Journal of Experimental Psychology:
Human Perception and Performance
2002, Vol. 28, No. 2, 251–269

0096-1523/02/$5.00 DOI: 10.1037//0096-1523.28.2.251

A Measurement Theory of Illusory Conjunctions

William Prinzmetal and Richard B. Ivry
University of California, Berkeley

Diane Beck
University College London

Naomi Shimizu
University of Hawaii at Manoa

Illusory conjunctions refer to the incorrect perceptual combination of correctly perceived features, such as color and shape. Research on the phenomenon has been hampered by the lack of a measurement theory that accounts for guessing features, as well as the incorrect combination of correctly perceived features. Recently, several investigators have suggested using multinomial models as a tool for measuring feature integration. The authors examined the adequacy of these models in 2 experiments by testing whether model parameters reflect changes in stimulus factors. In a third experiment, confidence ratings were used as a tool for testing the model. Multinomial models accurately reflected both variations in stimulus factors and observers' trial-by-trial confidence ratings.

The goal of the present research was to test a formal measurement theory of the perceptual phenomenon of illusory conjunctions. This phenomenon was first discussed by Treisman and Schmidt in 1982, providing one of the key pieces of evidence for Treisman's feature integration theory (Treisman & Gelade, 1980). Observers, when briefly presented with strings of colored letters, sometimes reported colors and letters in incorrect combinations. For example, when the stimulus consisted of a green *N*, red *X*, and blue *O*, the response might be "red N." Treisman and Schmidt termed these errors illusory conjunctions because they were incorrect (illusory) conjunctions of perceived features.

Illusory conjunctions are a robust phenomenon (for a complete review, see Prinzmetal, 1995) and can be found with many different stimulus features (e.g., Butler, Mewhort, & Browse, 1991; Gallant & Garner, 1988; Lasaga & Hecht, 1991; Treisman & Paterson, 1984). For example, Prinzmetal (1981) found that people sometimes incorrectly combine vertical and horizontal lines to form an illusory plus sign. Illusory conjunctions are found in whole report tasks, visual search tasks, and matching tasks. The observation that illusory conjunctions are found in visual search and matching tasks, as well as whole report tasks, has made clear that the errors are not due to memory limitations but appear to be perceptual in nature. Moreover, observers rate their confidence high on trials in which their report involves the incorrect combination of stimulus features. Illusory conjunctions can also be obtained under extended viewing conditions (2 s of exposure), even without diverting attention (Prinzmetal, Henderson, & Ivry, 1995; see also Prinzmetal, Presti, & Posner, 1986).

A variety of stimulus factors have been found to affect feature integration. A plethora of data have demonstrated that illusory conjunctions are more likely between adjacent than distant items (e.g., Ashby, Prinzmetal, Ivry, & Maddox, 1996; Chastain, 1982; Hazeltine, Prinzmetal, & Elliot, 1997; Keele, Cohen, Ivry, Liotti, & Yee, 1988; Prinzmetal & Millis-Wright, 1984; Prinzmetal & Keysar, 1989; Wolford & Shum, 1980). Organizational factors, such as similarity, good continuation, and common fate, affect illusory conjunctions (Baylis, Driver, & McLeod, 1992; Gallant & Garner, 1988; Ivry & Prinzmetal, 1991; Khurana, 1998; Lasaga & Hecht, 1991; Prinzmetal, 1981, 1995; Prinzmetal & Keysar, 1989). The general rule is that features that are part of the same perceptual group or object are more likely to be incorrectly combined (form an illusory conjunction) than features from different perceptual groups. These effects might not be surprising in that the purpose of correctly binding features is to enable object recognition in displays with several objects. Surprisingly, the effect of perceptual organization extends to syllable or syllable-like units in printed words. When presented with a string of colored letters such as *MAYBE*, observers are more likely to report that the *Y* is the color of the *M* or *A* (within a syllable) than to report that the *Y* is the color of the *B* or *E* (e.g., Prinzmetal, Treiman, & Rho, 1986; Prinzmetal, Hoffman, & Vest, 1991; Rapp, 1992; Seidenberg, 1987; also see Prinzmetal & Millis-Wright, 1984; Prinzmetal, 1990). Finally, we have begun to locate areas in the brain that appear to be critical for feature integration (e.g., Arguin, Cavanagh, & Joanette, 1994; Cohen & Rafal, 1991; Friedman-Hill, Robertson, & Treisman, 1995).

The empirical findings just reviewed are impressive, but most of the studies involved have lacked a theoretically justified, empirically verified method for measuring illusory conjunctions. This is a potentially serious limitation: As we discuss subsequently, re-

William Prinzmetal and Richard B. Ivry, Department of Psychology, University of California, Berkeley; Diane Beck, Department of Psychology, University College London, London, England; Naomi Shimizu, John A. Burns School of Medicine, University of Hawaii at Manoa.

This research was supported by National Science Foundation Grant SBR-9319103.

Correspondence concerning this article should be addressed to William Prinzmetal, Department of Psychology, University of California, Berkeley, California 94720-1650. E-mail: wprinz@socrates.berkeley.edu

porting an incorrect conjunction of features does not necessarily mean that an illusory conjunction of features occurred. Consider a task in which the stimulus for each trial consists of two colored letters, a colored target letter (*X* or *T*) and a colored nontarget letter (*O*). The colors of the letters are chosen without replacement from a set of four possible stimulus colors (red, green, blue, and yellow). The observer's task is to report the target letter (e.g., red *X*). Suppose the display contained a red *X* and a blue *O* and the response was "blue X." This might result from an illusory conjunction. That is, the observer might have incorrectly combined the color blue with the target letter *X*, perceiving a blue *X*. On the other hand, several other perceptual states could have led to the report "blue X." For example, observers might have perceived the letter *X*, but not its color, and simply guessed blue.

