



The structure of illusory conjunctions reveals hierarchical binding of multipart objects

Edward Vul¹ · Cory A. Rieth² · Timothy F. Lew³ · Anina N. Rich⁴

Published online: 24 October 2019
© The Psychonomic Society, Inc. 2019

Abstract

The world around us is filled with complex objects, full of color, motion, shape, and texture, and these features seem to be represented separately in the early visual system. Anne Treisman pointed out that *binding* these separate features together into coherent conscious percepts is a serious challenge, and she argued that selective attention plays a critical role in this process. Treisman also showed that, consistent with this view, outside the focus of attention we suffer from illusory conjunctions: misperceived pairings of features into objects. Here we used Treisman's logic to study the structure of pre-attentive representations of multipart, multicolor objects, by exploring the patterns of illusory conjunctions that arise outside the focus of attention. We found consistent evidence of some pre-attentive binding of colors to their parts, and weaker evidence of binding multiple colors of the same object. The extent to which such hierarchical binding occurs seems to depend on the geometric structure of multipart objects: Objects whose parts are easier to separate seem to exhibit greater pre-attentive binding. Together, these results suggest that representations outside the focus of attention are not entirely a “shapeless bundles of features,” but preserve some meaningful object structure.

Keywords Attention · Binding · Feature Integration Theory

Anne Treisman demonstrated that “binding” features that are represented in anatomically segregated areas, such as color and shape (Livingstone & Hubel, 1988), into consciously perceived objects is a challenge for the visual system, and that focused selective attention is crucial for solving it (Treisman, 1988; Treisman & Gelade, 1980; Treisman & Gormican, 1988; Wolfe & Cave, 1999; Treisman, 1998). When selective attention cannot isolate objects in order to bind their constituent features together, people tend to misperceive which features go together, yielding “illusory conjunctions” of features from different objects (e.g., Treisman & Schmidt, 1982). For example, a brief presentation of a red X and a yellow Y might

be misreported as a yellow X and a red Y. Such illusory conjunctions arise when attentional capacity is strained due to competing task demands or rapid presentation (Kanwisher, 1991; Treisman & Schmidt, 1982), damage to the parietal cortex (Cohen & Rafal, 1991; Friedman-Hill, Robertson, & Treisman, 1995), or neural stimulation disrupting parietal function (Ashbridge, Walsh, & Cowey, 1997). Experiments generating illusory conjunctions have provided evidence that attention is important for solving the “binding problem.”

In her feature integration theory, Treisman (Treisman, 1988; Treisman & Gelade, 1980) proposed that the means by which attention binds multiple features together is by forming a master location map that indexes the independent feature maps (Treisman, 2006). Thus, correct conjunctions are perceived only when attention selects a master map precise enough to correctly isolate the features of one object and to exclude the features from other locations. Although various modifications of feature integration theory have been proposed (Huang & Pashler, 2007; Huang, Treisman, & Pashler, 2007), they retain the core component: that attention defines a location map used to identify features. When attention does not form a map precise enough to isolate a single object, it seems that feature values are randomly sampled from the set selected by the map (Ashby, Prinzmetal, Ivry, &

✉ Edward Vul
evul@ucsd.edu

¹ Department of Psychology, University of California, San Diego, CA, USA

² Pacific Science and Engineering, San Diego, CA, USA

³ Quora, Mountain View, Mountain View, CA, USA

⁴ Perception in Action Research Centre (PARC), Department of Cognitive Science, Faculty of Human Sciences, Macquarie University, Sydney, New South Wales, Australia

Maddox, 1996; Vul, Hanus, & Kanwisher, 2009). Consequently, when attention is overloaded and imprecise, illusory conjunctions are likely to arise from nearby items (Cohen & Ivry, 1989; Emrich & Ferber, 2012). Such random feature sampling yields correct conjunctions only in the limited case in which attention selects a location precise enough to isolate one object (Vul & Rich, 2010). Without such precision, the link between the different feature maps and the master location map can be compromised.

There is conflicting evidence, however, about how multiple features of objects are represented outside the focus of attention. Some evidence suggests that outside the focus of attention, multiple spatial features, such as oriented lines, may be joined into emergent shapes that may then be misbound together (Treisman & Paterson, 1984). Moreover, the set of basic features in an object appears to be somehow bundled together pre-attentively (Wolfe & Bennett, 1997). Orientation and color seem to be conjointly represented outside the focus of attention or awareness, since orientation-contingent color aftereffects may be induced even when the inducing stimuli are imperceptible (Humphrey & Goodale, 1998; Vul & MacLeod, 2006). Furthermore, multiple features in the same domain appear to be encoded jointly, yielding accurate perception of the average feature of the “ensemble” (Chong & Treisman, 2003, 2005). Such feature ensembles can be formed outside the focus of attention (Alvarez & Oliva, 2008) and may bias perception of the features of individual objects (Brady & Alvarez, 2011). Although search for targets defined by two colors among distractors that share these colors is usually very inefficient (Wolfe et al., 1990), this is not the case if there is a hierarchical relationship between the two colors, such that one colored section is perceived as “part” of the other colored “whole” (Wolfe, Friedman-Hill, & Bilsky, 1994). This special case seems to hold for color and size (Wolfe et al., 1994), but not for orientation and form (Bilsky & Wolfe, 1995). Furthermore, although some visual search results are consistent with pre-attentive binding of conjunctions (Found, 1998), these results are also consistent with certain unbound feature selection strategies (Nordfang & Wolfe, 2014).

Together, this work suggests that even when spatial attention is not precise enough to select a specific feature conjunction, features do not seem to be represented as an undifferentiated feature soup. Instead, evidence of co-dependent, hierarchical coding of features into clusters and ensembles suggests that outside of very sparse, unstructured displays, the perception of conjunctions might not be a matter of merely sampling independent features.

Here we explored the statistical structure of illusory conjunctions to test for a pre-attentive hierarchical organization of features into multipart, multifeature assemblies. We showed subjects multipart objects such as a bulls-eye or a cross (Fig. 1). Each part had a distinct color, and we asked subjects to identify the colors of both parts of the cued target object. The

colors of the object parts near the target were unique, so that each incorrectly reported color could be identified as an intrusion from a particular part of a particular object.

The structure of the displays allowed us to identify the source of the reported color. If, for example, the subject reports two incorrect colors (e.g., a pink horizontal and brown vertical for the display shown in Fig. 1a), we would know in which of the surrounding objects and parts the reported colors were presented (e.g., from the same object clockwise from the target, but swapped across parts). The reported colors of the target object could therefore be categorized as one of 100 possible color–color conjunctions, defined by which part of which object the colors appeared on (Fig. 2). In this task, different pre-attentive representations of the multipart objects would yield different diagnostic patterns of illusory conjunctions. Consequently, on the basis of the distribution of reported conjunctions, we could identify whether, outside of attention, multipart objects are represented as independent features, as ensembles of features from the same object, as bound parts, or as completely bound whole objects.

Independent features

The default account of feature maps and illusory conjunctions predicts that outside of attention, features are represented independently. Consequently, when attention is too imprecise to select one feature value, all the features within the selected spatial region are mixed into a feature “soup” from which feature values are independently sampled, with no regard for their part or object origin. Thus, if attention is not focused sufficiently to isolate specific conjunctions, the features are effectively unbound, with no defined relationship to each other, to particular object parts, or to objects. According to this account, if attention is not sufficiently precise to isolate a particular conjunction, then color intrusions should be independent: There should be no systematic relationship to which part or which object (within the window of attentional selection) the two colors would come from (Fig. 3a). Correctly reported conjunctions would arise only when attention becomes so precise as to isolate a particular display element, yielding a feature soup made of only one ingredient feature.

