Research News

What boundaries tell us about binding

Michael Kubovy and Dale J. Cohen

Early in the process of perception the visual system decomposes the visual scene into features that are processed in parallel by identifiable subsystems. At some point in the process of scene recognition, these features must be recombined to allow the perception of objects. This recombination is called 'feature integration'. Over the past twenty years, cognitive scientists have devoted much energy to understanding the mechanisms underlying feature integration and, more generally, how perceptual systems recognize that disparate elements in the visual input should be considered parts of the same object. In some contexts the latter operation is called the 'binding problem', in other contexts it is called 'grouping'.

In a recent paper, Phillips and Craven[1] addressed the feature-integration problem with an experimental technique that is similar to ours[2]. However, they analyzed their data using quite different tools, and reached conclusions that were at variance with ours. Unfortunately they did not consider our work in their article. In this comment we outline the feature-integration problem, describe our methodology and results, and discuss why we do not believe that Phillips and Craven's work challenges our conclusions.

The feature-integration problem

The debate regarding feature integration has focussed on two issues: (1) When does feature integration occur?; (2) Is feature integration additive or interactive?

When does feature integration occur?

Does feature integration occur in early vision, which runs effortlessly and is preattentive, or does it occur late, and require the intervention of attention? A common method used to assess whether feature integration occurs in early vision[3] is the 'visual search' technique, which requires subjects to search a computer display for a single target hidden among a variable number of distractors. You can distinguish the target from a distractor only if you have integrated its features. If the reaction time (RT) it took you to find the target increased with the number of distractors, it is generally inferred that you examined the items one-by-one (serially), and concluded that you did not detect the target preattentively. If your RT did not change despite variation in the number of distractors, many researchers would conclude that you processed all the elements in the display at the same time (in parallel), which means that you detected the target preattentively. This is taken to mean that it occured in early vision.

A less common method of assessing whether feature integration occurs preattentively is the 'Gestalt detection technique'[2], which curtails the time available for feature integration. This is achieved by (1) presenting the stimulus briefly, and (2) minimizing post-stimulus processing by following the stimulus with a mask (a carefully designed jumbled array). To complete the method, a means is developed of assessing how well the subject has integrated the features of the stimulus, along with a criterion for the duration of preattentive display (around 200–300 ms).

Is feature integration additive or interactive?

Suppose that the task involves the processing of two features. We need to vary the duration of the stimuli below and above the minimum duration required for the activation of attention. The subsystems that process the features either interact during processing of the two features or they act independently. If we find that the subsystems that process the features are independent for preattentive durations, but feature integration occurs for longer durations, then our theory of feature integration must involve attention. If we find that the subsystems that process the features interact for preattentive durations, then there may be no need to invoke attention in a theory of feature integration.

The Gestalt detection paradigm

We now turn to our Gestalt detection technique, which allows us to assess the degree of preattentive dependence between any two subsystems. We tested whether the subsystems that process form and color function independently in processing brief (and therefore preattentively processed) stimuli. Each of our stimuli consisted of a sequence of frames. In some trials, one of the frames was a target and all the other frames were distractors. Each frame contained a 25-element square lattice of colored forms (see the upper panel of Fig. 1). Each element could have one of two colors and one of two forms. A frame was a 'color target' if its rows (or columns) were partitioned into two compact regions defined by a color change (a color boundary); it was a 'form target' if its rows (or columns) were partitioned into two compact regions defined by a form change (a form boundary). It was a 'color-and-form target' if it contained both a color and a form boundary. The relationship between the color and form boundaries could be coincident, parallel or orthogonal. On each trial we showed between 10 and 16 frames, with the onsets of successive frames separated by 50–250 ms (the stimulus-onset asynchrony).

We used two tasks. In our boundary-detection task, a target was present on only half of the trials, and observers were asked to detect the presence of a target. In the other experiments we used a boundary-localization task, in which a target was present on every trial and observers were asked for the location of *one* boundary, even if two were present.