What is needed is a measurement theory that distinguishes between incorrect combination of features and guesses that appear to be illusory conjunctions. We refer to objective reports in which features from different objects are combined together as *conjunction responses* (e.g., blue *X* in the example). We use the term *illusory conjunctions* to refer to the subset of these responses in which the percept is actually the result of incorrectly conjoining features from different objects. The task then is to estimate the rate of true incorrect feature combinations (illusory conjunctions) from the conjunction responses.

Different procedures have been adopted to correct for guessing in the context of illusory conjunction experiments. However, the theoretical justification for these methods is rather murky and, as shown by Ashby et al. (1996), can lead to erroneous conclusions. These researchers (see also Donk, 1999; Prinzmetal et al., 1995) developed an alternative procedure to correct for guessing. This correction procedure was part of a general theory of feature binding (Ashby et al., 1996). However, for the present purpose, we can restrict our discussion to the part of the theory concerned with estimating the true probability of correct feature binding.

To understand correction for guessing, consider the simple case of a student taking a four-alternative multiple-choice test. The probability correct, $P(C)$, is the sum of items the student knew, $P(K)$, and correct guesses:

$$P(C) = P(K) + \{1/4 \times [1 - P(K)]\}. \tag{1}$$

$P(K)$, the proportion of items the student knew, can be calculated algebraically. The model that we present here for analyzing the responses in illusory conjunction experiments does not have an algebraic solution, so $P(K)$ must be found by an iterative search through possible solutions. The simple problem just described is a binomial problem, because there are only two outcomes for each item (correct and incorrect). In most illusory conjunction experiments, there are many outcomes; hence, these experiments involve multinomial problems.

The multinomial approach was developed to analyze experiments in source memory (e.g., Johnson & Raye, 1981) and has been extended in a number of directions by Batchelder and Riefer (1990; Riefer & Batchelder, 1988; Riefer, Hu, & Batchelder, 1994; cf. Banks, 2000). There is an interesting computational similarity between source memory and illusory conjunction experiments. In a source memory experiment, participants must correctly combine the source of an item of information with that item (e.g., Did you discover that fact in the *National Inquirer* or the *New York Times*?). In an illusory conjunction experiment, observers must correctly combine information from two sensory features (e.g., Was the *X* red or blue?). The details of the models are, however, substantially different.

Table 1
Sample Stimulus Display and Response Categories in a Simple Feature Integration Experiment

	Stimulus: $X_{\mathrm{red}}\ O_{\mathrm{blue}}$	
Response	Code	Description
X_{red}	CT	Correct response
X_{blue}	CN	Conjunction response
X_{green}	CO	Color feature error
T_{red}	IT	Letter feature error
T_{blue}	IN	Letter feature conjunction error
T_{green}	IO	Color letter feature error

The first step in developing a multinomial model is to enumerate all of the possible outcomes of a trial (see Dodson, Prinzmetal, & Shimamura, 1998, for a tutorial on multinomial model development). In the example given earlier, there are two colored letters on each trial, a target letter (*X* or *T*) and a nontarget letter (*O*), with four possible colors. The observer's task is to name the color and identity of the target (a red *X* in the example). As can be seen in Table 1, there are six possible outcomes on each trial. We designate each response type with a two-letter code. The first letter indicates whether the letter identity is *correct* (C) or *incorrect* (I). The second letter indicates the type of color response. The observer could respond with the *target* color (*T;* red in the example), the color of the *nontarget* letter (*N;* blue in the example), or some *other* color not present in the display (*O;* e.g., green or yellow). Thus, CT indicates a totally correct response, CN indicates a conjunction response, CO indicates that the correct letter was reported along with a color that was not part of the display, and so on. The descriptions listed in Table 1 were adapted from Prinzmetal et al. (1995) and Ashby et al. (1996).

The second step in developing a multinomial model is to postulate psychological parameters that might lead to particular responses. In our basic illusory conjunction model, we postulate the following five parameters[1]: *TL* (the probability of correctly identifying the target letter), *TC* (the probability of correctly identifying the target color), *NC* (the probability of correctly identifying the nontarget color), α (the probability of correctly binding the target color to the target letter), and *g* (a guessing parameter, described subsequently).

The next step is to develop a tree diagram that depicts how the various response categories arise. Our basic model is presented in Figure 1. This model allows for the occurrence of illusory conjunctions but also has branches in which conjunction responses

[1] With six response categories and five parameters, it seems that a perfect fit of the data would be guaranteed regardless of the validity of the model. However, in our experiments there are always additional independent variables, so the number of parameters is always much less than the number of data points. Furthermore, even in the simplest case, there are internal constraints in the model, so a perfect fit is not guaranteed (see Riefer et al., 1994).