Color–color object ensembles

A second option is that there is some association between different adjacent instances of the same feature (e.g., multiple colors) outside of attention, but different types of features (e.g., color and shape) are represented independently. This might be the case if multiple adjacent values within a given feature domain form “ensembles” (Chong & Treisman, 2003, 2005). Although most work on ensemble statistics has focused

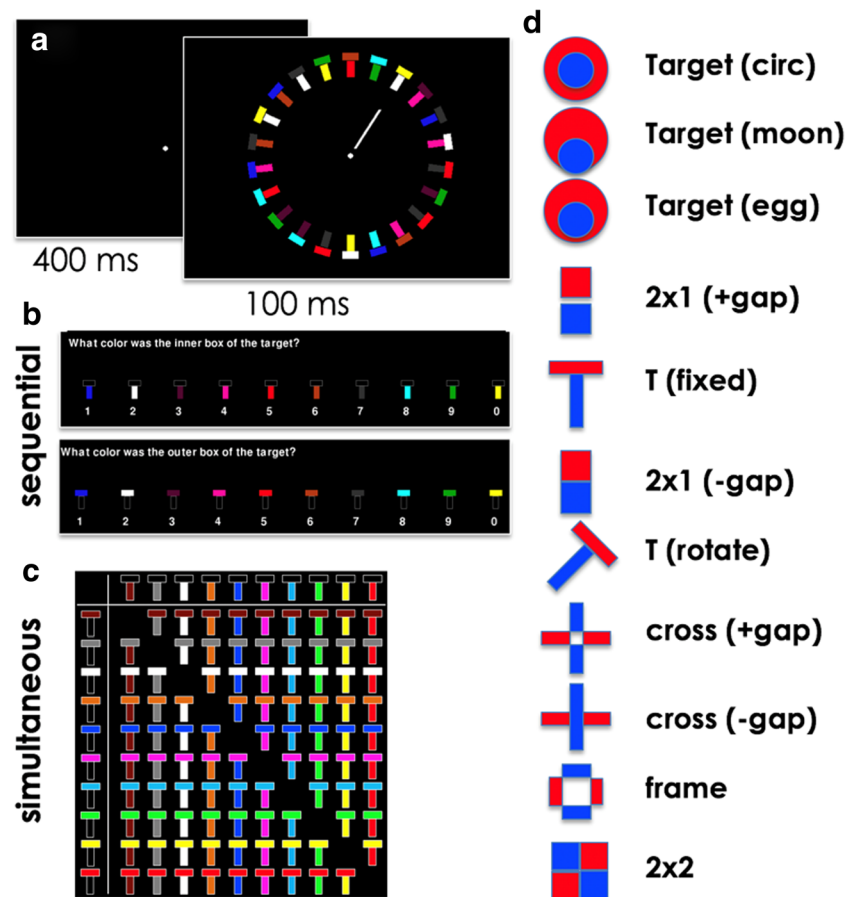


Fig. 1 Experiment procedure. **(a)** Subjects saw a fixation display for 400 ms, followed by a 100-ms display containing 22 two-part/two-color shapes, one of which was cued by a white line (here showing rotated Ts; other object shapes had a stable absolute orientation, rather than a stable orientation with respect to fixation, as is shown here for the rotated Ts). The ten colors of the five items centered around the cue were all unique. **(b)** In the sequential-report experiments, subjects were asked to

report the colors of the two parts of the shape sequentially (in a random order), each time picking from one of ten unique colors. **(c)** In the simultaneous-report experiments, subjects reported the two colors simultaneously, by clicking on one of 90 possible color conjunctions. **(d)** The 11 two-color shapes used here, arranged from highest (top) to lowest (bottom) average accuracy in reporting the correct conjunction (see Table 1).

on whole-scene statistics, there is evidence that we may extract group-specific statistics (Brady & Alvarez, 2011; Lew & Vul, 2015). If people represent object-specific color statistics, then these statistics can be used to reconstruct the two colors of the object. Thus, proximity-based color-color associations can arise from object-specific ensemble representations. Under this hypothesis, instead of sampling independent feature values from the selected region, one would sample these local object ensembles, and thus would tend to report two colors from the same object, even if they are not correctly matched to their shape-defined parts (Fig. 3b). The signature of this process would be that even when colors are swapped across parts, they would tend to come from the same object. In Fig. 2c, this would correspond to higher rates of reporting conjunctions marked with “c” (colors swapped, but both arising from the same object adjacent to the target) than “e” (colors swapped, and the two colors arising from different target-adjacent objects).

Part-color binding

A third possibility is that colors are somehow bound to parts outside the focus of attention, as might arise if some aspect of the shape of a given color region is pre-attentively represented in the color feature map. This third option would effectively suggest multiple levels of feature binding within a single object: Features are associated with their parts in the absence of attention, but the binding of parts into objects requires attention. The signature of such proto-binding would be correct binding of colors to object parts, but illusory conjunctions of bound color-parts across objects (Fig. 3c). In this case, when attention is imprecise, we should see errors at an object level; colors should be correctly bound to parts, but the object-parts could arise as often from different objects as from the same object. The signature of this process would be that even when colors are reported from different objects, they are more likely to arise from the same (rather than from different) parts. In Fig.

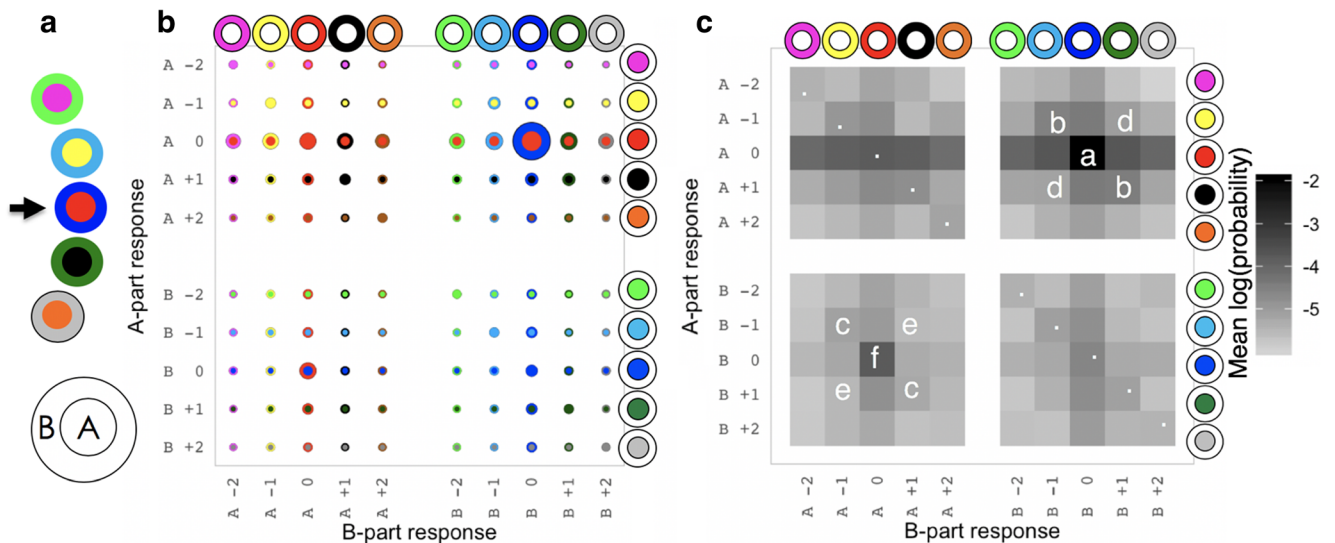


Fig. 2 Joint response distribution for bulls-eye (circular target) stimuli. (a) One example arrangement of ten unique colors onto the five objects surrounding the target object (red inside blue, indicated by the black arrow). Subjects could report any one of 100 possible conjunctions, corresponding to reports of any of the ten colors in this panel for each of the two parts of the target object. (b) This data panel shows the frequency with which each possible conjunction was reported, with the area of each object scaled to reflect the mean probability (across subjects) that this conjunction was reported in the sequential-report experiment for this shape. The *x*-axis represents which color was chosen for part B (here, the outer ring of the bulls-eye), and the *y*-axis represents which color was reported for part A (here, the center of the bulls-eye). Each color is identified as belonging to one part (a or b) of one of the five items around the cued location (−2 to 2). Thus, “A − 1” represents the color in part A of the item one position counterclockwise from the target (here, yellow). Together, the *X*- and *Y*-coordinates indicate the color conjunction reported in terms of the origin of the colors. The upper right quadrant (*X* = B* and *Y* = A*) corresponds to reporting the colors correctly paired to their parts, but not necessarily arising from any one object, whereas the lower left quadrant (*X* = A* and *Y* = B*) would mean that the colors were swapped across parts. For instance, the left diagonal of the upper right quadrant shows reports in which the two colors come from the same object, with either the correct color–part binding (the target object: *X* = B 0, *Y* = A 0) or a whole-object, nontarget intrusion (e.g., *X* = B − 1, *Y* =

A − 1). The left diagonal of the lower left quadrant shows reports in which the two colors come from the same object, but with colors swapped across parts (from the target object: *X* = A 0, *Y* = B 0), or a nontarget, object-ensemble intrusion (e.g., *X* = A − 1, *Y* = B − 1). This type of display illustrates the different potential report combinations and shows that the correct conjunction is reported more often than other conjunctions. (c) The same data presented as heatmaps of mean log probability illustrate more subtle patterns. Points of particular interest are marked with symbols: “a” corresponds to the correct target conjunction report, whereas “f” is the conjunction composed of both target colors, but swapped across parts. The upper right quadrant corresponds to reported conjunctions in which each part is labeled with a color that originated on the same part type (although not necessarily on the target object; responses in this region would be more likely under intrusions of bound color–part conjunctions). Cells marked with “b” correspond to matched part–color reports of one of the two objects adjacent to the target (as would arise from whole-object intrusions), whereas “c” indicates reports of the same colors, but swapped across parts (as would arise from intrusions of bound color–color object ensembles). “d” corresponds to reports of two colors from different objects adjacent to the target reports, but the colors are correctly matched to their parts (as would arise from binding of colors to parts, but not of parts to objects), whereas “e” indicates that the colors were also swapped across parts. The major diagonals marked with “.”s indicate reports of the same color for both parts

2c, this would correspond to higher rates of conjunctions marked with “d” (colors arising from different objects, but the colors are correctly matched to parts) than with “e” (colors arising from different objects and swapped across parts). Wolfe et al. (1994) found evidence that such part–color conjunctions may be more pre-attentively accessible, as it is easier to find a red house with yellow windows than a red-and-yellow house.

