimensions were ongoing. One basic problem that was addressed concerned the perceived structure of multidimensional stimuli. (For example, a particular stimulus might be described as having a specific orientation, spatial frequency, and brightness, or, as another example, a color patch might be described in terms of hue, saturation, and brightness.) Torgerson (1958), Attneave (1962), and Shepard (1964) had observed that the perceived structure of such stimuli, as shown by the results of multidimensional scaling studies, depended on whether the underlying

dimensions were integral or separable. The distinction, phenomenologically, is “...between dimensions which can be pulled apart, seen as unrelated, or analyzable, and those which cannot be analyzed but somehow are perceived as single dimensions” (Garner & Felfoldy, 1970, p. 225). In a multidimensional scaling study, separable dimensions will yield a city-block similarity metric while integral dimensions yield a Euclidean metric. An example of a pair of separable dimensions is color and shape; an example of a pair of integral dimensions is Munsell chroma and value (i.e., saturation and lightness). Another way to distinguish dimensions is whether they can be separately attended to or not. It is, for example, easy to attend to shape and color independently, but not color saturation and hue (Fig. 4). (Note that although color and shape actually have multiple constituent dimensions, such as the hue, chroma, and value of the Munsell system, in some contexts they are treated as simple dimensions, as we did when we described separable stimuli in terms of color and shape.)

Garner (1974) developed a set of three speeded classification tasks that were designed to assess the extent to which dimensions were treated independently of one another. These studies had a large influence on Treisman and they are worth some detailed inspection here. The studies in Treisman (1977), which is a key paper in the development

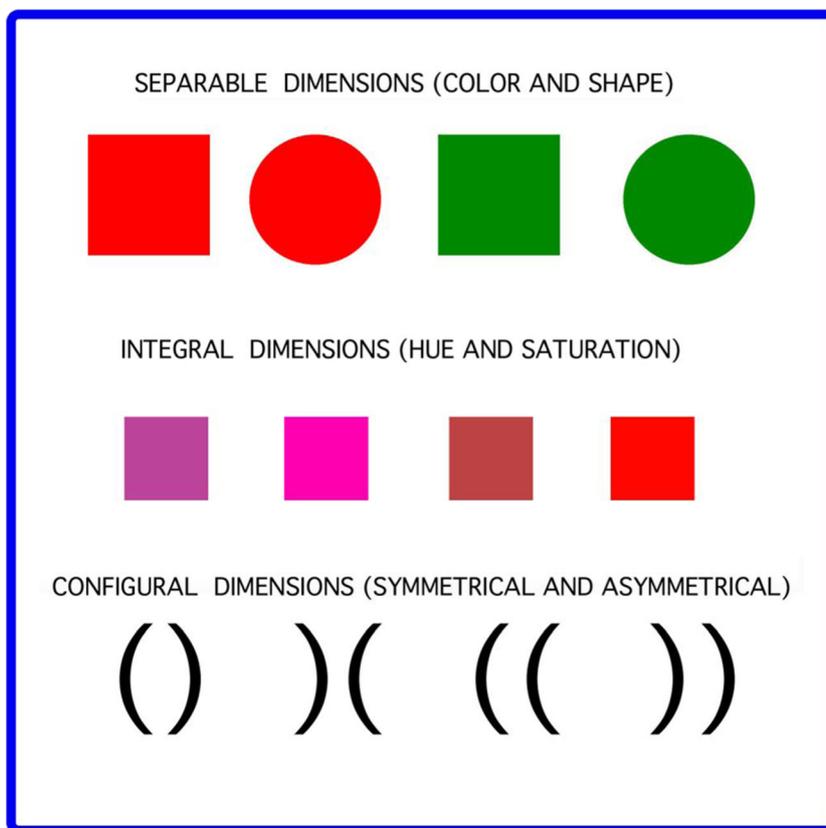


Fig. 4 Different dimensions types, defined by whether they are integral, separable or configural

of FIT, did not, for example, involve visual search, but were a response to Garner's studies. What follows is a simplified version of a hypothetical experiment in this tradition. Consider a set of four stimuli, created by the factorial combination of two dichotomous dimensions, say a bar that could be either vertical or horizontal and either red or blue. One task requires unidimensional discrimination. The test object is held constant on one dimension and subjects are asked to quickly identify the other dimension. For example, they may be asked to indicate whether a horizontal bar is red or blue. With such a stimulus set, there are clearly four such unidimensional tasks. Another task involves speeded classification of stimuli that vary in both dimensions, but the dimensions are correlated. For example, the horizontal red bar is to be discriminated from the vertical blue bar (and, in a separate condition, the vertical red bar is to be discriminated from the horizontal blue bar). The third task also involves both dimensions of variation, but now the variation is orthogonal, and subjects are to classify stimuli according to just one dimension. For example, subjects are instructed to make one response to red stimuli and another response to blue stimuli, and this response assignment holds regardless of whether the stimulus is horizontal or vertical. In essence, this is a selective attention condition; subjects are ignoring the dimension of orientation. Similarly, in another condition they respond on the basis of orientation and ignore color.

Garner and his colleagues (e.g., Garner & Felfoldy, 1970; Pomerantz & Garner, 1973) found that the results of these three tasks differed depending on the dimensions that were paired. One pattern of results was that (a) performance on the orthogonal classification task was just as fast as in the single dimension task, indicating highly efficient selectivity and (b) classification speed was not improved by combining the dimensions redundantly. Pairs of dimensions that yielded this pattern were those (like the color and shape in our example) that yielded a city-block metric in multidimensional scaling. These are separable dimensions. Another pattern of results was found with stimuli such as color chips that varied in lightness and saturation. In such cases (c) performance on the orthogonal classification task was slower than on the single dimension task, indicating inefficient selectivity, and (d) classification speed was improved by combining the dimensions redundantly. These pairs of dimensions, which yield a Euclidean similarity metric in multidimensional scaling, are integral. Other patterns of results may also occur.