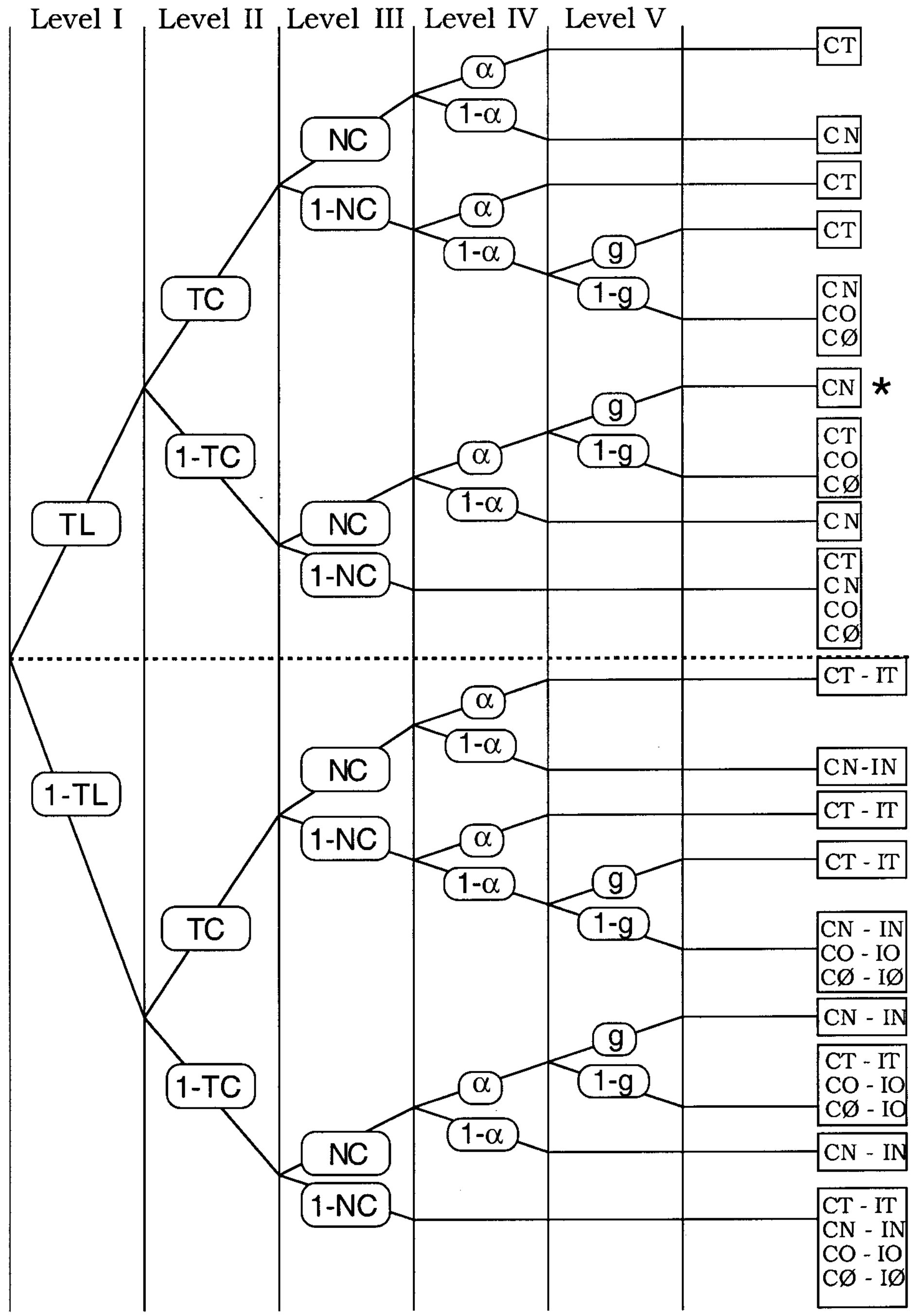

Figure 1. Tree diagram for simple illusory conjunction model. The parameters *TL, TC, NC,* α, and g are indicated in ovals. TL = probability of perceiving the target letter; TC = probability of perceiving the target color; NC = probability of perceiving the nontarget color; α = probability of correctly binding colors and letters; g = a guessing parameter. The responses are indicated in boxes on the right. The first letter of the response code indicates whether the letter was correct (C) or incorrect (I). The second letter indicates whether the color response was the target color (T), nontarget color (N), or one of the two other colors that were not part of the display (O or Ø). The path ending with an asterisk is explained in the text.

occur as a result of guessing. The diagram is not intended to be a process model; the order of the parameters in the tree is arbitrary. The tree simply describes that state of the parameters on a particular trial. Level I expresses the probability of perceiving the target letter (TL) or not ($1 - TL$). Levels II and III express the probability of perceiving the target and nontarget colors (TC and NC), respectively. Level IV expresses the probability of correctly binding the target color with the target letter (the parameter α).

The tree diagram in Figure 1 is actually simpler than it might appear. The branch $1 - TL$ is identical to the branch TL except that, when observers do not perceive the target letter, they guess a letter. On half of the trials the guess will be correct, and on half it will be incorrect.[2]

Each branch of the tree represents the probability for a specific set of events. For example, the probability that the observer will perceive both of the target features and bind them correctly is given by the product

$$TL \times CT \times NC \times \alpha. \tag{2}$$

This probability will lead to a correct response, *TC.* However, there are other ways to obtain a correct response. If an observer did not perceive the target color, target letter, or nontarget letter, he or she might guess the correct response.

Note that everywhere the response "O" occurs in Figure 1 (e.g., CO and IO), there is another color (designated Ø). The reason there are two response categories is that four different colors were used in the experiment. Thus, there were always two other colors, the two colors not present in the display.

If the observer did not perceive the target letter, target color, or nontarget color (bottom pathway), there are eight equally likely responses, one of which is a correct response (CT). The probability of a correct response is the sum of all of the pathways leading to the response CT. Similarly, the probability of reporting the target letter and a color not present in the display (a color feature error) is the sum of all pathways leading to the response CO.