Whole-object binding

Finally, implausible though it may be in light of the prior literature, it is possible that outside the focus of attention,

features, parts, and objects are all correctly bound together. If this were the case, then when attention was insufficiently precise to select only the target, we would expect to see whole-object intrusions of complete correctly bound objects, such that any reported conjunctions would correspond to one complete, presented, nontarget object (Fig. 3d). The signature of such a process beyond object-ensemble intrusions and bound part–color intrusions would be a greater same-object advantage if parts were matched than if they were swapped. In other words, in Fig. 2c, this would mean that the rate of “b” (matched part, same object) as compared to “d” (matched part, different object) conjunctions would be greater than the rate of “c” (swapped part, same object) as compared to “e” (swapped part, different object) conjunctions.

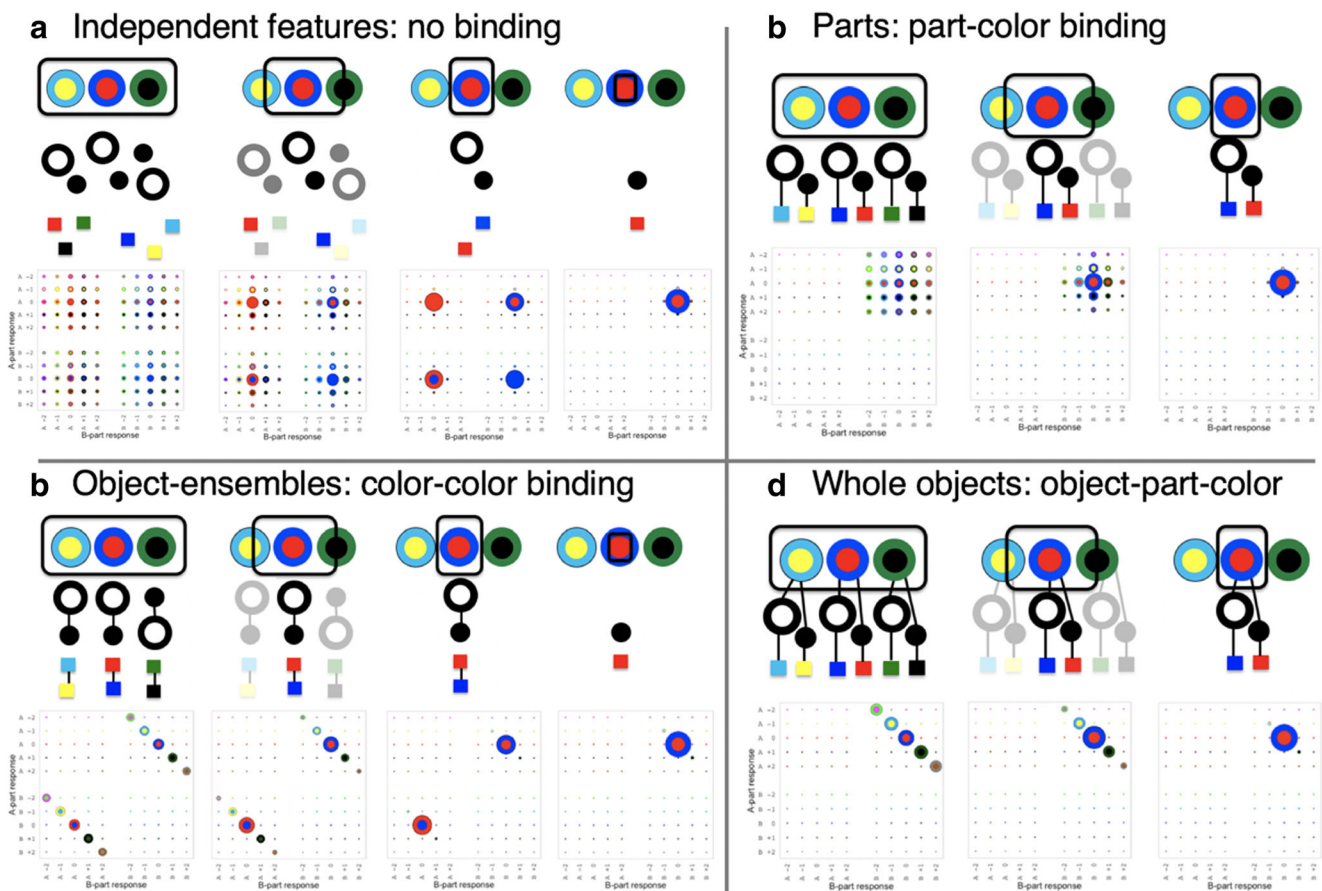


Fig. 3 Predictions of error patterns as attention becomes more precise, based on the different representations of multipart objects outside the focus of attention. The top row of each panel shows the stimuli, with the target in the center. The black rectangles show the precision of attention, from imprecise (capturing three adjacent objects, each with two parts) on the left to precise (capturing only the parts of the target object) on the right of each panel. The second row of each panel shows the features present within the selected region, along with any binding (connecting line). The third row of each panel shows the predicted joint report distributions, in the format described in Fig. 2b. The area of each dot shows the predicted frequency for each possible conjunction. **(a)** The independent-feature account, in which features float freely, unbound to each other, to parts, or to objects, unless attention is focused enough to isolate a particular conjunction. As the window of attention becomes more focused, the pattern of intrusions isolates specific objects, and eventually isolates the target and its specific parts. **(b)** Object ensembles, as would arise if the features of the same type (e.g., colors) are bound to each other in “ensembles,” but features of different types (e.g., color and

shape) are not bound to each other, and no features are bound to an object outside of attention. With imprecise attention, we would expect whole-object intrusions. With chance levels of part binding. As attention becomes more precise, eventually only the correct target is selected, but even then, colors are not correctly matched to parts; only when attention is precise enough to isolate a single part would the correct target conjunction be reported. **(c)** Features are bound to parts, but parts are not bound into objects. We would not expect above-chance levels of features arising from the same object, unless attention was precise enough to isolate that object. However, in all cases, colors would be correctly matched to their parts. **(d)** Parts are bound to objects, and features are bound to parts. Color intrusions would be correctly matched to their parts and would arise from the same object (complete-object intrusions). As attention becomes more precise, only the target object would be reported. To report the correct part-color conjunction, representations of independent features and object ensembles would require that the selected region isolate a specific part, whereas part-color and whole-object representations would require only that the target object be selected

Method

Subjects

All subjects were UCSD undergraduates (mean age = 20 years; 64% female) and had normal or corrected-to-normal vision and participated for course credit. This study was approved by the UCSD Institutional Review Board. For each object type and response modality, we aimed to collect at least

16 subjects, but we achieved this by collecting data for one week, yielding slightly different numbers of subjects in each case. Although such convenience-based subject-sampling strategies have fallen out of favor since Anne Treisman’s time, given the exploratory, model-based nature of this investigation, targeting a sample size based on a hypothesized effect size was impractical, and we believe that Treisman’s strategy of internal replications (as we have done with sequential and simultaneous response types for each object type) offers more

useful assurance of the reliability of results than do invariant subject counts. Table 1 shows the numbers of subjects, the range of trials/subject, and the overall accuracy for each stimulus and experiment. Two subjects were excluded because they terminated the experiment before completing 200 trials.

Stimuli

The main display had 22 stimuli arranged in a circle of radius 5 deg from central fixation (a white dot) on a black background (Fig. 1a). Each stimulus was ~1 deg of visual angle in diameter at a viewing distance of ~60 cm. A white line extending ~3 deg from fixation indicated the target on each trial.

There were ten possible colors, which were unique and randomly assigned to the ten parts of the five items around the target. Every target (T), the adjacent items (T + 1, T – 1) and the next ones along (T + 2, T – 2) had unique colors for both of their parts, allowing us to identify the source of any color intrusions.

We tested 11 different two-part/two-color shapes (Fig. 1d). These varied in the degree to which the parts were

overlapping, the complexity of the object, shape topology, and other variables that might identify what makes some shapes easier to bind than others (in the end, none of the specific hypotheses about what determines the extent to which a shape is bindable outside of attention was decisively borne out). There were three bulls-eye-style targets, varying in the offset of the inner circle from the outer circle: centered (target circle), offset but contained (target egg), and offset past the outer circle boundary (target moon). There were four objects composed of two rectangles: in a T configuration with a fixed absolute orientation (T-fixed), in a T configuration with a fixed orientation relative to fixation (T-rotate), and two squares one atop the other, with (2×1 + gap) and without (2×1 – gap) a gap between them. Finally, four object types comprised four rectangles varying in aspect ratio and configuration: a plaid of four squares (2×2); a cross in which the front rectangles occluded the rear (cross – gap); a cross in which the rectangles did not abut each other, leaving a central gap (cross + gap); and the same cross + gap configuration, but inverting the aspect ratio of the rectangles, thus yielding something that looked like a window frame (frame). All the shapes retained a fixed absolute orientation, with the exception of the rotated Ts, which had a fixed orientation relative to the fixation point.

