Visual Feature Integration in a World of Objects

William Prinzmetal
University of California, Berkeley

Visual Feature Integration in a World of Objects

Take away the sensations of softness, moisture, redness, tartness, and you take away the cherry, since it is not a being distinct from sensations. A cherry, I say, is nothing but a congeries of sensible impressions.
- Berkeley, 1713

Throughout the history of psychology, the study of perception has been marked by a conflict between analytic and holistic views. According to the analytic view, object recognition involves analysis of the visual input of stimulus features. A stimulus is recognized to the extent that it contains the appropriate features and these features are registered by the visual system. Early examples of the analytic approach include the British Empiricists and the Structuralists. The holistic view, described by the Gestalt school, claims that recognition involves more than analysis of a stimulus into parts. According to this view, a collection of features is not sufficient for object recognition. These apparently conflicting views lead to the following question: Do we recognize a cherry because we register the features round and red, or is a cherry more than an assortment of features? In studying the perceptual phenomenon of illusory conjunctions
we have found that rather than being in opposition, the analytic and holistic approaches to perception complement each other. Veridical object recognition depends not only on the perception of features, but also on correctly combining features. How features are combined is largely determined by principles enumerated by the Gestalt school.

Treisman and Schmidt first described the phenomenon of illusory conjunctions in 1982. In several experiments they found that subjects perceived colors and shapes in incorrect combinations. For example, in one experiment, subjects were briefly presented with stimuli consisting of three colored letters, flanked by two black digits. On each presentation, subjects had two tasks. The first task was to report the two digits, and the second task was to report as many of the colored letters as they could. Treisman and Schmidt found that subjects reported colors and letters in incorrect combinations on 40% of the trials. For example, if a subject was shown a pink X next to a green T, the subject might report seeing a green X. They called these errors illusory conjunctions because there are illusory (i.e., incorrect) combinations of presumably correctly perceived features. Illusory conjunctions of color and shape have been observed by a number of investigators using a variety of tasks.

Illusory conjunctions between features of shape have also been observed. For example, Prinzmetal briefly presented subjects with stimuli that consisted of an array of circles that contained vertical or horizontal lines (see Figure 1). Subjects reported whether or not a plus sign was present in the display. On target-absent trials, subjects incorrectly reported a plus sign more often when the stimulus contained the two features of a plus sign (i.e., vertical and horizontal lines, Figures 1b and 1e) than when it contained only one feature (i.e., two vertical or two horizontal lines, Figures 1c and 1f). Thus a plus sign could be falsely perceived if the stimulus contained its constituent features.

---

It should be noted that illusory conjunctions are a genuine perceptual phenomenon and they can not be accounted for by short-term memory limits or simple guessing. The hypothesis that illusory conjunctions are due to short-term memory limits is as follows: Although subjects correctly perceive a stimulus display, they forget what they have seen by the time they report it. In a whole report task, where a subject is asked to report everything present in a display, it is easy to see how this might happen. For example, in Treisman and Schmidt’s experiment described above, subjects had to keep two digits, three colors, three letters, and the combination of colors in letters in short-term memory. However, the memory limitation hypothesis does not explain the occurrence of illusory conjunctions in a detection task. In a typical detection task, the subject might be asked whether or not the display contained a particular colored letter (e.g., a green T). Subjects are more likely to report incorrectly seeing a target if the display consists of items that include all of the features of the target (e.g., green X and red T) than if the display does not
contain the features of the target (e.g., red X, blue T). In the detection task, subjects need only keep one color and one letter in memory to report the target correctly. Unlike the whole report task, subjects are not required to keep several colors and letters in memory. Hence the illusory conjunctions that are observed in a detection task cannot be attributed solely to limits in short-term memory.

Although illusory conjunctions cannot be attributed to memory limitation, not all of the responses labeled illusory conjunctions are due to the incorrect combination of features in perception. For example, a subject might report the presence of a green X in a display that contains a green T and red X because he or she incorrectly combined the features green and X. Alternatively, the subject might not have perceived one or both of the features. If the subject perceived the target letter (i.e., X), but not the color (e.g., red), the subject would have to guess a color. On some trials the subject would guess green and the response would be counted as an illusory conjunction.

Ashby, Prinzmetal, Ivry, and Maddox compared many different probabilistic (i.e., stochastic) response models. One class of models assumed that all of the illusory conjunction reports are due to guessing, and the other class of models assumed that some of the illusory conjunction reports were due to genuine incorrect feature combination. In general, models that assumed that some of the illusory conjunction reports were due to feature integration errors provided a better fit of the data than models that assumed all of the illusory conjunctions were due to guessing. The only exception to this finding was when the letters were far apart. When the letters were far apart, there were relatively few illusory conjunction reports and these responses could be entirely accounted for by guessing. In general, however, illusory conjunction reports in detection tasks cannot be accounted for by guessing.