The critical parameter concerning illusory conjunctions is α, the probability of correctly joining features. This parameter can vary from .5, the situation in which the likelihood of binding the target features together is random or at chance, to 1.0, in which case binding is perfect (Ashby et al., 1996; Prinzmetal et al., 1995). In studies of illusory conjunctions, the focus has generally been on the response category called conjunction responses, CN, the situation in which the target letter is correctly identified but is reported in the nontarget color. However, as can be seen in Figure 1, there are many paths resulting in CN responses, and most of these paths are not the result of incorrect binding. An important aspect of our modeling approach is that estimates of the parameter α provide a better measure of feature integration than estimates based directly on the ratio of CN responses to CO responses.

Level V represents a guessing bias, *g.* This parameter comes into play when only one color is perceived. In the model shown in Figure 1, *g* represents the probability that the perceived color will be reported as the target color. Consider the path that leads to CN followed by an asterisk (CN*). The observer perceived the target letter (e.g., *X*) and the nontarget color (e.g., blue) but not the target color (e.g., red). Furthermore, suppose the observer correctly bound the nontarget color with the nontarget letter. Phenomenologically, the observer perceived a blue *O* and an *X* of unknown color. At one extreme, the observer might always guess the one color that was perceived, that is, the nontarget color. In this situation, *g* would equal 1.0, and the response would be categorized as a conjunction response. However, these would not be true illusory conjunctions, because binding was correct. At the other extreme, observers would never guess the nontarget color (g = .00); that is, they would apply an exclusionary strategy.

For each of the six response categories listed in Table 1, the probability is simply the sum of all of the paths that lead to the particular response category. For example, the predicted probability that the response will be correct is given by the following:

$$\begin{aligned} P(CT) = {} & [TL \times TC \times NC \times \alpha] + \\ & [TL \times (1 - NC) \times \alpha] + \\ & [TL \times TC \times (1 - NC) \times (1 - \alpha) \times g] + \\ & [TL \times (1 - TC) \times NC \times \alpha \times g] + \\ & [TL \times (1 - TC) \times (1 - NC)] + \\ & [(1 - TL) \times TC \times NC \times \alpha \times 0.5] + \\ & [(1 - TL) \times (1 - NC) \times \alpha \times 0.5] + \\ & [(1 - TL) \times TC \times (1 - NC) \times (1 - \alpha) \times g \times 0.5] + \\ & [(1 - TL) \times (1 - TC) \times NC \times \alpha \times g \times 0.5] + \\ & [(1 - TL) \times (1 - TC) \times (1 - NC) \times 0.5]. \end{aligned} \tag{3}$$

In a similar manner, a formula is obtained for each of the other response categories. Unlike the simple example given in Equation 1, there is no algebraic solution for the parameters (except for the parameter *TL*). Hence, we adjust the parameters (i.e., *TL, TC,* and so on) so as to maximize the fit between the predicted response frequency and the actual response frequency. In multinomial models, the appropriate measure of goodness of fit is G^2, which is defined as follows:

$$G^2 = \sum \left[(2 \times \mathrm{ObsFreq}_i) \times \ln\left(\frac{\mathrm{ObsProp}_i}{\mathrm{PredProp}_i} \right) \right]. \tag{4}$$

For each of the *i* response categories, ObsFreq is the observed frequency, ObsProp is the observed proportion, and PredProp is the predicted proportion. The smaller the value of G^2, the better the fit. There are several methods of finding the best-fitting parameters to minimize G^2.[3]

In a recent review of more than 80 applications of multinomial models, Batchelder and Riefer (1999) warned that "a key question in the development of an MPT (multinomial processing tree) model is whether the model's parameters are, in fact, valid measures of their respective cognitive capacities. In general, validity of a model's parameters is essential if one wishes to have confidence in an MPT model as a valid measurement tool" (p. 76). Thus, as with any measurement tool, it is critical to empirically evaluate the validity of the underlying model and, in particular, assess whether the parameters reflect the operation of hypothesized cognitive processes.

---

[2] This treatment of perceiving the target letter is slightly different from that of Prinzmetal et al. (1995) and Ashby et al. (1996). However, in our experiments, the parameter *TL* is always very high, and consequently this branch contributes little to the results. When the same structure is used for the *TL* and $1 - TL$ branches, the representation is much simpler.

[3] There are several easily available methods for fitting the data to a multinomial model. See Dodson, Prinzmetal, and Shimamura (1998) for a tutorial on a simple method for fitting multinomial data using the Solver function in Excel. This method has the advantage that it is available on a wide variety of platforms. Xiangen Ho has a program for PCs that has a graphic interface and is designed especially for this kind of model (Batchelder & Riefer, 1999). The program is available at http://irwin.psyc.memphis.edu/gpt/.

The multinomial models that we have used in our studies of illusory conjunctions (Ashby et al., 1996; Prinzmetal et al., 1995; see also Thompson, Hall, & Pressing, 2001) encompass aspects of a threshold theory. In Figure 1, the parameters represent binary states: The target letter is either seen or not seen. Such theories have been criticized as providing parameter estimates that are incorrect or misleading (e.g., Kinchla, 1994). This problem is particularly critical in the case of guessing parameters. Although this criticism has been directed at studies using present–absent detection tasks and source monitoring experiments, it is also likely to apply to illusory conjunctions experiments. Thus, the primary motivation for the current study was to test the validity of our multinomial approach for measuring feature integration. As shown subsequently, in most cases the model performed well. However, there were some exceptions, and these led to interesting revisions in the model.