Table 1 Number of subjects in each combination of object type and experiment (response type), as well as the range of completed trials/subject and the average accuracy for reporting the correct conjunction

| Object Type | Response Type | Subjects | Trials/Subject | Accuracy |
|---------------|---------------|----------|----------------|----------|
| Target (circ) | Sequential | 24 | 500–500 | 22% |
| Target (circ) | Simultaneous | 26 | 400–400 | 26% |
| Target (moon) | Sequential | 28 | 300–500 | 22% |
| Target (moon) | Simultaneous | 26 | 400–400 | 26% |
| Target (egg) | Sequential | 27 | 300–500 | 18% |
| Target (egg) | Simultaneous | 26 | 400–400 | 27% |
| 2×1 (+ gap) | Sequential | 25 | 300–400 | 14% |
| 2×1 (+ gap) | Simultaneous | 26 | 345–400 | 16% |
| T (rotated) | Sequential | 30 | 300–400 | 13% |
| T (rotated) | Simultaneous | 18 | 400–400 | 17% |
| 2×1 (– gap) | Sequential | 32 | 300–400 | 15% |
| 2×1 (– gap) | Simultaneous | 17 | 400–400 | 14% |
| T (fixed) | Sequential | 37 | 300–400 | 9% |
| T (fixed) | Simultaneous | 25 | 289–400 | 13% |
| Cross (+ gap) | Sequential | 27 | 450–500 | 13% |
| Cross (+ gap) | Simultaneous | 24 | 400–400 | 7% |
| Cross (– gap) | Sequential | 24 | 300–400 | 8% |
| Cross (– gap) | Simultaneous | 25 | 400–400 | 8% |
| Frame | Sequential | 34 | 300–400 | 7% |
| Frame | Simultaneous | 24 | 400–400 | 8% |
| 2×2 | Sequential | 30 | 300–400 | 4% |
| 2×2 | Simultaneous | 24 | 322–400 | 11% |

Chance performance is 1%. Objects are arranged from highest to lowest overall average accuracy across both response types

Procedure

Each trial started with a fixation dot for 400 ms, followed by the main display for 100 ms (Fig. 1a). Because the cue was presented simultaneously with the stimulus display, this created some uncertainty in the target selection (see Vul & Rich, 2010, for a manipulation of cue–target interval). Subjects then reported the two colors of the cued target item.

In the “sequential” experiments, there were two response screens, presented in random order on each trial. On each response screen, the ten possible colors were shown in one part of the object (Fig. 1b). The subject had to select the target color for that part on each response screen. In the “simultaneous” experiments, we presented all the possible conjunctions simultaneously (Fig. 1c), and the subjects had to select the target stimulus by clicking on it.

Results

Our data yielded a distribution over all the possible color–color conjunctions that could be reported. With all five colors for part A and the five colors for part B on the items around the target as candidate reports for both parts of the target, there were 100 possible conjunctions in the sequential reports, and 90 in the simultaneous reports (because same-color reports were not allowed). Figure 2 shows the frequency distributions for reported conjunctions of bulls-eye-shaped objects under sequential report, and Fig. 4 shows these joint distributions

for all objects and response types. Different types of intrusions yield different patterns in these histograms. In the Model Decomposition of Errors section, we explicitly estimate the preponderance of each type of intrusion. First, however, we conducted focused analyses to assess the extent to which the pattern of intrusions reflects some pre-attentive binding of two colors from the same object into a sort of object ensemble (Fig. 3b), the binding of parts to colors (Fig. 3c), or the binding of complete objects (Fig. 3d). Each of these forms of pre-attentive binding would create diagnostic dependencies in the distribution of illusory conjunctions that would not arise from sampling independent features.

Independent feature coding or color–color object ensembles?

When attention is too imprecise to support accurate report, the pattern of errors can distinguish between independent sampling of features and the sampling of color–color object ensembles. The ensemble account predicts that intruding colors would be more likely to arise from the same object than from different objects, creating a relationship between the origins of the two intrusions. In our previous work (Vul & Rich, 2010), we found that intrusions of color and letter identity from items adjacent to the target were unrelated, reflecting independent sampling of features. However, two instances of the same feature might be encoded into some sort of object ensemble, and thus would not be independent. This would cause some binding of colors within the same object and would yield higher rates of color intrusions arising from the same object. To isolate this effect, we could ask whether two colors that both originated from one of the two objects adjacent to the target would be more likely to come from the same object than

from different objects. To consider only the effect of such color-ensemble intrusions, we made this comparison only in the subset of trials in which both colors were incorrectly paired to parts (such that the color of part A was misreported as the color of part B, and vice versa). Figure 5 shows that intrusions are more likely to arise from the same object than from different objects, even when they are incorrectly bound to parts. This suggests that sometimes a color ensemble intrudes from an adjacent item, with the part–color binding unknown.

Are colors bound to parts outside of attention?

If illusory conjunctions arise because spatial attention is too imprecise to select the features that correspond to the target object, we would expect that color intrusions from adjacent objects would not be correctly bound to their parts. However, if colors are bound to parts outside of attention, we would expect that even when the two intruding colors originated from different objects, they would be more likely to be correctly paired to their parts than to be swapped. Specifically, if spatial attention were so imprecise that you report two colors from the two items adjacent to the target, they could be correctly bound to parts, be swapped across parts, or come from the same part; if there were no local binding of colors to parts, then the rate of swapped-part intrusions would be the same as the rate of matched-part intrusions. Figure 6 shows that, on average across object types, and for many specific object types, the probability that two colors reported from two different objects were correctly matched to their parts (rather than swapped) was greater than the 50% that would be expected under independent-feature intrusions. This means that for those objects, some amount of local part–color binding preceded attentional selection.

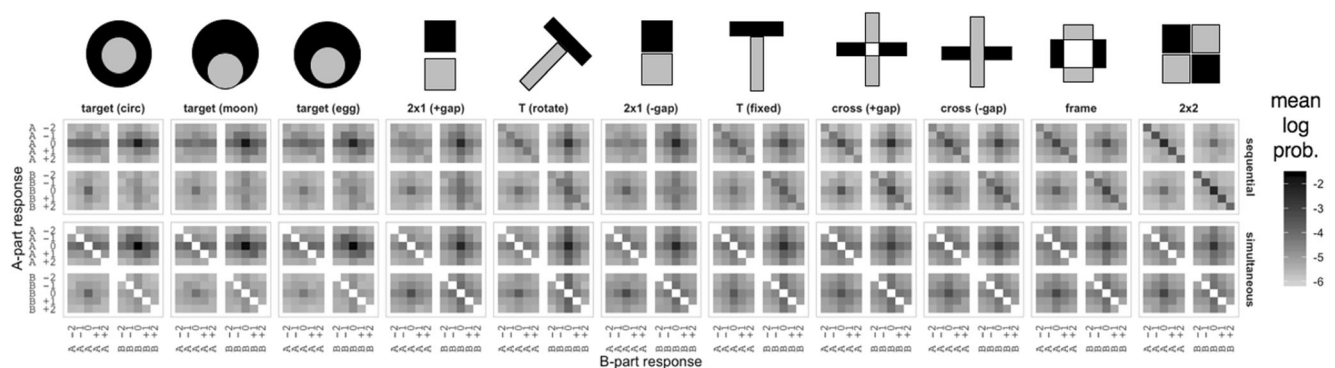


Fig. 4 Results. The joint frequency distributions of which colors were reported in the two experiments (rows: sequential vs. simultaneous), for each of the 11 different shapes (columns). The order of shapes in the figure follows average accuracy (reporting the correct target conjunction), from highest accuracy (easiest shapes) on the left to lowest accuracy (hardest shapes) on the right. Several coarse trends are evident here. First, harder shapes involve more repetition of the same color when that is an option in the sequential experiment (the main diagonal has a greater probability mass for shapes on the right than on

the left). Second, for the easier shapes, more colors are reported as correctly bound to their parts than as swapped, even when they do not originate from the target object (the upper right quadrants are darker than the lower left quadrants). Third, for easier shapes, intrusions of the two features tend to come from the same object, whether or not the colors are correctly matched to parts (the diagonals corresponding to the same origin object—those sloping down—are darker than the orthogonal diagonals in both the upper right and lower left quadrants). Each of these patterns is demonstrated with focal analyses below