Illusory conjunctions are a more general phenomenon than previously supposed. Heretofore, all of the experiments involving illusory conjunctions have used a brief exposure. Furthermore, in most of the experiments subjects had to perform an attention demanding task in addition to the task in which illusory conjunctions were obtained. For example, Treisman and Schmidt had subjects report two black digits before reporting the colored letters. Treisman proposed that diverting attention was a prerequisite for obtaining illusory conjunctions. However, we found that illusory conjunctions can be obtained readily with relatively long exposure durations (i.e., 1.5 seconds) and without diverting attention.

In one experiment, we compared performance in two conditions: with an attention demanding task and without such a task. In the attention demanding task, subjects were presented with a rapidly changing sequence of single digits, presented in the center of a computer monitor, for five seconds (see Figure 2). The primary task was to press a button whenever the digit '1' appeared. Approximately 3.5 seconds into the five second trial, two colored letters, a target letter and a nontarget letter, appeared in the periphery for 1.5 seconds. The target letter was either an X, T, or L and the nontarget letter was
an O. (The colored target and nontarget letters were flanked by two white O’s.) After the five second trial, the screen became blank and the subject reported the color (red, green, or blue) and identity (X, T, or L) of the target letter. In the condition without the attention demanding task, subjects merely had to maintain fixation in the center of the screen for 1.5 seconds (eye movements were monitored). To control overall accuracy, the eccentricity of the colored letters (d in Figure 2) was varied for each subject so that for each subject and condition, performance was about 75% correct. The average eccentricity, with and without the attention demanding task, was 8.4 and 7.6 degrees of visual angle, respectively. We obtained as many illusory conjunctions as in previous experiments using a brief exposure, and there was no difference in the percent of illusory conjunctions with or without diverting attention.7

Insert Figure 2 about here

The robustness of the phenomenon of illusory conjunctions in the laboratory argues for the analytic view of perception. Subjects occasionally incorrectly recognize a target green X when the stimulus contains the features green and X. Thus a green X might seem to consist of "congeries of sensory impressions", as proposed by Berkeley. However, there are a number of questions about object features in illusory conjunction experiments. For example, Treisman claimed that these features represent "early" perceptual processes. However, it is not clear what "early" means. From the present research, we can not determine where these features are represented in the visual system. There is evidence that the "features" in many psychological experiments do not correspond to the tuning functions of cells in the primary visual cortex (V1).8 Furthermore, different types of features could be represented in different cortical areas: Vertical and horizontal lines might be represented anywhere from the primary visual area (V1) to the temporal lobe, whereas letter shapes are probably not represented in the primary visual area.9 It might be possible to discover where these features are represented in the visual system by relating feature integration to other perceptual processes. For example, as discussed below, illusory conjunctions are more likely between items that are similar in color than items that are dissimilar in color. If the similarity effect is influenced by color constancy, then the features are represented in a part of the visual system that shows color constancy (e.g., V4).10

If illusory conjunctions can be obtained so easily in the laboratory, whether or not attention is diverted or exposure duration is brief, the question that must be addressed by visual science is why they do not often occur outside the laboratory. In order to understand veridical feature integration, it is necessary to understand why illusory conjunctions occur. Our working hypothesis is that illusory conjunctions occur because the perceived location of features is sometimes in error. If a feature of one object is perceived to be in the location of another object, an illusory conjunction may occur. In many cases, such coarse location information is
sufficient for perception. If one looks at a leopard, it is important to note that the leopard has spots but it is not necessary to register the precise location of each spot. The spots can migrate without changing the identity of the animal. However, it is important that the visual system keep the features of one object from migrating to another object: The spots of the leopard should not perceptually migrate to the tiger in an adjacent cage. In other words, the visual system should use information about perceptual organization to constrain where features are localized and therefore how they are joined.

The insights of the Gestalt school of psychology give us a hint as to how the visual system correctly combines or integrates features. The Gestalt school proposed a series of "laws" that describe how objects form perceptual groups. For example, we tend to group items that are close together (i.e., proximity grouping - the grouping of circles in Figure 1 into columns), are similar (similarity grouping), or that form a line (good continuation grouping). The laws of perceptual organization describe how the visual system forms objects - we group the leopard’s spots into a single creature. Given this, the laws of perceptual organization should constrain feature integration in the following way: Features from different perceptual groups should be less likely to form illusory conjunctions than features from the same perceptual group. Hence, any factor that affects perceptual organization should also affect the occurrence of illusory conjunctions. Referring to Figure 1, subjects were approximately 1/3 more likely to incorrectly combine the vertical and horizontal lines in 1b than in 1e because the features in 1b are part of the same perceptual group (i.e., objects) whereas the features in 1e are not.