In one striking case, Pomerantz and Garner (1973) used stimuli with two stimulus dimensions (left and right element), each with two levels (left and right curvature of a parenthesis (See Fig. 4)). Subjects had grave difficulties responding selectively to the left or the right parenthesis; however, they showed no benefit in a redundantly correlated condition. This was taken to exemplify *configural* processing. The left and right parentheses were not truly integral, but the data showed that physically separate stimulus elements may be

treated as belonging to a single emergent configuration. It is of some interest to note that in a second condition, when the elements were perpendicular to one another, rather than parallel, selectivity was excellent; there was no interference at all from an irrelevant curved line.

This body of research was well known to Treisman. In Treisman and Gelade (1980, p. 132) it is suggested that the new findings of that paper (with respect to search, texture segmentation, etc.) "...offer a new set of criteria for determining which features are perceptually 'separable,' which may be added to the criteria listed by Garner" (Treisman & Gelade, 1980, p. 132).

### Early studies of visual search and visual scanning: The question of parallel versus serial processing

From reading the literature on visual search since FIT was proposed, one might be forgiven for thinking that Anne Treisman and her collaborators had invented visual search. This is far from being the case. Treisman's theory entered into a rich literature that contained many examples of visual search findings that would constitute a core part of her theory. One key question involved uncovering whether observers scan or search the environment serially or in parallel. How many items can visual attention process at a given moment in time? Contrast, for example the search types in Fig. 2A on the one hand and the search type in Fig. 2B on the other.

During the 1960s and 1970s many studies on visual search were published. Observers were, for example, asked to find a certain letter in a set of other letters, a digit among letters or vice versa, and so on. Ulric Neisser was one of the earliest researchers to investigate such visual scanning (Neisser, 1963, 1964). Neisser's aim was to measure the "the speed of thought" by measuring how long it took to make a decision whether a candidate item was a target or not. The general logic was simple. Imagine a vertical list 50 lines long, each line being a string containing six alphanumeric characters. Subjects might be given the task of searching from the top of the list down until they find the string that contains a critical property and then pressing a switch. Response time was measured as a function of the position of the critical item in the 50-item list. Not surprisingly, the function relating response time to position in the list was strongly linear; the slope of that function was the dependent variable of chief interest. Also of interest was the effect of different instructions on search slope. For example, the "critical feature" might be the presence of the letter Z; the subject had to find the one line that contained a Z. Then again, the critical feature might be the absence of a Z; the subject had to find the one line that did not contain a Z. (Detection of absence was slower than detection

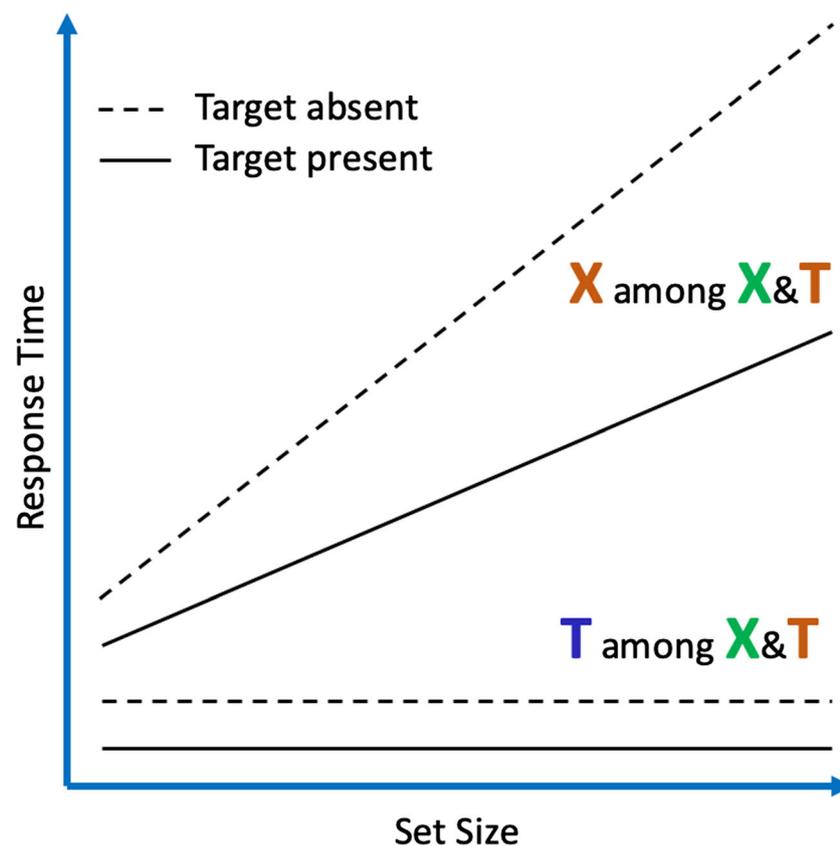
of presence, presumably because the former required an exhaustive analysis of each line.) Perhaps the most interesting result of this research was that, with extensive practice, subjects could search for ten targets as easily as they could search for one target (e.g., Neisser, Novick & Lazar, 1963). They proposed that a set of “analyzers” (in the sense of the Pandemonium model of Selfridge, 1959) are devoted to the important properties of the targets, but importantly that with practice such mechanisms can begin to operate in parallel.

Neisser argued that “some combination of feature detectors is presumably sufficient [...] to stimulate activity in some subsystem sensitive to the letter that is sought” (Neisser, 1964; p 97). Neisser emphasized that it was almost certainly not the case that observers processed each letter to the point of identification to determine if it is a target – Neisser’s observers claimed that they did not see the letters that were not specifically attended but that they “see only a blur” (Neisser, 1963, p. 380) until the target jumps out, and Neisser argued that this was evidence for a preattentive processing stage.

The large practice effects Neisser and colleagues found on scanning rates, and specifically the ability after practice to search for ten targets as quickly as for one target, argues

against a strictly serial notion of the search process. Indeed, memory scanning here appears to be parallel, which stands in contrast to the serial processing suggested by Sternberg’s (1966, 1967) results. But note that this disparity in outcomes refers to memory scanning, not visual scanning. The visual scanning in Neisser’s experiments was relentlessly serial, but this is entirely reasonable, as his experimental paradigms were highly biased in favor of serial scanning since his stimuli consisted of linearly organized lists of letters and subjects were instructed to search the list from top to bottom (Fig. 5).