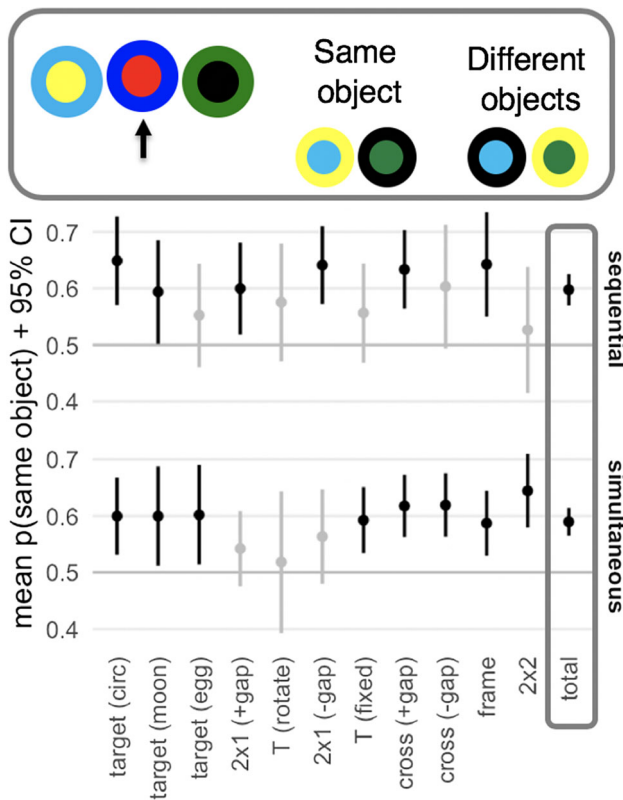


Fig. 5 Testing for the signature of object-ensemble intrusions. If two colors from one object are somehow associated outside of attention, then we would expect two intruding colors to arise from the same object more often than chance. We compared the rate at which two colors arose from the same object when both colors were incorrectly paired with parts and both originated from one of the two items adjacent to the target (this amounts to comparing the probability of points “c” in Fig. 2b to the combined probability of points “c” and “e”). The y-axis shows the mean (across-subjects) probabilities and 95% confidence intervals on the probability that such intrusions arose from the same object for each object type and each response type. Points significantly different from chance (50%) are black (rather than gray). The rate of such color–color intrusions is higher than chance, on average across object types, as well as for most individual object types, in both the sequential and simultaneous report conditions

Are colors and parts bound together into whole objects outside of attention?

Whole-object intrusions differ from object-ensemble intrusions only in that they predict intrusions from the same object would be correctly bound to their parts. Consequently, the signature of whole-object intrusions would be a greater rate of two colors arising from the same object when they were correctly matched to their parts than when they were incorrectly matched to their parts. As in our analysis for the signature of object ensembles, we considered only conjunctions in which both colors arose from one of the objects adjacent to the target and in which the colors were either correctly matched to parts or swapped. Here we compared the probability that the two colors came from the same object when they were matched to

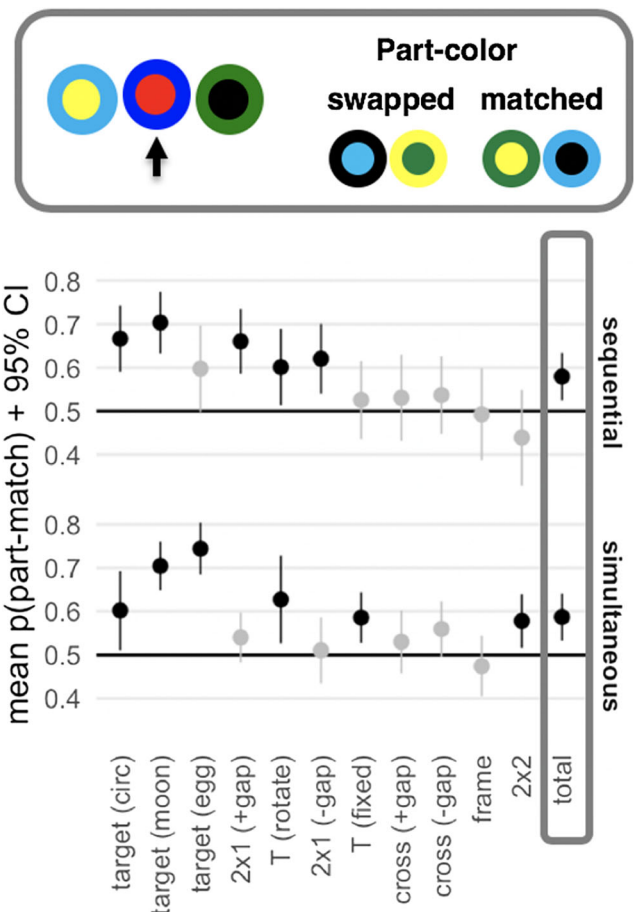


Fig. 6 Two colors arising from different parts of different objects adjacent to the target are more likely than chance to be correctly bound to parts. The y-axis shows the mean (across subjects + 95% confidence interval) probabilities that colors are correctly matched to their parts (rather than swapped), when both colors came from different parts of different objects adjacent to the target (this amounts to comparing the probability of points “d” in Fig. 2b to the combined probability of points “d” and “e”). Estimates significantly different from chance (50%) are in black. Across both the sequential and simultaneous experiments, many objects meet this criterion for part–color binding outside the focus of attention

their parts to the probability that the two colors came from the same object when they were swapped across parts. Whole-object intrusions would predict a greater same-object advantage for matched-part intrusions than for swapped-part intrusions. Figure 7 shows that such a whole-object effect is small, unreliable, and largely limited to the easiest object geometries (bulls-eye-shaped targets). This suggests that whole-object intrusions are very rare, if they happen at all.

Discussion

The targeted analyses in this section isolated specific, diagnostic comparisons from the full distribution of all conjunctions to detect key signatures of specific types of pre-attentive representations. We found that, as predicted from pre-attentive formation of object ensembles, two intruding colors swapped

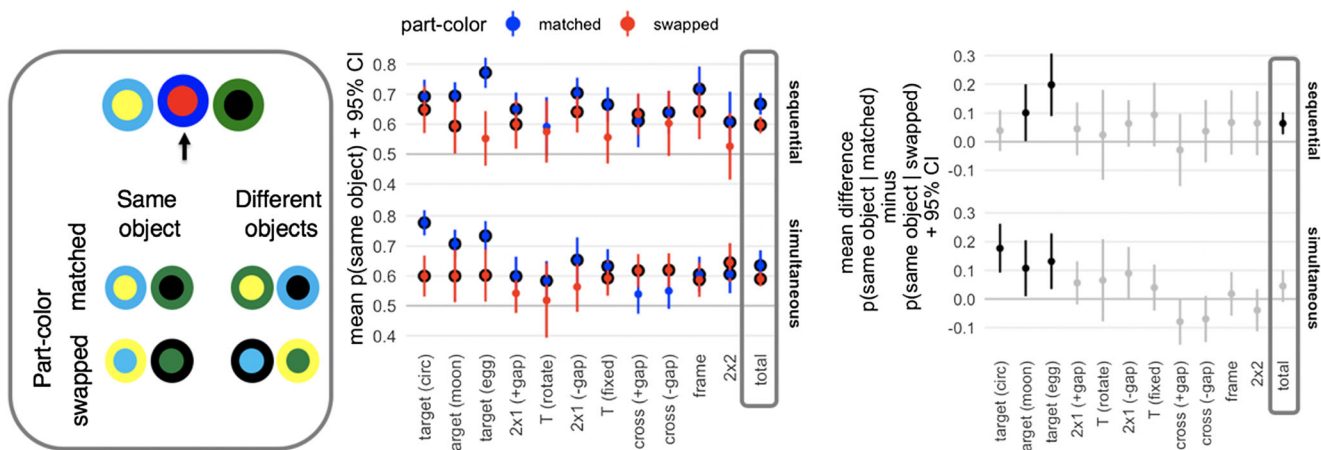


Fig. 7 Are there whole-object intrusions? If so, the rate at which two colors correctly matched to their parts would arise from the same object should be higher than the rate at which two colors swapped across parts would arise from the same object. When both colors reported for the target conjunction are intrusions from one of the two target-adjacent objects, they might come from either the same object or different objects, and they might be either correctly matched with their parts or swapped (disregarding the cases in which the two colors arise from the same part type). (Middle) Here we measure the probability that two reported colors came from the same object, given that both colors originated from one of the two objects adjacent to the target. The y-axis shows the across-subjects mean probability (and 95% confidence interval) that two reported colors came from the same object, as a function of the object shape (x-axis); whether the colors were correctly matched to their parts or were not;

and experiment (panels). The 95% intervals that do not cross the .5 line reflect significant differences from chance, and those dots are marked with a black outline. For both the sequential and simultaneous experiments, and regardless of whether the colors were correctly matched to their parts, most objects show that intrusions are more likely to come from the same object. (Right) The critical question for evaluating whether there is any evidence of whole-object intrusions is whether the rate of such same-object intrusions is higher when the two colors were correctly paired to parts (as would be the case under a whole-object intrusion) than when they are swapped (as would be equally likely under an object-ensemble intrusion). We found that the difference between these two conditional probabilities (γ ; across-subjects means and 95% confidence intervals) is mostly indistinguishable from zero, aside from the case of bullseye-shaped objects

across parts were more likely to arise from the same object than from two different objects. Moreover, consistent with pre-attentive binding of colors to parts, intrusions from two adjacent items were more likely to be correctly matched to their parts than to be swapped. We found no consistent evidence of an extra same-object advantage for matched-part as compared to swapped-part errors, as would be expected from whole-object intrusions. These results provide some evidence that outside the focus of attention, colors are bound to parts, and two colors from the same object are bound together.