We have examined the effect of several of the Gestalt laws on feature integration. One of the strongest grouping principles is proximity. Treisman and Schmidt reported that the distance between items did not affect the occurrence of illusory conjunctions. However, since then numerous investigators have found that illusory conjunctions are more likely to occur between items that are close together than between items that are far apart. Other factors that affect perceptual organization are similarity and proximity. Illusory conjunctions are more likely to occur between items that are similar in color or in shape. In addition, illusory conjunctions are affected by good continuation when good continuation is manipulated independently of proximity and similarity. In all of these cases, the visual system avoids combining features from different perceptual groups.

Perceptual grouping is a consequence of how an observer processes a stimulus, but all of the above examples involved stimulus manipulations. Attneave demonstrated that an evenly spaced matrix of items, such as floor tiles, has an ambiguous organization. For example, the matrix of items in Figure 3 can be perceptually organized into either rows or columns. If one simultaneously reads the horizontally aligned digits (3 and 8) the matrix tends to organize into rows. If one reads the vertically aligned digits (7
and 2), the matrix tends to organize into columns. Prinzmetal and Keysar\textsuperscript{14} found that illusory conjunctions were more likely to occur between letters within these subjectively defined units (rows or columns) than between items in different units. That is, when subjects read the horizontally aligned digits, illusory conjunctions were more likely to occur between items within the same row. When they read the vertically aligned digits, illusory conjunctions were more likely to occur between items within the same column.

Insert Figure 3 about here

The perceptual organization of a stimulus can be influenced by acquired, cognitive structures. Printed words have an organization due to orthographic rules and morphology. For example, most observers would parse the words VODKA and SUNUP after the third letter and the words ADMIT and TODAY after the second letter. The question is whether this kind of perceptual organization affects feature integration. This question was investigated in a series of experiments in which subjects were briefly presented words composed of colored letters.\textsuperscript{15} The task was to indicate whether or not a target letter was present and, if so, to identify its color. The structure of printed words had a marked influence in feature integration. For example, with a word like VODKA, subjects were two to three times more likely to incorrectly respond that the D was the color of the V or O than the K or A. We found that orthographic structure, as determined by two consonants that do not co-occur within an English syllable, affect illusory conjunctions (e.g., VOD\textbullet KA). Morphological structure also affects feature integration with compound words (e.g., PIN\textbullet UP; FAT\textbullet HEAD) and bound morphemes (e.g., END\textbullet ED). Surprisingly there is no influence of purely phonological structure. For example, with words like PI\textbullet VOT subjects are not more likely to respond that the V is the color of the O or the T than the P or the I. It is important to note that these results are not due to a guessing bias and they represent the operation of unconscious processes.\textsuperscript{16} When subjects are questioned about the factors that affect their responses, they almost never mention anything related to syllable-like structure. The visual system unconsciously parses the letter array into groups, and uses information about the perceptual structure to combine or integrate features.

Rather than being in opposition, the analytic and holistic views nicely complement each other. On the one hand, we can show that the functional features subjects use to recognize objects can be incorrectly combined to form an illusory object. On the other hand, "the intrinsic nature of the whole" (i.e., the Gestalt laws) is one factor that constrains how the visual system combines this information. Indeed, it is difficult to imagine a complete theory of object recognition that does not combine elements of both analytic and Gestalt views of perception.
Notes
7. This is not to claim that attention has no influence on feature integration {e.g., see Prinzmetal, W., Presti, D. E., & Posner, M. I. (1986). Does attention affect visual feature integration? Journal of Experimental Psychology: Human Perception and Performance, 12, 361-369.}
However, it is not necessary to divert attention to obtain these errors.


Figure Captions
1. Illusory conjunctions occur between features of shape. When the stimulus contained all of the features of a target plus (b and e), subjects were more likely to perceive an illusory plus than when the stimulus contained only one feature of a plus (c and f).
2. Sample stimuli for illusory conjunctions with a long exposure. In the attention demanding condition, subjects monitored a rapid sequence of changing digits in the center of the screen for a target digit. Colored letters were presented in the periphery for 1.5 seconds. Over accuracy was controlled by varying the distance from the fixation (d in the figure).
3. Ambiguously organized stimulus used to investigate illusory conjunctions. When subjects read the horizontal digits, they were more likely to report seeing a yellow X than a green X, but when they read the vertical digits, they were more likely to report seeing a green X than a yellow X. Illusory conjunctions were more likely within rows when rows were the subjectively defined groups and within columns when columns were the subjectively defined groups.