Our approach was similar to that used to test measurement theories of recognition memory (Snodgrass & Corwin, 1988) and source memory (Bayen & Erdfelder, 1996). We experimentally manipulated variables to determine whether the effects would be restricted to specific parameters in the model identified a priori. In Experiment 1, we manipulated the display configuration in a way that we predicted would uniquely affect the binding probability, α. We also included an independent variable that was expected to affect the guessing parameter, g. In Experiment 2, we focused on independent variables that were expected to affect the probability of perceiving the target and nontarget colors. We took a different tack in Experiment 3. Here we obtained confidence judgments, with the expectation that high values for parameters reflecting perception of the features would be associated with high levels of confidence and high values associated with guessing would be associated with low levels of confidence.

Experiment 1

A primary goal in Experiment 1 was to demonstrate that the probability of correctly conjoining color and shape features could be manipulated independently of the probabilities associated with perceiving these features. Our previous work had indicated that this situation is not easy to obtain. Ashby et al. (1996) had shown that the binding parameter α varies with the distance between the target and nontarget letters, with lower values occurring when the items are close together. However, the distance manipulation also affected the probability of correctly identifying the target letter. We attributed this effect to lateral masking, assuming that masking is reduced when the interitem distance is large.

In the current experiment, we again varied distance, but, we also manipulated the configuration of the displays (Figure 2). We expected that one display would lead to many illusory conjunctions, whereas the other display would lead to few illusory conjunctions. Each display contained one colored target letter (T or X) and one colored nontarget letter. Within each display configuration, the two items were either close together (*near;* Figures 2A and 2B) or far apart (*far;* Figures 2C and 2D).

We expected that there would be more feature integration errors in the near condition for three reasons. First, feature integration is

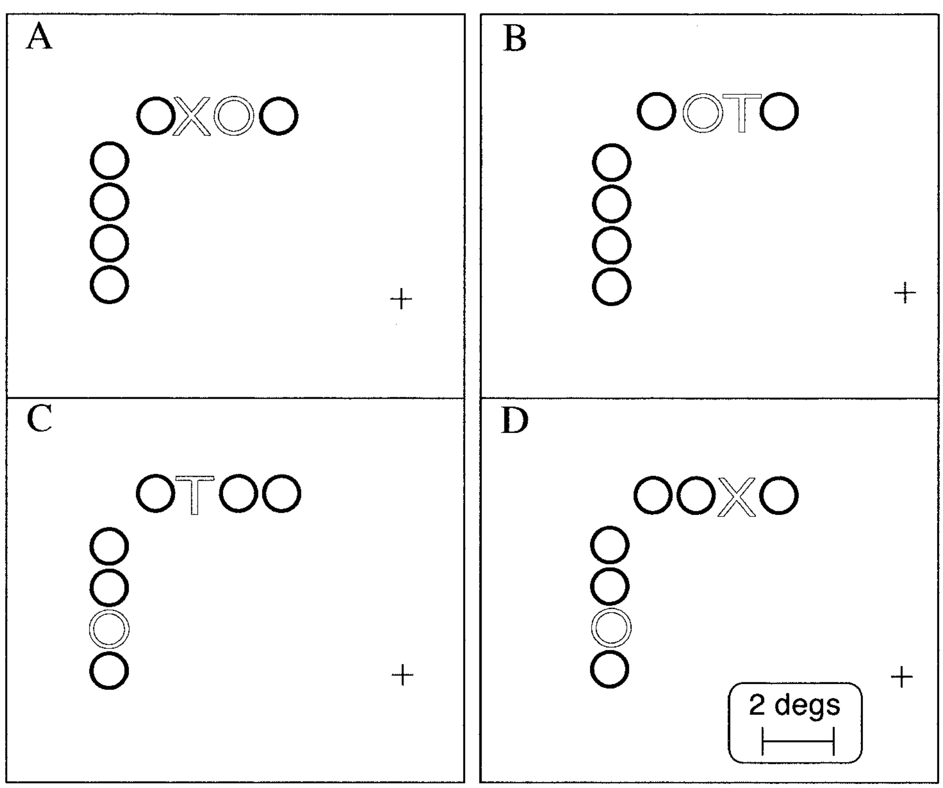

Figure 2. Sample stimuli from Experiment 1. A and B: Examples of the near condition; C and D: Examples of the far condition. Plus signs mark the center of the monitor. degs = degrees.

worse when items are close together than far apart (e.g., Ashby et al., 1996). Second, in the near condition the target and nontarget letters were part of the same perceptual group, whereas in the far condition they were part of different perceptual groups. Perceptual organization has been shown to be a powerful determinant of feature integration (e.g., Prinzmetal, 1981, 1995; Prinzmetal & Keysar, 1989). Third, illusory conjunctions are more likely between features that are vertically or horizontally aligned (Lasaga & Hecht, 1991). We hoped to keep other errors constant because the eccentricity from fixation and amount of masking were approximately the same in the two conditions. We use the terms *near* and *far* as a short notation for all of the preceding stimulus variables.

The second independent variable was designed to evaluate the guessing parameter, g. This parameter comes into play when the observer has perceived the color of only one of the items. As noted earlier, g can vary from 0 to 1.0. A probability of 0 would correspond to the situation in which the observer uses the partial information to constrain the guessing set. In particular, the color of the perceived item is never reported for the color of the item for which color information is absent. At the other extreme, a probability of 1.0 would correspond to the situation in which the perceived color is always reported for both items. It is important to recognize that a high value of g will lead to many conjunction responses, even if binding is perfect (Donk, 1999). Whenever the participant only perceives the color of the nontarget letter, a conjunction response will occur if the target letter is perceived or correctly guessed.