Data analysis

We obtained a measure of accuracy of detection (A_z) for each observation in the two types of tasks (duration and boundary type). We used the color and the form trials to compute predicted A_z of 'color-and-form' targets based on the assumption that independence between the feature subsystems held, and compared this prediction to the data. The results showed that the detectability for

 1364-6613/01/$ – see front matter PII: S1364-6613(00)01604-1

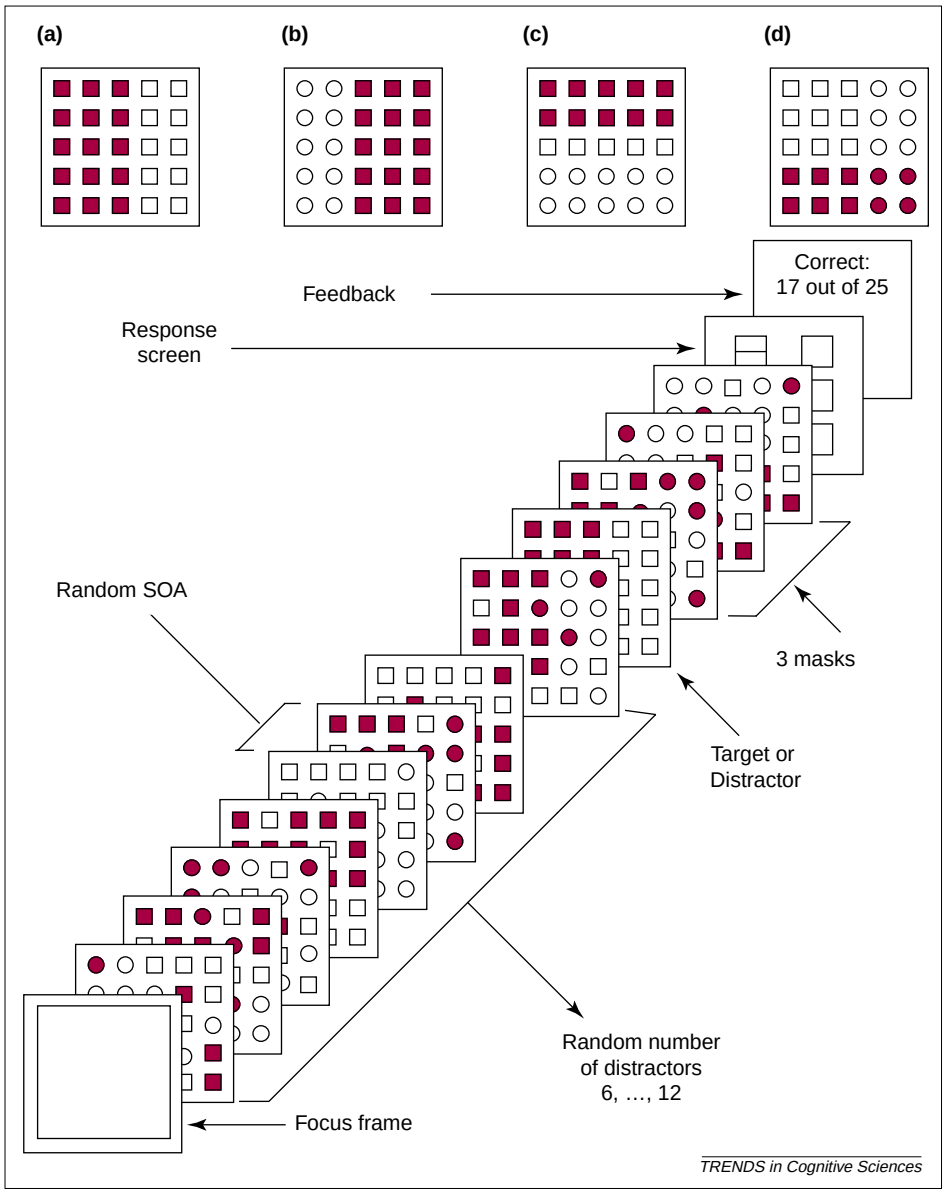

Fig. 1. Experimental design of Kubovy *et al.*[2] Upper panel: four targets (a) a color target with vertical boundary; (b) a 'color-and-form' target with vertical coincident color and form boundary; (c) a color-and-form target with parallel horizontal boundaries; (d) a color-and-form target with orthogonal boundaries. Lower panel: one trial of the boundary-detection procedure in which target (a) appeared.

coincident boundaries (Fig. 2, top) fell in the region of module synergy – this means that when the modules receive consistent information, they operate in synergy. The detectability for parallel and orthogonal boundaries fell below the region of module synergy – this means that when the modules receive inconsistent information, they operate antagonistically. The findings from boundary-localization experiments confirm this pattern of interaction. Thus, our data show that the form and color modules interact.