How is visual search accomplished when the organization of displays and experimenter’s instructions do not so strongly bias subjects towards serial processing? Egeth, Jonides, and Wall (1972) scattered alphanumeric characters randomly on the circumference of an imaginary circle centered at fixation. In one experiment subjects had to indicate if a 4 was present or not in a display; the other elements were Cs, varying in number from one to five. The slope of the target-present and target-absent functions were both close to zero, indicating parallel processing. The same result was obtained when the task was changed from search to same versus different discrimination. Displays varied in size from two to six and the elements were



**Fig. 5** Typical data patterns from feature-based search and conjunction-based search. For searches such as shown in Fig. 2A, where the target can be distinguished by a single feature (a blue T among green X’s and brown T’s), response times tend to be unaffected as set size increases, while for

searches requiring conjunctions of features (a brown X among green X’s and brown T’s), search times typically increase linearly with set size. These set size effects formed a core part of Treisman’s feature integration theory

either homogeneous, (all Cs or all 4s), or heterogeneous (one 4 and the others all Cs, or one C and the others all 4s). Reaction times on both same and different trials were independent of display size (see also Connor, 1972; Donderi & Zelnicker, 1969). When the task was made more complex (search for any digit in a background composed of varied letters), target-absent reaction times increased with display size (slope = 28 ms/item), strongly suggestive of serial processing. However, over four days of practice that slope was reduced to a nonsignificant 3 ms/item. Intriguingly, the slope for target-present trials in this task requiring a conceptual categorization (digit vs. letter) was essentially flat even on day 1.

Processing mode in search experiments proved to be highly dependent on the circumstances of testing. For example, using tachistoscopic exposures, Estes and Taylor (1964, 1966) also examined search for scattered items. They had observers indicate which of two predesignated letters was present in arrays of letters presented for just 50 ms. They found that only a small number of items can be apprehended in 50 ms (cf. Sperling, 1960), but also that the number of elements processed appeared to increase with increased set size. They argued for a serial selection model. The number effectively perceived is small because the "...elements registered by the sensory apparatus must be scanned one at a time by a central mechanism, and the number scanned is limited by the decay of the stimulus traces activated by a brief display" (Estes & Taylor, 1964, p. 451). These results (see also Estes & Taylor, 1966) supported a limited capacity conception of visual search.

Other investigators came down on the side of parallel processing at the perceptual level. For example, Gardner (1973; see also Eriksen & Spencer, 1969; Shiffrin & Gardner, 1972) showed that subjects could search a display for a target equally well regardless of whether the items were presented simultaneously or sequentially. These data were taken to support a class of models (independent parallel channels) that assumed no limit on perceptual capacity. (For a pre-FIT review of this literature, see Egeth, 1977). A particularly important contribution was made by Hoffman (1979), who proposed a detailed two-stage model that included both parallel and serial stages (see also Rumelhart, 1970). Hoffman's model was partly based on Neisser's model of visual search. An initial "preattentive" stage was responsible for segmentation of the scene. This stage was spatially parallel, but error prone. The outcome of this first stage guided a second stage that performed fine discriminations on a single object at a time. In Hoffman's version the first stage compared in parallel each item in the display to all items in the memory set; the outcome was a similarity measure for each display item. The display items are then transferred in order of decreasing similarity to a comparison process. Hoffman's model obviously foreshadows many features of FIT.

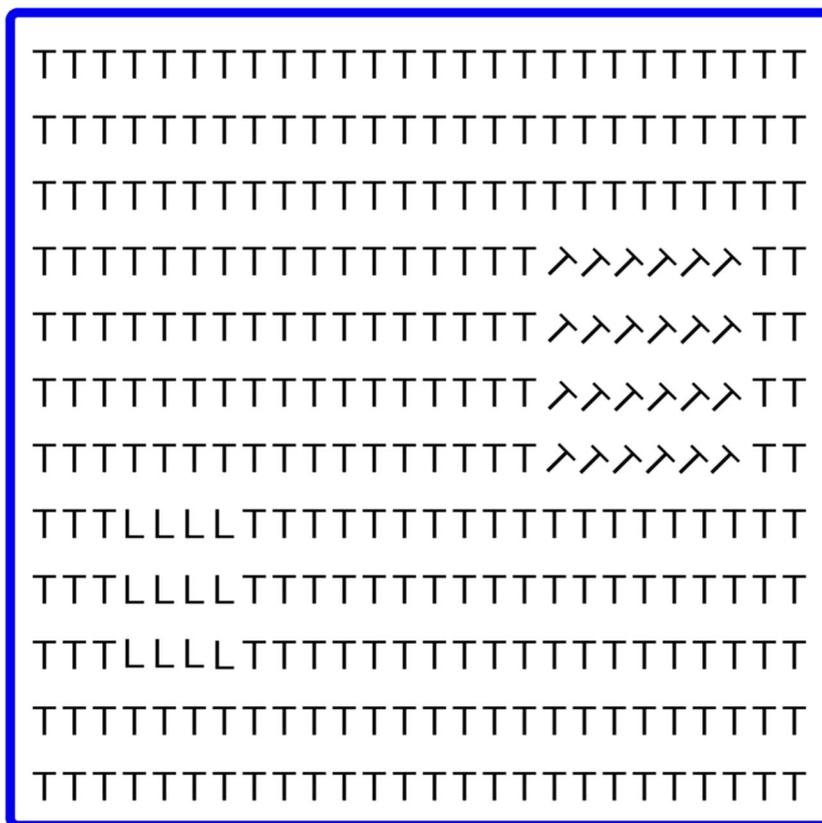
To summarize, rather than trying to determine whether selection in multielement arrays occurs serially or in parallel, instead the questions increasingly became *when* is selection serial and when is it parallel, and *which aspects of processing* are serial and which are parallel, which are some of the major questions addressed by FIT.

We earlier alluded to modifications of FIT or alternatives to its central claims. Although our focus is on the antecedents of FIT, we would like to at least briefly point readers who may be new to the field to some of the issues that have been raised since the theory's initial articulation.