We would expect that g would be influenced by the probability that colors are repeated in a trial. In most illusory conjunction studies, the target and nontarget colors are never the same (i.e., the colors are selected without replacement). Assuming that the observers are informed of this constraint or become sensitive to it, g should be zero. Observers should exclude the perceived color from the guessing set, and they should not guess that two colors are the same. However, if the target and nontarget colors are the same on some of the trials (i.e., colors selected with replacement), then observers may include the perceived color in the guessing set. If the guessing probabilities were matched to the actual probabilities, then g should equal the actual probabilities. For example, if there are four possible colors, g should equal .25. We examined this issue in a between-groups manipulation. For the nonrepeat group, the target and nontarget colors were never the same. For the repeat group, the target and nontarget colors were the same on one fourth of the trials.

In summary, this experiment involved a 2 $\times$ 2 design manipulating display configuration and repetition. The display configuration manipulation was expected to affect the probability of correctly joining features, α. Given that we did not expect this manipulation to have any effect on feature perception, we did not expect the parameter g to differ for the near and far conditions. In contrast, whether the colors can repeat or not should affect g but not α. Manipulating these variables in the same experiment had one added advantage. The ideal way to measure a guessing bias, g, would be in a situation in which observers did not make illusory conjunctions. We expected this situation to hold in the far condition.

One problem with the model described for Figure 1 is that the parameters g and α are not mutually constrained: The same fit can be obtained for different pairs of these values. To overcome this problem, we used a more complex experiment and model. In addition to reporting the identity and color of the target letter, observers were required to report the color of the nontarget letter. The model for this more complex experiment is illustrated in Figure 3. Only the TL branch is illustrated; the $1 - TL$ branch is identical except that the observer must guess the target letter and will be correct on only half of the trials.

With this expanded response information, we can assess the guessing parameter g for the nontarget color as well as for the target color. In doing so, we have a slightly more complicated definition of g. Rather than representing the probability of guessing that the only color perceived belongs to the target, g now represents the probability of guessing that the color of both letters is the same. If one color is perceived, it is assigned to both letters. If no colors are perceived, the probability is that the same color is guessed for both letters. When observers are required to report the nontarget color, g becomes more constrained. In comparing Figures 1 and 3, one can observe that g plays a role in many more response categories.

Requiring the observers to report the nontarget color confers a couple of additional advantages. First, there are 20 response categories instead of 6. For each of the 6 response categories shown in Table 1, there are 3 additional categories, designated by a third letter (see Table 1). The third letter represents the observer's performance in judging the color of the nontarget letter. The letter *N* indicates that the report of the *nontarget* color was correct. The letter *T* indicates that the *target* color was reported as the nontarget color. The letter *O* indicates that a color not in the display was reported as the nontarget color (*other*). Because there were four possible display colors, whenever the two reported colors were not in the display, they could either be the same color (indicated as COO) or different colors (indicated as COØ). Note that the number of parameters remains fixed at five despite the increase to 20 response categories. Thus, the additional report requirement makes the model more highly constrained.

A second advantage of the expanded report requirement is related to the parameter NC. Ashby et al. (1996) found that NC was well constrained only when information about the nontarget color was included in the observer's report. In many situations, an accurate estimate of the probability of perceiving the nontarget color is not essential. However, given that our focus here was on testing the model, it would seem important to have as accurate an estimate as possible of all of the parameters. Including a report of the nontarget color has been shown to have a minimal impact on the overall pattern of results, despite what might seem to be a greater cognitive load (Ashby et al., 1996).

Method

Procedure. Each trial began with a white rectangle centered in the middle of the screen. The rectangle appeared before and after the stimulus and thus served as an energy mask. The exposure duration of the letter display was adjusted between blocks to maintain approximately 75% correct responding. On days in which data were collected (described subsequently), the exposure duration averaged 96 ms (5.78 computer refreshes at 60 Hz), with a standard deviation of 20.79 ms. Following the stimulus, the white masking rectangle reappeared.

In the center of the poststimulus rectangle was a 3 $\times$ 4 matrix of colored square buttons (see Figure 4). The colors were the same within each of the three rows and differed across the four columns. The matrix served as a