Model decomposition of errors

In the previous section we compared specific points in the distribution of conjunctions (as shown in Fig. 2), to ask whether particular kinds of correctly bound intrusions occur more often than chance. We found above-chance rates of intrusions of two colors from the same object, even if they are incorrectly bound to their parts, suggesting that pairs of colors from the same object are somehow bound together. Furthermore, we found that even when colors are reported from different adjacent objects, they are more likely than chance to be correctly bound to their parts, suggesting some amount of pre-attentive part-color binding. Here we endeavor to characterize the full distribution of reported color conjunctions in order to estimate the rates of such whole-object and bound part-color intrusions.

We characterize the distribution of reported conjunctions as a mixture model of six distinct processes: uniform guessing, correct target reports, independent sampling of features, sampling of bound part-color conjunctions, sampling of bound object ensembles, and sampling of whole objects. These six components are illustrated in Fig. 8. Under uniform guessing, both colors are sampled from the set of ten colors with equal probabilities; this is a useful baseline component for estimating overall difficulty and the rate of random guessing. Independent-feature sampling corresponds to the predictions of a pre-attentive feature soup: Colors are sampled with a probability inversely proportional to their distance from the target, but correct binding of colors to parts happens only by chance. Part sampling corresponds to intrusions of colors correctly bound to parts, but the parts not bound to objects: Matched part-color pairs are sampled with a probability inversely proportional to their distance from the target in the display, but critically, sampling of two part-color conjunctions is independent of which object they occurred on. Under object-ensemble sampling, two colors are sampled from one of the presented objects, with a probability inversely proportional to the distance of that object from the target, but the colors are not bound to their parts. Whole-object sampling differs from object-ensemble sampling only in that the colors are correctly bound to their parts.

Another mechanism that might generate structured intrusions is cue mislocalization. If the cue is misperceived as

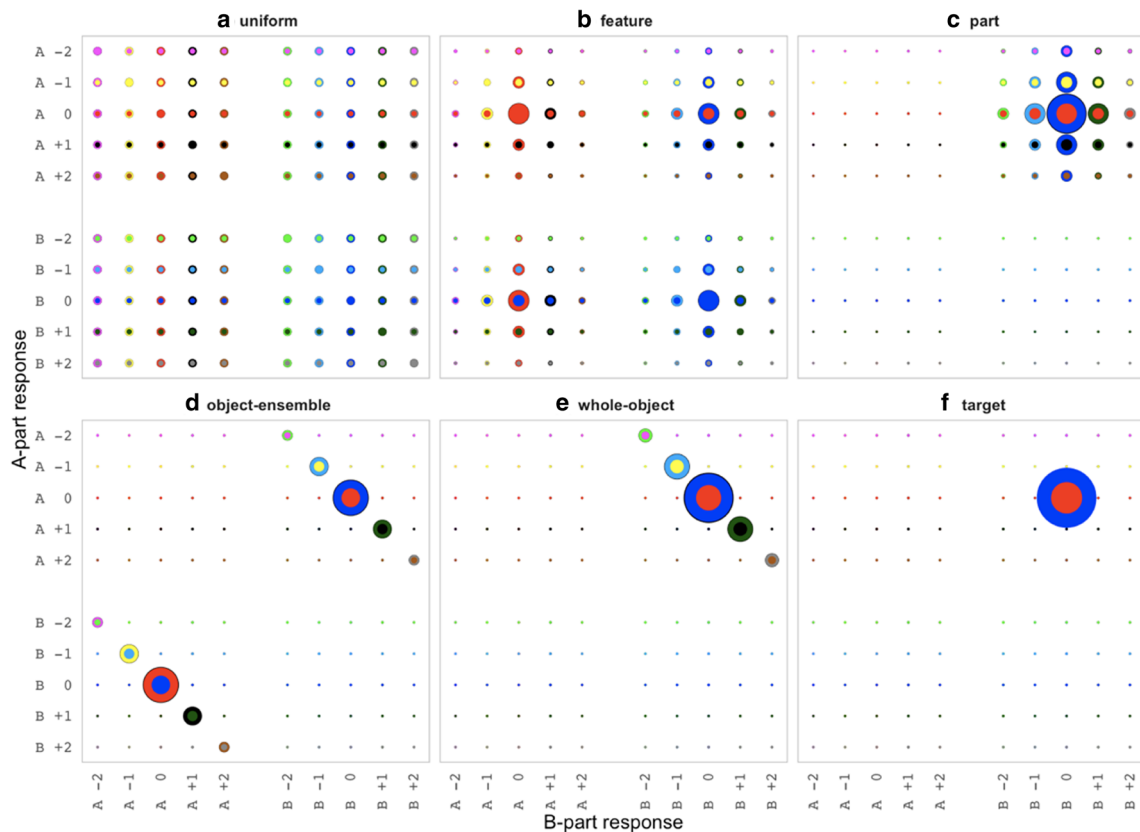


Fig. 8 The components of our error model. Our goal was to estimate what fraction of the time objects arose from these different processes. (a) Uniform guessing. (b) Sampling unbound features within some spatial selection window. (c) Sampling features correctly bound to parts (but parts not bound to objects). (d) Sampling object ensembles: pairs of colors arising from the same object but not bound to their part shapes.

This would arise if features were bound to objects or to each other in feature ensembles, but not necessarily to parts. (e) Sampling whole objects, wherein both colors were correctly matched to their parts and arose from the same object. (f) Sampling just the correct target conjunction. For the feature, part, object-ensemble, and whole-object intrusions, the spatial precision of attentional selection was a free parameter

being slightly clockwise on some trials and slightly counter-clockwise on other trials, then the selected region would have trial-to-trial variability, or noise. Such spatial variability in the selected region would result in a spatial correlation of feature intrusions: if the selected region were shifted clockwise on a given trial, then all features would be more likely to be reported from objects clockwise from the target. Vul and Rich (2010) measured the magnitude of this correlation as they varied the precueing time from 0 to 200 ms, and they found no evidence of such a correlation (although they did find that such a correlation could be easily induced by explicitly adding noise to the cue location). In the present experiments and models, cue mislocalization would be manifest as either color-ensemble intrusions or whole-object intrusions (i.e., two colors reported from the same object, whether or not they were bound). Critically, cue mislocalization would be constant for all object types, as it ought to depend only on the cue and its presentation time, not on what object was being selected. We did not add an explicit term to our models to estimate cue mislocalization noise, because Vul and Rich found no evidence of such noise, and our results indicated that the rates

of whole-object and object-ensemble intrusions were very low and far from stable across object types. That said, whatever stable rate of object-ensemble and whole-object intrusions we found across object types might be interpreted as cue mislocalization noise.

The simplest way to evaluate whether part-color, object-ensemble, or whole-object intrusions occur at any nonnegligible rate was to ask whether the addition of these error components yielded a better fit to the distributions of reported conjunctions than did a model without these components. To this end, we compared six models, each of which corresponded to the addition of one or more nonindependent error components: bound part-color sampling, object-ensemble sampling, or whole-object sampling. Figure 9a shows the improvements per subject in the Akaike information criterion (i.e., reduction in AIC; Akaike, 1974) for each object type and experiment over a model that includes uniform guessing, independent feature intrusions, and target responses. Regardless of response type, for the easier objects, adding bound part-color intrusions offers a reliable improvement in model fit. However, as is shown in Fig. 9b, after adding such