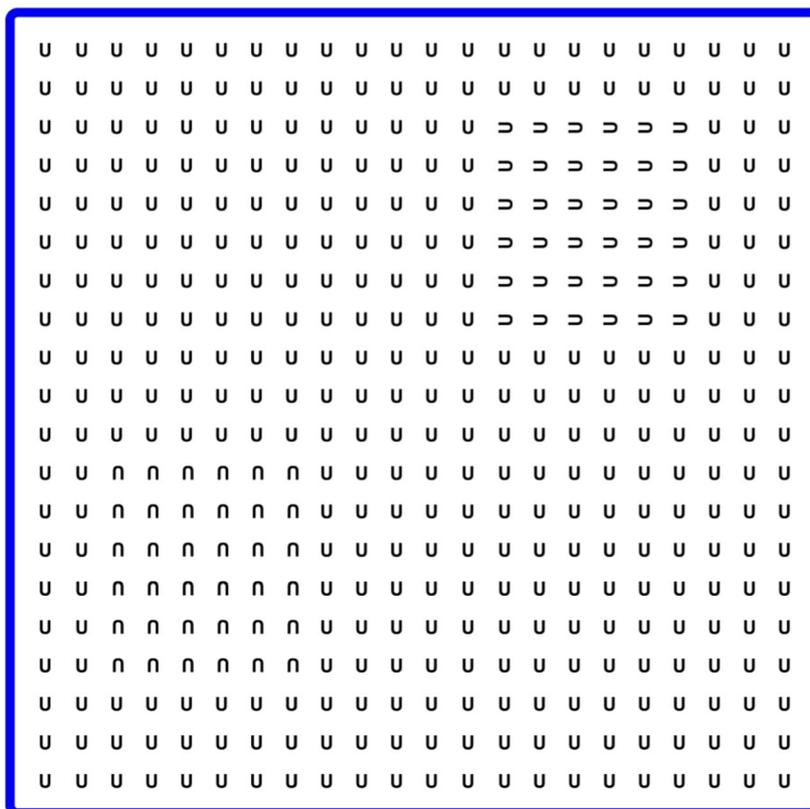
The flat search functions observed in feature search tasks seem to be pretty clear evidence of parallel processing. What is more problematic is the interpretation of the steep functions relating response time to display size in search tasks involving conjunctions of features. It turns out that a variety of parallel processing models can account for the increasing reaction time functions (for pre-FIT examples, see Atkinson, Holmgren, & Juola, 1969; Townsend, 1971.) The simplest to think about is a fixed capacity model; the larger the display size the less the capacity that can be devoted to any particular element in the display. This will increase the amount of time required to process elements – even if they are handled simultaneously. More recent papers have considered the factors that affect search and that must be considered in evaluating models. These include target-distractor similarity (e.g., Duncan & Humphreys, 1989), retinal eccentricity, and density of display elements (e.g., Carrasco, Evert, Chang, & Katz, 1995), and the role of eye movements as display size increases (e.g., Zelinski, Sheinberg, & Bülthoff, 1993). When these factors were carefully controlled, Eckstein (1998) found that the conjunction-feature dichotomy was due to noisy processing of features in the visual system. His results did not support the idea of a serial attentional mechanism that binds information across feature dimensions (see also Jenkins, Grubert, & Eimer, 2017; Palmer, Verghese, & Pavel, 2000; Rozenholtz, 2017).

Another direction that subsequent researchers have taken is to reconsider the nature of the proposed random, self-terminating search mechanism. Consider Panel B in Fig. 2. If participants are told that the target is the brown X and they search randomly, the target would be found on average after searching about half of the 16 items in the display. If, however, the subjects could affect top-down control of their search they could effectively ignore the green items and just examine the brown items, the target could be found after search about half of the 8 brown items. If display size were varied, this kind of selective search would result in a slope half as steep as a completely random search. Egeth, Virzi, and Garbart (1984) showed that such selectivity was indeed possible. The first, parallel stage of search can provide information that effectively guides attention towards locations that are likely to contain the target. Wolfe and his colleagues (e.g., Wolfe, Cave, & Franzel, 1989) developed this idea into the influential guided

**a** Effortless Texture Segmentation



**b** First-order and second-order statistics



◀ **Fig. 6** (A) “Effortless” texture segregation. The patches with Ts that are oriented away from upright are easily segregated from the other stimuli. (B) Two six by six item patches are embedded within the other stimuli – which contain identical first-order statistics but not second-order statistics (upper right) or identical first- and second-order statistics (lower left). Only the former are easily discernable from the background. Stimuli in (B) are based on Julesz (1975)

search model (see Treisman & Sato, 1990, for the incorporation of this basic idea into FIT, and see Wolfe & Horowitz, 2017, for a recent review of attentional guidance).

We have so far emphasized the steep slopes for conjunction searches found by Treisman and her colleagues, and successfully replicated by others (e.g., Egeth et al., 1984). However, using different display modes (in particular, computer monitors rather than tachistoscopic presentations of printed displays) there are circumstances in which conjunction searches have been found to yield flat slopes (e.g., McLeod, Driver, & Crisp, 1988; Nakayama & Silverman, 1986; Wolfe, Cave, & Franzel, 1989). One reason the notion of search guidance is so important to the development of FIT-like theories is that it provides a way to explain such evidence of parallel processing. The basic idea is that if stimuli are discriminable enough the preattentive process can for example, work on color and shape simultaneously. For Treisman and Sato (1990), guidance operated by inhibition of features (in the models proposed by Wolfe and his colleagues, the process works by enhancement of target features). Thus, in Fig. 2B inhibition of the nontarget features green and T would leave the brown X as “the sole survivor on the battlefield” (p. 470); it could be found directly without examining any other items and thus the slope might well be zero.

As a final point about reaction time functions, we should mention that not all feature searches yield flat slopes. When target-non-target similarity is high, reaction time can increase significantly with the number of elements in the search display. This was observed by Treisman and Gormican (1988) and was incorporated into a revised version of FIT. They assumed that subjects checked groups of items in parallel with group size depending on target-nontarget discriminability. Effects of target-nontarget discriminability are important to other models as well (e.g., Duncan & Humphreys, 1989). For an extended recent discussion of search models and, in particular, the serial/parallel processing issue see Liesefeld and Müller (in press).