Phillips and Craven (2000)

The experiments of Phillips and Craven were motivated by two pieces of theoretical research. The first is the work of Phillips *et al.* on neural networks in which contextual information modulates the processing of a target[4]. The second is the work of Smyth *et al.* who showed how information-theoretic measures could be used to analyze and interpret the results of such psychophysical experiments[5].

Phillips and Craven conducted several experiments: they presented observers with a 12×12 matrix of elements, in which they defined boundaries by varying two different features: line length and line orientation. Each of their stimuli contained either one boundary (horizontal or vertical) or two, which could be coincident or orthogonal. They told observers which feature to attend to, thus specifying which feature was the target and which the contextual information. For example, observers could be asked to identify the orientation of length-defined boundaries, in the presence of (1) a coincident orientation-defined boundary, (2) an orthogonal orientation-defined boundary, or (3) no orientation-defined boundary. Each stimulus was visible until the observer responded (≤ 1000 ms). Using the information-theoretic tools developed by Smyth *et al.*, Phillips and Craven reached the conclusion that these feature subsystems do not interact.

Despite these results, we do not believe that they have shown that, in general, feature-processing modules do not modulate each other's output. First, it is not clear whether Phillips and Craven's results were due to preattentive processing or the intervention of attention. We have two main reasons to have doubts on this matter. (1) Phillips and Craven did not limit the amount of time available to their observers to process the stimulus. They presented each stimulus until the observer responded (≤1000 ms), and did not mask the stimulus after offset. Therefore, not only was the presentation time too long to assess whether the processing was preattentive, but processing could have continued even after the stimulus was removed. Indeed, the reported mean reaction times were all greater than 1 s, indicating that observers responded after the stimulus was removed. (2) In some of their experiments[4,5] they found evidence that their results were influenced by late cognitive processes. Because the procedure in these experiments was nearly identical to the procedure they used in the first three experiments, it is possible that all their data were affected by late cognitive processes.

Second, it is not clear how information-theoretic approaches relate to the more standard signal-detection analyses of module interaction. Whereas it is clear that the information-theoretic definition of modulatory interaction involves a non-independence between feature subsystems, it is unclear how modulatory interaction relates to an observer's ability to detect a stimulus, and therefore to independence as conventionally defined in this field.

Finally, Phillips and Craven assessed the perception of the parameters line length and orientation. Even if they