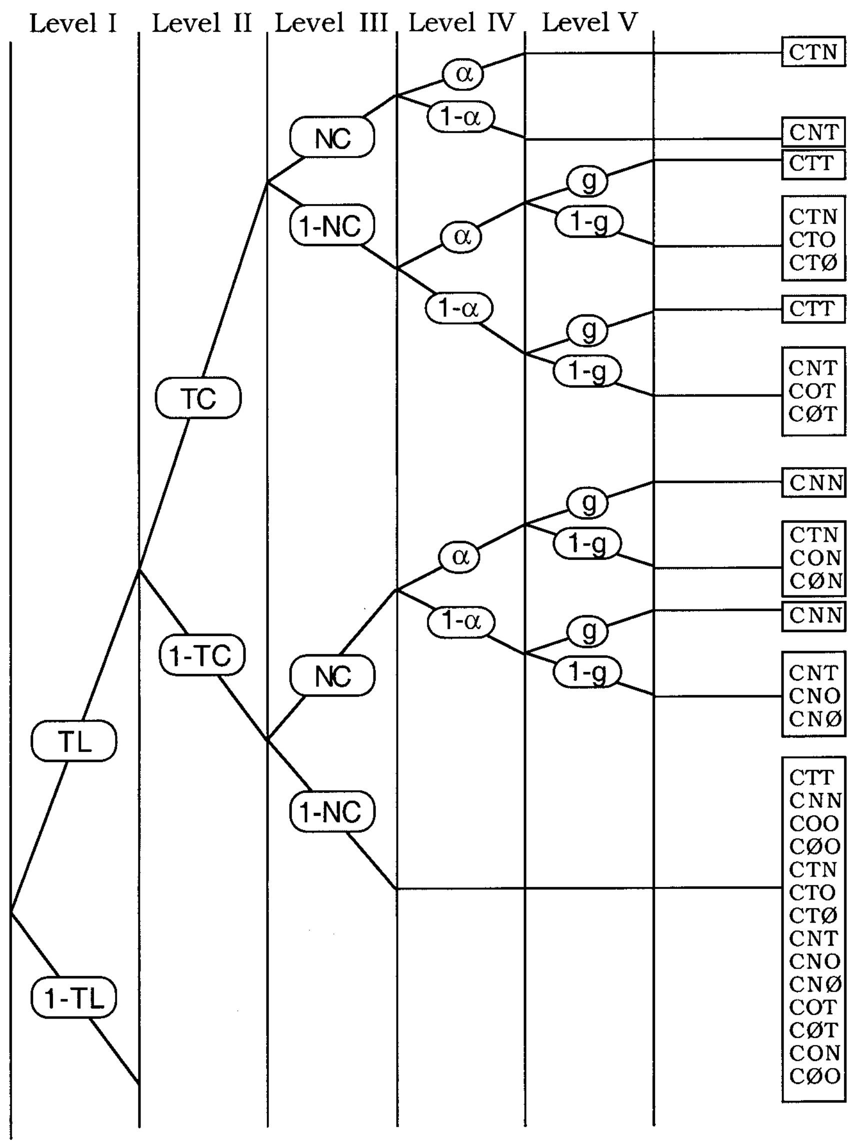

Figure 3. Model for illusory conjunctions when observers responded to the target letter, target color, and nontarget color. Only trials in which the observer correctly identified the target letter are included. *TL* = probability of perceiving the target letter; *TC* = probability of perceiving the target color; *NC* = probability of perceiving the nontarget color; α = probability of correctly binding colors and letters; g = a guessing parameter. The first letter of the response code indicates whether the letter was correct (C) or incorrect (I). The second letter indicates the type of color response for the target: the target color (T), the nontarget color (N), or one of two other colors that were not part of the display (O or Ø). The third letter indicates the type of color response for the nontarget: T, N, O, or Ø.

response palette. The observers were required to first indicate the target letter and color by selecting one of the eight buttons on the top two rows of the response palette. They would click on the top row for a "T" response (buttons with –) and the second row for an "X" response (buttons with ∞). The letters *T* and *X* were not used on the response buttons because it has been found that following the stimulus string with another alphanumeric string can cause errors (see Dixon, 1986). Within each row, observers were to use the button that corresponded to the target color. Next, they would indicate the color of the nontarget letter, the colored O, by selecting a button on the bottom row (buttons with *). On trials in which observers were correct on all three responses, the computer emitted a brief tone.

There were 96 trials in a block. In each block, the two configurations (near and far) and four positions (the stimulus in Figure 2 reflected about the vertical and horizontal meridians) occurred equally often. The colors were selected at random with the following constraints. For the no repeat group, the target and nontarget colors were never the same. For the repeat group, the target and nontarget were the same on 25% of the trials within a block. Observers were not informed about these constraints. Within the string, the colored *O* (nontarget item) was always in the same relative position, but there were two potential target positions (e.g., Figures 2A and 2C vs. Figures 2B and 2C). This position was randomly determined on each trial, as was the target letter (*T* or *X*). The order of trials within each

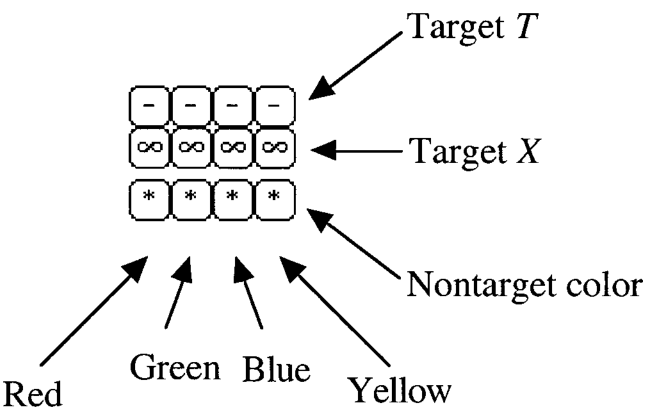

Figure 4. Response palette used in Experiments 1 and 2. The observer responded by clicking the mouse on buttons located on the screen.

block was random. After each block, observers were told their overall percentage correct. Each participant was tested in five 1-hr sessions with six blocks of trials per session. The first session was practice, and the data from this session were not included in the analysis. After the last session, observers were asked the following two questions: (a) Did you see any displays with repeating color (i.e., two colors the same) and, if so, on what percentage of trials did this occur? and (b) Did you see any displays with two target letters and, if so, on what percentage of trials did this occur?