While a discussion of visual search theories could, in essence, go on forever, our hope is that we have simultaneously been able to convey the importance of Treisman’s contributions while still preserving the fact that FIT was not the “big bang” of the visual search literature (Wolfe, 2003). Her contributions entered into a rich literature that is still vibrant.

## Comparison and texture segregation tasks

While visual search studies are a cornerstone of FIT, such tasks were not the only experimental paradigms used by Treisman to support the theory. This multifaceted testing is a major strength of Treisman’s theory (Driver, 2001). In fact, one of the clearest explications of FIT can be found in Treisman (1977), which involves studies on sequential matching tasks. These studies were inspired by Garner’s studies of the nature of different feature dimensions (Garner, 1974), discussed above. It is not particularly well known that these studies were a major source of inspiration for FIT, where Garner attempted to match the channels approach from neurophysiology to the viewpoint of cognitive psychology. These studies are, regrettably, little-known today.

Here too, Treisman’s work fit in well with an existing literature. Egeth (1966) carried out experiments on the matching of pairs of simultaneously presented stimuli that differed in color, shape, and the orientation of a line inscribed within them. The questions at issue included whether comparisons are made between overall (template-like) stimulus representations or between the individual dimensions constituting each stimulus and, if the latter is the case, whether dimensions are processed serially or in parallel, and whether those comparisons are self-terminating or exhaustive. The results favored a serial self-terminating model. However, the study required subjects to ignore variation in irrelevant dimensions, which they proved unable to do. In a study designed to overcome this problem, Hawkins (1969) came down on the side of a parallel self-terminating model. Other investigators explored how subjects determined whether a given multidimensional stimulus matched a specified criterion such as “large, red, square” (e.g., Nickerson, 1967; Saraga & Shallice, 1973).

Another source of inspiration for FIT came from studies of texture segregation. In a typical texture segregation task a large display containing many elements is presented. Subjects may be asked to indicate the location of a segment of the display that differs from the other segments (Fig. 3A). For example, three quadrants of a display might contain Ts while the remaining quadrant contains Ts all rotated 45° to the right. Alternatively, the odd quadrant might contain upright Ls. Early work by Beck and his colleagues (e.g., Beck, 1966, 1972; Beck & Ambler, 1973) showed that elements that differ from one another on the basis of a simple feature difference (such as the orientation difference between Ts and tilted Ts) are readily distinguished and are presumably picked up preattentively. However, grouping is weak when the elements do not differ on a primitive feature, but only in the arrangement of those features (like the horizontal and vertical lines that make up upright Ts and Ls). Such discriminations require more effortful processing by focal attention.

But Beck also found that the rules governing texture segregation differed from those governing ratings of similarity of

items. Beck (1972) proposed that feature density detection (the number of features in a given area) determined performance. Julesz (1975) noted that texture segregation could be effortless, assuming that it occurred preattentively. Julesz proposed initially that simple rules based on statistics determined how easily textures could be segregated (Julesz, 1975). A first-order statistic corresponds to the average brightness of a texture. It is measured by calculating the probability that a point thrown randomly on, say, a texture made up of black and white dots (think pixels) lands on a black dot. A second-order statistic corresponds to the clumpiness of a texture. It is assessed by calculating the probability that dipoles (line segments or pins) of all possible lengths and orientations thrown randomly on a texture would land with both ends on black dots. So, for example, stimuli containing identical first-order statistics but different second-order statistics (the 90° rotated Us in the upper right patch in Fig. 6B), can easily be distinguished from the background, but stimuli containing identical first-order *and* second-order statistics could not (the 180° rotated Us in the lower left part of Fig. 6B). Later, counterexamples surfaced (e.g., Martin & Pomerantz, 1978), which prompted Julesz to revise his proposal, focusing more on feature density ideas in the spirit of Beck's proposals, in his texton theory (Julesz, 1981), which involved fundamental elements that he proposed had dedicated detectors in the visual system, in that sense echoing Treisman's proposals.

## Conclusions

According to FIT, attention was the glue that bound different features together. If we stick with her adhesive metaphor, we could argue that Treisman's FIT was the glue that bound together a number of disparate disciplines and theoretical conceptions that had developed in parallel over the decades that preceded its first formulation. It could be argued that because of FIT, attention now plays a major role in any account of visual perception. FIT may, among other things, have been instrumental in helping to make the concept of attention "respectable" even among those who do research on the responses of single units or who study low-level psychophysics (see, e.g., Buracas & Boynton, 2007; Kristjánsson et al., 2013; Martínez-Trujillo & Treue, 2002; Moran & Desimone, 1985; Reynolds, Pasternak, & Desimone, 2000; Yeshurun & Carrasco, 1998). Attention has consequently been shown to play a large role at almost every step of perceptual processing (Kristjánsson, 2006; Noudoost et al., 2010).

While many aspects of FIT have survived the test of time, others have not. It is not our aim here to provide any overview of this, and we can point to others (Wolfe 1989; Nakayama & Martini, 2011; Kristjánsson, 2015; Treisman, 1988) for discussion on how FIT has fared since it was originally proposed. The theory undoubtedly involved a brilliant solution to many

of the problems that had been posed in the literature, involving such questions as what grabs our attention, and why. Findings had indicated that not all stimuli are equal in capturing our attention and that this depends on the context that they appear within. Another major problem that FIT had a solution to was how we perceive whole objects rather than free-floating features. The elegant solution is that attending to stimuli assembles them, brings them into attentional focus, and makes sense of them. For them to be assembled they need to be in the same location in the feature map. But these attentional processes tax attentional capacity. As explained earlier, preattentively detected features will stand out if they are unique against the background stimuli, but to make sense of them they need to be integrated with other stimuli. This is the job of attention. This is an elegant and beautiful concept that has shaped the subsequent four decades (and counting) of research on attention and perception.