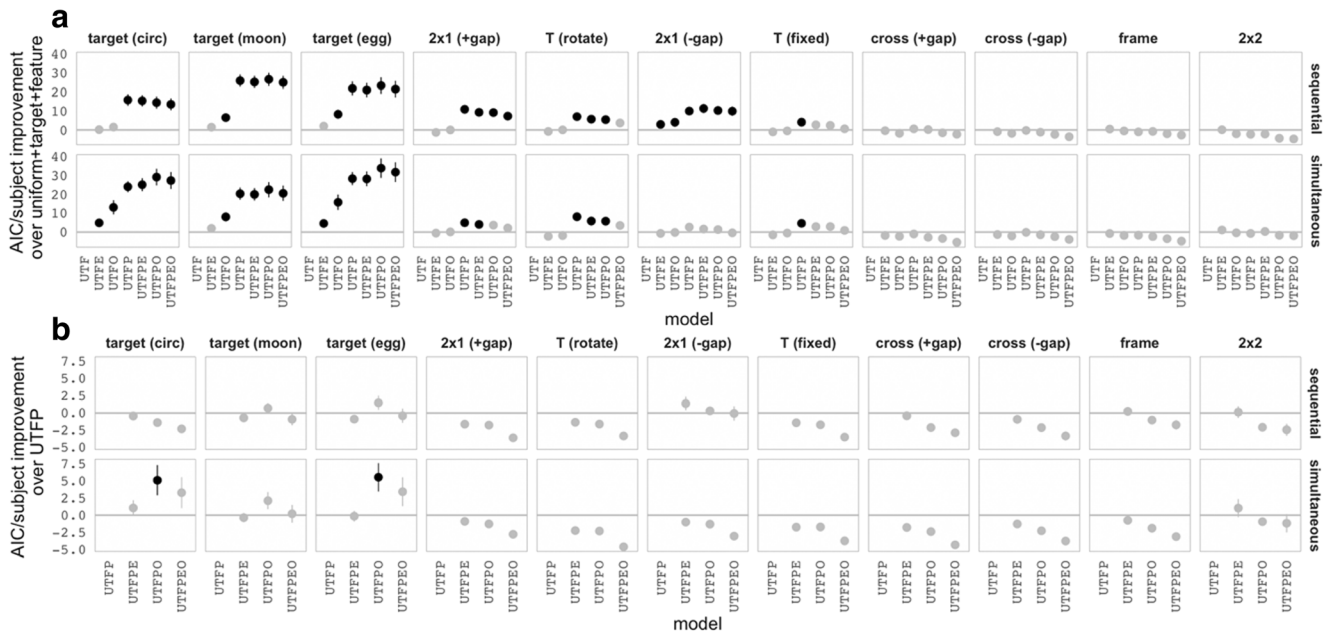


Fig. 9 Comparison of models with different error components, indicated by the abbreviations U (uniform), T (target), F (feature sampling), P (part sampling), E (object-ensemble sampling), and O (whole-object sampling). **(a)** Average (across-subject) improvement in Akaike information criterion (AIC; higher is better) over a model that includes the uniform-guessing, feature-sampling, and target components. For many object types, a model that adds bound part–color intrusions (UFP) is reliably

part–color intrusions, there are no consistent further improvements from adding object-ensemble or whole-object intrusions, suggesting that insofar as those types of intrusions occur, they are fairly rare.

A second way to characterize the response distributions is to estimate the mixture proportions in a complete model that includes all six components: uniform guesses, independent feature intrusions, target responses, bound part–color intrusions, object-ensemble intrusions, and whole-object intrusions. Figure 10 shows the average (across-subjects) mixture proportions for each object type and response type. Across most object types, the rate of object-ensemble and whole-object intrusions is very low (the combined rate for these two is below 3% for all object types), whereas the rate of bound part–color intrusions is roughly at the rate of correct target reports (as high as 17% for some target configurations, and above 10% for half of all object types).

Discussion

The model-based characterization of the full response distributions shows considerable rates of bound part–color intrusions—intrusions wherein the color is correctly matched to its originating part—but negligible rates of other possible structured intrusions, such as object ensembles or whole-object intrusions. These model-based results are in slight conflict with the focused analyses in the previous section, which

better than one that includes only the independent intrusion processes (uniform and feature). **(b)** Improvements in AIC over a model that includes uniform guessing, feature sampling, targets, and bound part–color sampling. For nearly all object types, adding whole-object or object-ensemble intrusions does not yield a better fit, indicating that such structured intrusions are fairly rare

found significant signatures of object-ensemble intrusions and some evidence of whole-object intrusions. These results may be reconciled by considering that the rates of these types of intrusions are estimated to be very low, yet nonzero. The focused analyses may have been picking up on these slight rates by isolating the specific conjunctions that are expected to be disproportionately more frequent, whereas a characterization of the overall distribution is largely insensitive to capturing these slight differences in the frequencies of specific types of conjunctions.

General discussion

In a series of experiments, using both sequential and simultaneous report methods, we asked subjects to identify the colors of each part of two-part objects that varied in different ways. Across 11 different object types, we examined the extent to which the errors were suggestive of bound part–colors, part–objects, or complete objects. We asked whether the distribution of intrusions when reporting the colors of two-part objects revealed any structure indicative of pre-attentive binding of features. Specifically, we looked for signatures of the binding of multiple colors from the same object into a sort of object ensemble, binding of colors to parts, and binding of whole objects. We also characterized the full joint distribution of the reported color–color conjunctions with a mixture model

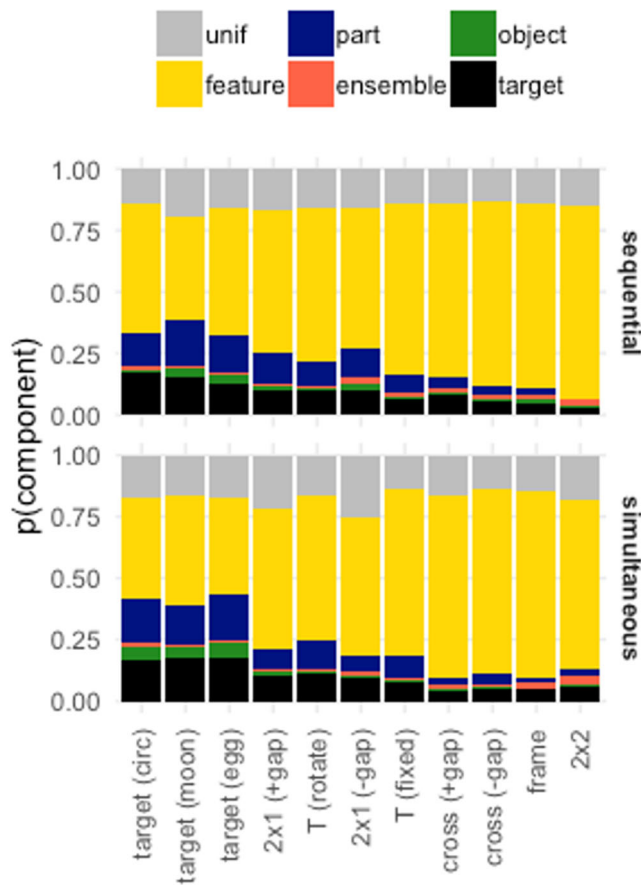


Fig. 10 Mixture proportions of the different response components (uniform guessing, independent feature sampling, target responses, bound part–color intrusions, bound object–ensemble intrusions, and whole-object intrusions) in a full model that allows them all. For most object types, particularly the easier ones, a considerable fraction of the reported conjunctions seem to arise from sampling bound part–color conjunctions

of the error patterns predicted by these types of intrusions. We found consistent evidence that intrusions from adjacent objects tend to be correctly matched to their parts, indicating

some amount of pre-attentive part–color binding; this was also borne out in the considerable proportions estimated for such part–color intrusions in our mixture model. Although we found evidence of the signature of object–ensemble intrusions, the rate of such intrusions seems to be very low in the mixture model estimates (Fig. 10).

These features of the distribution of illusory conjunctions suggest that binding may be a hierarchical process. Colors may be bound to object parts, without those parts being bound to objects, and to a lesser extent, co-occurring colors may be bound to each other but not associated with particular shapes. In the end, it remains the case that correct apprehension of a multifeature, multipart object relies on an object being isolated by selective attention, as was proposed by Treisman (Treisman & Gelade, 1980), but outside the focus of attention, different types of features appear not to be represented as a completely undifferentiated soup, but instead exhibit some coherent structure, consistent with the part–whole search efficiencies reported by Wolfe and colleagues (Bilksy & Wolfe, 1995; Wolfe et al., 1994).

The extent to which colors are pre-attentively bound to parts, or to each other, seems to vary considerably as a function of the geometry of the objects in which they appear. Target/bulls-eye-shaped objects appear to have the greatest degree of pre-attentive binding and structured illusory conjunctions, and they are also the easiest (in terms of overall accuracy in reporting the correct target conjunction). This is consistent with the results of Wolfe et al. (1994), who found that conjunctions wherein one color is surrounded by another are easier to identify than other configurations. However, the next most bindable (and easiest) sets of objects are Ts and two abutting squares (2×1 in our stimulus labels), which seem more conducive to such pre-attentive binding than crosses or 2×2 shapes. One possibility is that the advantage of target-shaped objects is a consequence of the imbalance of total area of the two parts in these circular geometries; however, that

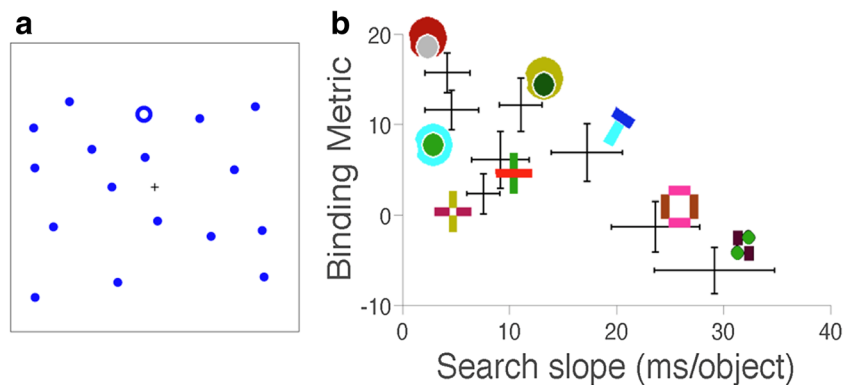


Fig. 11 An informal experiment to evaluate the relationship between apparent hierarchical binding and shape discriminability. We measured the ease of discriminating part shapes with a visual search experiment in which subjects needed to find one singleton part type among distractors that consisted of the other part of the given shape. (a) One such search

display for circular target objects. (b) Search slopes for finding one part among others were inversely related to the propensity for binding colors to those parts, suggesting that whatever aspects of object geometry make colors more likely to bind to the object parts also make those parts more easily discriminable

would not explain why Ts and abutting squares have greater rates of part–color binding than crosses. Another possibility is that some geometries have fewer parts; crosses and 2×2 shapes may be construed as four-part objects, which might yield a greater binding problem. Although these explanations seem plausible, they should be considered only speculative for now.