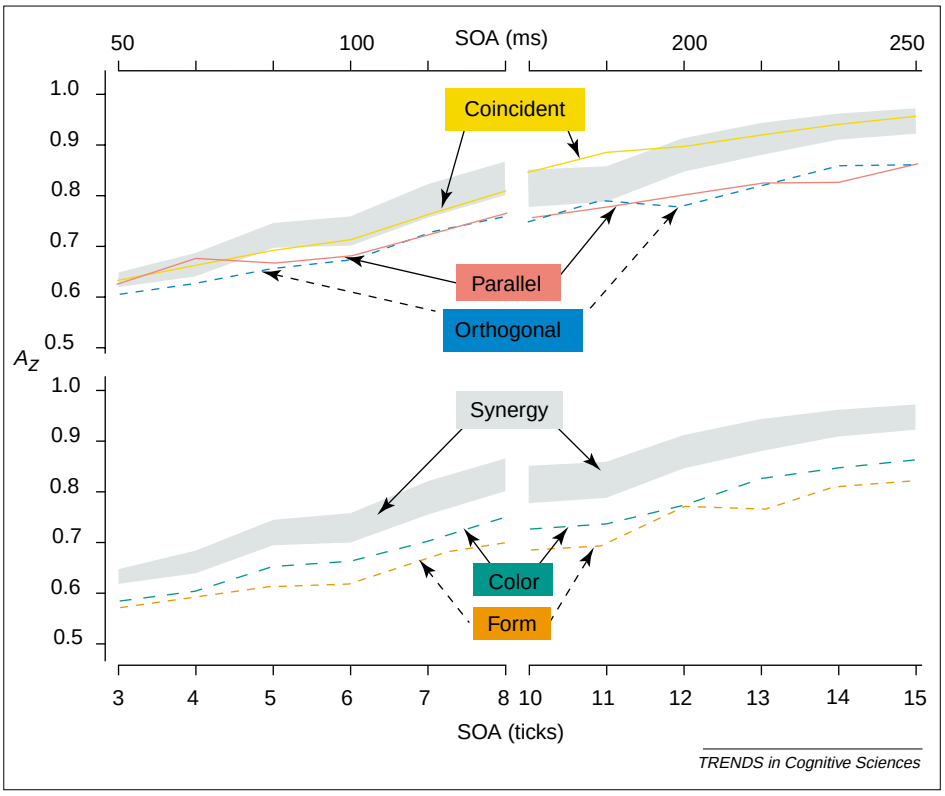

Fig. 2. Detectability (A_z) of targets as a function of target stimulus-onset asynchrony (SOA) in 'ticks' (each tick = 60 s^{-1}). The value in ms is given at the top of the graph. Left: target durations 50–133 ms. Right: target durations 167-250 ms. Bottom: performance for single-boundary targets, colour (green dashed line) and form (orange dashed line). The region of module synergy is the shaded area above these curves (duplicated in the upper panel). Top: performance for dual-boundary targets (orthogonal, blue line; parallel, pink line; coincident, yellow line). It can be seen that the detectability for coincident boundaries falls in the region of module synergy.

followed our recommended procedure, it is possible that they happened to choose two modules that do not interact, and that we happened to chose two modules that do. In such a case they would obtain results different from ours.

Conclusion

Our qualms suggest that three steps need to be taken in order to resolve the inconsistency between our work and that of Phillips and Craven: (1) there needs to be a comparison of their modeling tools with ours, to determine whether conditions exist under which our model would show independence and their model would show interactivity; (2) we should use our experimental technique to see whether we obtain interactivity of line-length and line-orientation modules; and (3) we should explore whether experimental techniques that insure preattentive processing can be modified to allow the application of the Smyth *et al.* analytic tools.

References

1 Phillips, W.A. and Craven, B.J. (2000) Interactions between coincident and orthogonal cues to texture boundaries. *Percept. Psychophys.* 62, 1019–1038
2 Kubovy, M. *et al.* (1999) Feature integration that routinely occurs without focal attention. *Psychonomic Bull. Rev.* 6, 183–203
3 Treisman, A. and Gelade, G. (1980) A feature-integration theory of attention. *Cognit. Psychol.* 12, 97–136
4 Phillips, W.A. *et al.* (1995) The discovery of structure by multi-stream networks of local processors with contextual guidance. *Network: Compilation Neural Syst.* 6, 225–246
5 Smyth, D. *et al.* (1996) Measures for investigating the contextual modulation of information transmission. *Network Compilation Neural Syst.* 7, 307–316

Michael Kubovy*
Psychology Dept, University of Virginia, Charlottesville, VA 22904-4400, USA.
*e-mail: kubovy@virginia.edu

Dale J. Cohen
Psychology Dept, University of North Carolina at Wilmington, NC, USA.

Contextual modulation and dynamic grouping in perception

Response from William A. Phillips

Evidence for specialization of function within and between modules dominates research in cognition and neuroscience. Interactions that coordinate those specialized activities are also necessary, however, and, although often taken for granted, they are much less studied and much less well understood. Two major forms of coordination can be distinguished: dynamic grouping and contextual modulation[1]. I use the term 'dynamic grouping' to be more precise about issues that are commonly discussed under the heading of 'binding'. Grouping can be divided into two fundamentally distinct classes: pre-specified grouping and dynamic grouping. Pre-specified grouping is due to the convergence of signals in a hierachy of feature detectors. It is ubiquitous in neural computation, and is the primary mechanism by which feature or object detectors compute whatever it is that they detect. It is pre-specified in the sense that although it adapts gradually to the statistical structure of input, it is specified as a possible grouping prior to the occurrence of the particular inputs processed at each moment. Dynamic grouping, by contrast, forms groupings that cannot be specified prior to the particular input being processed, and are computed by processes that configure themselves to that input. Dynamic grouping was emphasized by the Gestalt psychologists, and can occur pre-attentively[2]. Processes that group features dynamically can therefore be clearly

 1364-6613/01/$ – see front matter PII: S1364-6613(00)01617-X