**Acknowledgements** We thank Daniel Kahneman, Jeremy M. Wolfe, and two anonymous reviewers for highly useful comments on earlier versions of this manuscript.

## References

- Albrecht, D. G., De Valois, R. L., & Thorell, L. G. (1980). Visual cortical neurons: Are bars or gratings the optimal stimuli?. *Science*, 207(4426), 88–90.
- Atkinson, R. C., Holmgren, J., & Juola, J. F. (1969). Processing time as influenced by the number of elements in a visual display. *Perception & Psychophysics*, 6(6), 321–326.
- Attneave, F. (1962). Perception and related areas. In S. Koch (Ed.), *Psychology: A study of a science, Vol. 4. Biologically oriented fields: Their place in psychology and in biological science* (pp. 619–659). New York, NY: McGraw-Hill.
- Barlow, H. B. (1953). Summation and inhibition in the frog's retina. *The Journal of Physiology*, 119(1), 69–88.
- Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology?. *Perception*, 1(4), 371–394.
- Barlow, H. B., & Levick, W. R. (1965). The mechanism of directionally selective units in rabbit's retina. *The Journal of Physiology*, 178(3), 477–504.
- Barlow, H. B., Blakemore, C., & Pettigrew, J. D. (1967). The neural mechanism of binocular depth discrimination. *The Journal of Physiology*, 193(2), 327–342.
- Bashinski, H. S., & Bacharach, V. R. (1980). Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations. *Perception & Psychophysics*, 28(3), 241–248.
- Beck, J. (1966). Effect of orientation and of shape similarity on perceptual grouping. *Perception & Psychophysics*, 1(5), 300–302.
- Beck, J. (1972). Similarity grouping and peripheral discriminability under uncertainty. *The American Journal of Psychology*, 85(1), 1–19.
- Beck, J., & Ambler, B. (1973). The effects of concentrated and distributed attention on peripheral acuity. *Perception & Psychophysics*, 14(2), 225–230.
- Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *The Journal of Physiology*, 203(1), 237–260.

- Blakemore, C., & Nachmias, J. (1971). The orientation specificity of two visual aftereffects. *The Journal of Physiology*, *213*(1), 157–174.
- Blakemore, C., Nachmias, J., & Sutton, P. (1970). The perceived spatial frequency shift: Evidence for frequency-selective neurones in the human brain. *The Journal of Physiology*, *210*(3), 727–750.
- Boring, E. G. (1924). Attributes and sensation. *American Journal of Psychology*, *35*, 301–304.
- Broadbent, D. (1958). *Perception and communication*. Oxford: Pergamon Press.
- Buracas, G. T., and Boynton, G. M. (2007). The effect of spatial attention on contrast response functions in human visual cortex. *Journal of Neuroscience*, *27*, 93–97. <https://doi.org/10.1523/JNEUROSCI.3162-06.2007>
- Carr, T. H., & Bacharach, V. R. (1976). Perceptual tuning and conscious attention: Systems of input regulation in visual information processing. *Cognition*, *4*(3), 281–302.
- Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Perception & Psychophysics*, *57*(8), 1241–1261.
- Cherry, C. (1953). Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America*, *25*, 975–979.
- Colegate, R. L., Hoffman, J. E., & Eriksen, C. W. (1973). Selective encoding from multielement visual displays. *Perception & Psychophysics*, *14*(2), 217–224.
- Connor, J. M. (1972). Effects of increased processing load on parallel processing of visual displays. *Perception & Psychophysics*, *12* (1B), 121–128.
- Day, R. S. (1967). Fusion in dichotic listening. *Psychonomic Bulletin*, *1*, 18.
- De Valois, R. L., Abramov, I., & Jacobs, G. H. (1966). Analysis of response patterns of LGN cells. *Journal of the Optical Society of America*, *56*(7), 966–977.
- De Valois, R.L. & De Valois, K.K. (1975) Neural coding of color. In E.C. Carterette & M.P. Friedman (eds.), *Handbook of perception: Vol 5*, (pp. 117-166). New York: Academic Press.
- Donderi, D. C., & Zelnicker, D. (1969). Parallel processing in visual same-different decisions. *Perception & Psychophysics*, *5*(4), 197–200.
- Donders, F. C. (1868). On the speed of mental processes. Translated from the Dutch by W. G. Koster. In Koster, W. G. (Ed.). (1969). Attention and performance II. *Acta Psychologica*, *30*, 412–431.
- Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, *92*(1), 53–78.
- Dubner, R., & Zeki, S. M. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Research*, *35*(2), 528–532.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458.
- Eckstein, M. P. (1998). The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science*, *9*(2), 111–118.
- Egeth, H. (1977). Attention and preattention. In G. H. Bower (Ed.), *The Psychology of learning and motivation* (Vol. 11, pp. 277–320). New York: Academic Press.
- Egeth, H., Jonides, J., & Wall, S. (1972). Parallel processing of multielement displays. *Cognitive Psychology*, *3*(4), 674–698.
- Egeth, H. E. (1966). Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, *1*(4), 245–252.
- Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(1), 32–39.
- Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology*, *187*(3), 517–552.
- Eriksen, C. W., & Hoffman, J. E. (1972). Some characteristics of selective attention in visual perception determined by vocal reaction time. *Perception & Psychophysics*, *11*(2), 169–171.
- Eriksen, C. W., & Spencer, T. (1969). Rate of information processing in visual perception: Some results and methodological considerations. *Journal of Experimental Psychology*, *79*(2), 1–16.
- Estes, W. K., & Taylor, H. A. (1964). A detection method and probabilistic models for assessing information processing from brief visual displays. *Proceedings of the National Academy of Sciences*, *52*(2), 446–454.
- Estes, W. K., & Taylor, H. A. (1966). Visual detection in relation to display size and redundancy of critical elements I. *Perception & Psychophysics*, *1*(1), 9–16.
- Gardner, G. T. (1973). Evidence for independent parallel channels in tachistoscopic perception. *Cognitive Psychology*, *4*, 130–155.
- Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Lawrence Erlbaum Associates.
- Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*(3), 225–241.
- Gegenfurtner, K. R., Kiper, D. C., & Fenstemaker, S. B. (1996). Processing of color, form, and motion in macaque area V2. *Visual neuroscience*, *13*(1), 161–172.
- Geisler, W. S., Perry, J. S., & Najemnik, J. (2006). Visual search: The role of peripheral information measured using gaze-contingent displays. *Journal of Vision*, *6*(9), 858–873.
- Gibson, J.J. (1941). A critical review of the concept of set in contemporary experimental psychology. *Psychological Bulletin*, *38*, 781–815.
- Gouras, P. (1974). Opponent-colour cells in different layers of foveal striate cortex. *The Journal of Physiology*, *238*(3), 583–602.
- Graham, N., & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: A comparison of single-channel and multiple-channels models. *Vision Research*, *11*(3), 251–259.
- Hartline, H. K. (1940). The receptive fields of optic nerve fibres. *American Journal of Physiology*, *130*, 690–699.
- Hawkins, H. L. (1969). Parallel processing in complex visual discrimination. *Perception & Psychophysics*, *1969*(1), 56–64.
- Hoffman, J. E. (1979). A two-stage model of visual search. *Perception & Psychophysics*, *25*(4), 319–327.
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, *148*(3), 574–591.
- Hubel, D.H. & Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, *160*, 106–154.
- Hulleman, J., & Olivers, C. N. (2017). The impending demise of the item in visual search. *Behavioral and Brain Sciences*, *40*, 1–69.
- Jameson, D., & Hurvich, L. M. (1968). Opponent-response functions related to measured cone photopigments. *Journal of the Optical Society of America*, *58*(3), 429–430.
- Jenkins, M., Grubert, A., & Eimer, M. (2017). Target objects defined by a conjunction of colour and shape can be selected independently and in parallel. *Attention, Perception, & Psychophysics*, *79*(8), 2310–2326.
- Julesz, B. (1975). Experiments in the Visual Perception of Texture. *Scientific American*, *232* (April), 34–43.
- Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature*, *290*(5802), 91–97.
- Keren, G., & Skelton, J. (1976). On selecting between theories of selective attention. *Attention, Perception, & Psychophysics*, *20*(1), 85–86.
- Kristjánsson, Á. (2006). Rapid learning in attention shifts – A review. *Visual Cognition*, *13*, 324–362
- Kristjánsson, A. (2015). Reconsidering visual search. *i-Perception*, *6*(6), 2041669515614670.
- Kristjánsson, A. (2016). The slopes remain the same: Reply to Wolfe (2016). *i-Perception*, *7*(6), 2041669516673383.