Another possibility is that some aspects of shape or extent are coded together with color. If so, we would expect that part shapes that are more discriminable in the basic features that are coded along with color would exhibit more part–color binding. We ran an informal experiment to assess whether there is such a relationship. Specifically, for a subset of object types, we measured the extent to which colors are bound to parts; this binding metric was the logarithm ratio of the rates of part–color and whole-object intrusions over feature intrusions. We also ran visual search experiments in which subjects were asked to identify a singleton part A among part B distractors for that object type (or vice versa; Fig. 11a shows one search display for finding the outer ring of the circle–target among inner-ring distractors), and we estimated the search slope as we varied the number of distractors for that object type. Figure 11b shows that objects that had greater amounts of binding also had shorter search slopes, but with considerable variability. This result suggests that whatever aspects of object geometry make colors more likely to bind to parts are the same features that make parts more easy to discriminate visually. This result seems consistent with the notion that the same properties of part shape that make them easy to discriminate also are coded with color, and thus enable some amount of part–color binding. However, the specific features that vary across these object types are not really consistent with such a story: Orientation-contingent color aftereffects suggest that orientation is coded along with color in early stages of visual processing (Vul & MacLeod, 2006), which should make crosses or Ts more likely to exhibit part–color binding than circles, but that is not the case. Thus, this remains another plausible, but only speculative, explanation.

Understanding why some geometries are more conducive to such pre-attentive binding might resolve the tension between these results and the existing literature on illusory conjunctions. For instance, if outside the focus of attention parts are bound to colors, then joint reports of the color and identity of a cued target ought to be somewhat correlated, but we found zero evidence of such correlations when people were asked to report both the color and specific letter of a cued target (Vul & Rich, 2010). It is possible that whatever caused the variation in pre-attentive binding across our geometries here also conspires to reduce or eliminate such binding in the classic color–letter stimuli used in binding experiments.

Although we cannot adequately characterize why some object geometries are more conducive to pre-attentive binding, or why our results seem to differ from the results for

color–letter conjunctions, it remains the case that such binding outside the focus of attention seems to be a consistent feature for some objects. This means that, at least for some stimulus configurations, features outside of attention are not represented independently but show some reliable structure.

Acknowledgements All data and analysis code for this article are available in this public github repository: <https://github.com/vullab/ColorBinding/>. ANR is supported by an Australian Research Council Discovery Project Grant (DP170101780).

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, AC-19*, 716–723. doi:<https://doi.org/10.1109/TAC.1974.1100705>
- Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*, 392–398. doi:<https://doi.org/10.1111/j.1467-9280.2008.02098.x>
- Ashbridge, E., Walsh, V., & Cowey, A. (1997). Temporal aspects of visual search studied by transcranial magnetic stimulation. *Neuropsychologia, 35*, 1121–1131.
- Ashby, F. G., Prinzmetal, W., Ivry, R., & Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review, 103*, 165–192. doi:<https://doi.org/10.1037/0033-295X.103.1.165>
- Bilsky, A. B., & Wolfe, J. M. (1995). Part–whole information is useful in visual search for size × size but not orientation × orientation conjunctions. *Perception & Psychophysics, 57*, 749–760.
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science, 22*, 384–392. doi:<https://doi.org/10.1177/0956797610397956>
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*, 393–404. doi:[https://doi.org/10.1016/S0042-6989\(02\)00596-5](https://doi.org/10.1016/S0042-6989(02)00596-5)
- Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research, 45*, 891–900. doi:<https://doi.org/10.1016/j.visres.2004.10.004>
- Cohen, A., & Ivry, R. (1989). Illusory conjunctions inside and outside the focus of attention. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 650–663. doi:<https://doi.org/10.1037/0096-1523.15.4.650>
- Cohen, A., & Rafal, R. D. (1991). Attention and feature integration: Illusory conjunctions in a patient with a parietal lobe lesion. *Psychological Science, 2*, 106–110.
- Emrich, S. M., & Ferber, S. (2012). Competition increases binding errors in visual working memory. *Journal of Vision, 12*(4), 12. doi:<https://doi.org/10.1167/12.4.12>
- Found, A. (1998). Parallel coding of conjunctions in visual search. *Perception & Psychophysics, 60*, 1117–1127. doi:<https://doi.org/10.3758/BF03206162>
- Friedman-Hill, S. R., Robertson, L. C., & Treisman, A. (1995). Parietal contributions to visual feature binding: Evidence from a patient with bilateral lesions. *Science, 269*, 853–856.
- Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review, 114*, 599–631. doi:<https://doi.org/10.1037/0033-295X.114.3.599>
- Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the limits of human visual awareness. *Science, 317*, 823–825.

- Humphrey, G. K., & Goodale, M. A. (1998). Probing unconscious visual processing with the McCollough effect. *Consciousness and Cognition*, 7, 494–519.
- Kanwisher, N. G. (1991). Repetition blindness and illusory conjunctions: Errors in binding visual types with visual tokens. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 404–421. doi:<https://doi.org/10.1037/0096-1523.17.2.404>
- Lew, T. F., & Vul, E. (2015). Ensemble clustering in visual working memory biases location memories and reduces the Weber noise of relative positions. *Journal of Vision*, 15(4), 10. doi:<https://doi.org/10.1167/15.4.10>
- Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, 240, 740–749. doi:<https://doi.org/10.1126/science.3283936>
- Nordfang, M., & Wolfe, J. M. (2014). Guided search for triple conjunctions. *Attention, Perception, & Psychophysics*, 76, 1535–1559. doi:<https://doi.org/10.3758/s13414-014-0715-2>
- Treisman, A. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology*, 40A, 201–237. doi:<https://doi.org/10.1080/02724988843000104>
- Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society B*, 353, 1295–1306. doi:<https://doi.org/10.1098/rstb.1998.0284>
- Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, 14, 411–443. doi:<https://doi.org/10.1080/13506280500195250>
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136. doi:[https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological Review*, 95, 15–48. doi:<https://doi.org/10.1037/0033-295X.95.1.15>
- Treisman, A., & Paterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 12–31. doi:<https://doi.org/10.1037/0096-1523.10.1.12>
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107–141. doi:[https://doi.org/10.1016/0010-0285\(82\)90006-8](https://doi.org/10.1016/0010-0285(82)90006-8)
- Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: Selection is probabilistic; responses are all-or-none samples. *Journal of Experimental Psychology: General*, 138, 546–560. doi:<https://doi.org/10.1037/a0017352>
- Vul, E., & MacLeod, D. I. (2006). Contingent aftereffects distinguish conscious and preconscious color processing. *Nature Neuroscience*, 9, 873–874. doi:<https://doi.org/10.1038/nn1723>
- Vul, E., & Rich, A. N. (2010). Independent sampling of features enables conscious perception of objects. *Psychological Science*, 21, 1168–1175.
- Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research*, 37, 25–43. doi:[https://doi.org/10.1016/S0042-6989\(96\)00111-3](https://doi.org/10.1016/S0042-6989(96)00111-3)
- Wolfe, J. M., & Cave, K. R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron*, 24, 11–17. doi:[https://doi.org/10.1016/s0896-6273\(00\)80818-1](https://doi.org/10.1016/s0896-6273(00)80818-1)
- Wolfe, J. M., Friedman-Hill, S. R., & Bilsky, A. B. (1994). Parallel processing of part-whole information in visual search tasks. *Perception & Psychophysics*, 55, 537–550. doi:<https://doi.org/10.3758/BF03205311>
- Wolfe, J. M., Yu, K. P., Stewart, M. I., Shorter, A. D., Friedman-Hill, S. R., & Cave, K. R. (1990). Limitations on the parallel guidance of visual search: Color × Color and Orientation × Orientation conjunctions. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 879–892. doi:<https://doi.org/10.1037/0096-1523.16.4.879>

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