- Kristjánsson, Á., Heimisson, P.R., Róbertsson, G.F. & Whitney, D. (2013). Attentional priming releases crowding. *Attention, Perception & Psychophysics*, *75*(7), 1323–1329.
- Kuffler, S.W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, *16*, 37–68.
- Külpe, O. (1904). Versuche über Abstraktion. *Bericht über den Ie Kongress für Experimentale Psychologie*, 56–58.
- Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the IRE*, *47*(11), 1940–1951.
- LeVay, S., Hubel, D. H., & Wiesel, T. N. (1975). The pattern of ocular dominance columns in macaque visual cortex revealed by a reduced silver stain. *Journal of Comparative Neurology*, *159*(4), 559–575.
- Lewis, J. L. (1970). Semantic processing of unattended messages using dichotic listening. *Journal of Experimental Psychology*, *85*, 225–228.
- Mackay, D. G. (1973). Aspects of the theory of comprehension, memory and attention. *Quarterly Journal of Experimental Psychology*, *25*, 22–40.
- Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *275*, (942), 483–519.
- Martínez-Trujillo, J., and Treue, S. (2002). Attentional modulation strength in cortical area MT depends on stimulus contrast. *Neuron*, *35*, 365–370. [https://doi.org/10.1016/S0896-6273\(02\)00778-X](https://doi.org/10.1016/S0896-6273(02)00778-X)
- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*(4), 287–330.
- McLeod, P. Driver, J., & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature*, *332*, 154–155.
- Michael, C. R. (1978). Color vision mechanisms in monkey striate cortex: dual-opponent cells with concentric receptive fields. *Journal of Neurophysiology*, *41*(3), 572–588.
- Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, *41*(5), 329.
- Mollon, J. (1974). After-effects and the brain. *New Scientist*, *61*(886), 479–482.
- Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782–784.
- Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56–60.
- Mowrer, O. H. (1938). Preparatory set (expectancy)—a determinant in motivation and learning. *Psychological Review*, *45*(1), 62–91.
- Mowrer, O. H., Rayman, N. N., & Bliss, E. L. (1940). Preparatory set (expectancy)—an experimental demonstration of its' central locus. *Journal of Experimental Psychology*, *26*(4), 357–372.
- Nakayama, K. & Martini, P. (2011). Situating visual search. *Vision Research*, *51*(13), 1526–1537.
- Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, *320*, 264–265.
- Neisser, U. (1963). Decision-time without reaction-time: Experiments in visual scanning. *The American Journal of Psychology*, *76*(3), 376–385.
- Neisser, U. (1964). Visual search. *Scientific American*, *210*(6), 94–103.
- Neisser, U., Novick, R., & Lazar, R. (1963). Searching for ten targets simultaneously. *Perceptual and Motor Skills*, *17*(3), 955–961.
- Nickerson, R. S. (1967). Categorization time with categories defined by disjunctions and conjunctions of stimulus attributes. *Journal of Experimental psychology*, *73*, 211–219.
- Noudoost, B., Chang, M. H., Steinmetz, N. A., and Moore, T. (2010). Top-down control of visual attention. *Current Opinion in Neurobiology*, *20*, 183–190.
- Pack, C. C., Born, R. T., and Livingstone, M. S. (2003). Two-dimensional substructure of stereo and motion interactions in macaque visual cortex. *Neuron*, *37*, 525–535. [https://doi.org/10.1016/S0896-6273\(02\)01187-X](https://doi.org/10.1016/S0896-6273(02)01187-X)
- Palmer, J. (1995). Attention in visual search: Distinguishing four causes of a set-size effect. *Current Directions in Psychological Science*, *4*(4), 118–123.
- Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(1), 108–130.
- Palmer, J., Verghese, P. & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, *40*(10-12), 1227–1268.
- Pomerantz, J. R., & Garner, W. R. (1973). Stimulus configuration in selective attention tasks. *Perception & Psychophysics*, *14*(3), 565–569.
- Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. In H.I Pick, Jr. & E. Saltzman (eds.) *Modes of perceiving and processing information* (pp. 137–158). Hillsdale, NJ: Erlbaum.
- Prinzmetal, W., & Millis-Wright, M. (1984). Cognitive and linguistic factors affect visual feature integration. *Cognitive Psychology*, *16*(3), 305–340.
- Quinlan, P. T. (2003). Visual feature integration theory: past, present, and future. *Psychological Bulletin*, *129*(5), 643–673.
- Rensink, R. A. (2001). Change blindness: Implications for the nature of visual attention. In *Vision and attention* (pp. 169–188). Springer, New York, NY.
- Rentzeperis, I., Nikolaev, A. R., Kiper, D. C., & van Leeuwen, C. (2014). Distributed processing of color and form in the visual cortex. *Frontiers in Psychology*, *5*, 932.
- Reynolds, J. H., Pasternak, T., and Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron*, *26*, 703–714. [https://doi.org/10.1016/S0896-6273\(00\)81206-4](https://doi.org/10.1016/S0896-6273(00)81206-4)
- Rosenholtz, R. (2017). What modern vision science reveals about the awareness puzzle: Summary-statistic encoding plus decision limits underlie the richness of visual perception and its quirky failures. arXiv preprint arXiv:1706.02764.
- Rumelhart, D. E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, *7*(2), 191–218.
- Saraga, E., & Shallice, T. (1973). Parallel processing of the attributes of single stimuli. *Perception & Psychophysics*, *13*(2), 261–270.
- Sekuler, R., & Ball, K. (1977). Mental set alters visibility of moving targets. *Science*, *198*(4312), 60–62.
- Selfridge, O.G. (1959) Pandemonium: a paradigm for learning. In D.V. Blake & A.M Uttley (eds.), *Proceedings of the symposium on mechanisation of thought processes* (pp. 511–529). London: Her Majesty's Stationary Office.
- Selfridge, O.G. and Neisser, U. (1960) Pattern recognition by machine. *Scientific American*, *20*, 60–68
- Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*(1), 54–87.
- Shiffrin, R. M., & Gardner, G. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology*, *93*, 72–82.
- Shiffrin, R. M., McKay, D. P., & Shaffer, W. O. (1976). Attending to forty-nine spatial positions at once. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(1), 14–22.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological monographs: General and Applied*, *74*(11), 1–29.
- Sternberg, S. (1966). High speed scanning in human memory, *Science*, *153*, 652–654.
- Sternberg, S. (1967). Two operations in character recognition: Some evidence from reaction time measurements. *Perception & Psychophysics*, *2*, 45–53.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276–315.

- Taylor, D. A. (1976). Stage analysis of reaction time. *Psychological Bulletin*, 83(2), 161–191.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. Oxford, England: Wiley.
- Townsend, J. T. (1971). A note on the identifiability of parallel and serial processes. *Perception & Psychophysics*, 10(3), 161–163.
- Townsend, J. T. (1972). Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology*, 25(2), 168–199.
- Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science*, 1(1), 46–54.
- Treisman, A. (1970). Perception and recall of simultaneous speech stimuli. In A. F. Sanders (Ed.), *Attention and Performance III*. *Acta Psychologica*, 33, 132–148.
- Treisman, A. (1977). Focused attention in the perception and retrieval of multidimensional stimuli. *Perception & Psychophysics*, 22(1), 1–11.
- Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology*, 40(2), 201–237.
- Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, 6(2), 171–178.
- Treisman, A. (2006). How the deployment of attention determines what we see. *Visual cognition*, 14(4–8), 411–443.
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95(1), 15–48.
- Treisman, A., Kahneman, D., & Burkell, J. (1983). Perceptual objects and the cost of filtering. *Perception & Psychophysics*, 33(6), 527–532.
- Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, 16(3), 459–478.
- Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, 114(3), 285–310.
- Treisman, A., Squire, R., & Green, J. (1974). Semantic processing in dichotic listening? A replication. *Memory & Cognition*, 2(4), 641–646.
- Treisman, A., Sykes, M., & Gelade, G. (1977). Selective attention and stimulus integration. In S. Dornic (Ed.), *Attention and performance VI* (pp. 333–361). Hillsdale, NJ: Lawrence Erlbaum.
- Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 12(4), 242–248.
- Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, 76(3), 282.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136.
- Van der Heijden, A. H. C., & Eerland, E. (1973). The effect of cueing in a visual signal detection task. *Quarterly Journal of Experimental Psychology*, 25(4), 496–503.
- Virzi, R. A., & Egeth, H. E. (1984). Is meaning implicated in illusory conjunctions? *Journal of Experimental Psychology: Human Perception and Performance*, 10(4), 573–580.
- von der Malsburg, C. (1981). *The Correlation Theory of Brain Function*, (Internal Report, No. 81-2). Max-Planck-Institute for Biophysical Chemistry, Department of Neurobiology, Göttingen.
- Webster, J. C., & Thompson, P. O. (1953). Some audio considerations in air control towers. *Journal of the Audio Engineering Society*, 1(2), 171–175.
- Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, 1(2), 202–238.
- Wolfe, J. M. (2016). Visual search revived: the slopes are not that slippery: A reply to Kristjansson (2015). *i-Perception*, 7(3), 2041669516643244.
- Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, 1(3), 0058, 1–8.
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human perception and performance*, 15(3), 419–433.
- Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, 396(6706), 72–75.
- Zeki, S. M. (1973). Colour coding in rhesus monkey prestriate cortex. *Brain Research*, 53(2), 422–427.
- Zelinsky, G. J., Sheinberg, D. L., & Bühlhoff, H. H. (1993, May). Eye movements during visual search. In *Annual Meeting of the Association for Research in Vision and Ophthalmology (ARVO 1993)*.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.