Stimuli. The stimuli were presented on a 13-in. (33-cm) Apple monitor controlled by a Macintosh II computer. The monitor had a screen resolution of 72 pixels per inch (approximately 28 pixels per centimeter). The letters were created with a custom font such that the height and width of each letter were equal. Each four-letter string subtended approximately 5° of visual angle in length. The background of the monitor was black (7.4 cd/m^2), and the rectangular mask was white (94.0 cd/m^2). The CIE coordinates of the four colors, measured with a Minolta Chroma meter, were as follows: red, x = .46, y = .33; green, x = .28, y = .48; blue, x = .17, y = .13; and yellow, x = .40, y = .48. In the Macintosh computer code, the color values were as follows: red, r = $FF00, g = $2C00, b = $2C00; green, r = $2A00, g = $F200, b = $2A00; blue, r = $2C00, g = $2C00, b = $FC00; and yellow, r = $FF00, g = $FC00, b = $0500. The luminance values were 35.0 cd/m^2 (red), 60.0 cd/m^2 (green), 24.0 cd/m^2 (blue), and 79.0 cd/m^2 (yellow).

The viewing distance was 40 cm, and a chin rest was used to minimize head movements. The room was illuminated with overhead fluorescent lighting.

Observers. Twelve observers, 4 male and 8 female, were recruited at the University of California, Berkeley. They ranged in age from 19 to 23 years, and they had normal or corrected-to-normal vision and no known visual deficits. Observers were paid $5 per hour. Six observers were randomly assigned to the repeat condition, and the other 6 were assigned to the no repeat condition.

Results

Raw response data. The response proportions for each of the 20 response categories, averaged over observers, are shown in Table 2. Because we wanted the analysis for the repeat and no repeat groups to be identical, we do not include the data from the repeat group on trials in which the target and nontarget colors were the same.

Performance in the far condition was excellent, with the proportion of correct responses (CTN) averaging approximately 88% for both the repeat and no repeat groups. Many more errors occurred in the near condition, and this was almost entirely due to an increase in the proportion of conjunction reports (CNT). The remaining errors for both the near and far groups were due to erroneous reports of either the nontarget color (CTO) or the target letter.

The raw data argue against a recent proposal by Donk (1999) that conjunction errors are the result of confusing the nontarget letter with the target letter. Any responses with our category label CN– (e.g., CNO and CNN) would be conjunction responses. According to Donk, these responses occur because observers confuse the colored nontarget letter (the letter *O*) with one of the target letters (*T* or *X*). However, this explanation cannot account for the numerous responses in the CN– categories, especially the high number of CNT responses (see Prinzmetal, Diedrichsen, & Ivry, 2001). Our observers reported the incorrect letter on fewer than 2.5% of the trials (all of the categories in Table 2). Donk's theory predicts that conjunction reports should be just as likely when the target letter is incorrectly reported (IN–) as when the target letter is correctly reported (CN–). This prediction arises because these trials represent instances in which the nontarget *O* is mistakenly perceived as the target letter and yet observers must respond "X" or "T." If they are guessing *X* or *T* in these cases, then they should get the target letter wrong as often as they get it right. As can be

Table 2
Proportions of Each Response Category: Experiment 1

Response type	Target letter correct: Near	Target letter correct: Far	Response type	Target letter incorrect: Near	Target letter incorrect: Far
		Nonrepeat group			
CTT	.0006	.0003	ITT	.0000	.0000
CTN	.6421	.8782	ITN	.0056	.0146
CTO	.0580	.0638	ITO	.0025	.0027
CNT	.2325	.0148	INT	.0081	.0019
CNN	.0006	.0000	INN	.0001	.0003
CNO	.0233	.0039	INO	.0025	.0012
COT	.0056	.0025	IOT	.0012	.0006
CON	.0139	.0120	ION	.0007	.0023
COØ	.0020	.0009	IOØ	.0006	.0001
COO	.0001	.0000	IOO	.0000	.0000
		Repeat group			
CTT	.0199	.0110	ITT	.0014	.0014
CTN	.6578	.8818	ITN	.0104	.0139
CTO	.0272	.0313	ITO	.0010	.0012
CNT	.2149	.0262	INT	.0075	.0029
CNN	.0318	.0079	INN	.0015	.0037
CNO	.0116	.0025	INO	.0008	.0006
COT	.0027	.0014	IOT	.0002	.0000
CON	.0093	.0131	ION	.0010	.0000
COØ	.0000	.0002	IOØ	.0002	.0000
COO	.0008	.0002	IOO	.0002	.0010

Note. For the repeat group, proportions are based on 2,880 trials per observer; for the nonrepeat group, proportions are based on 2,160 trials per observer. Trials on which a color repeated are not included. The first letter of the response code indicates whether the letter was correct (C) or incorrect (I). The second letter indicates the type of color response for the target: the target color (T), the nontarget color (N), or one of two other colors that were not part of the display (O or Ø). The third letter indicates the type of color response for the nontarget: T, N, O, or Ø.

Table 3
Fits for Experiment 1

Parameters	P1	P2	P3	P4	P5	P6	*M*
				Nonrepeat group			
TL near	.951	.972	.984	.943	.910	.984	.957
TL far	.967	.983	.981	.878	.920	.986	.953
TC near	.922	.959	.950	.923	.872	.980	.934
TC far	.955	.984	.981	.953	.955	.981	.968
NC near	.917	.956	.937	.893	.650	.987	.890
NC far	.971	.946	.913	.863	.684	.963	.890
α near	.737	.689	.713	.760	.828	.650	.729
α far	.984	.978	.989	.966	.968	.988	.979
g near	.000	.008	.014	.000	.000	.080	.017
g far	.000	.010	.000	.000	.002	.014	.004
G^2	65.558	88.838	73.835	240.382	212.536	43.611	120.793
SSE	6.34E-04	1.07E-03	3.59E-04	3.81E-03	3.78E-03	1.13E-04	1.63E-03
				Repeat group			
TL near	.937	1.000	.924	.935	.958	.963	.953
TL far	.935	.988	.984	.912	.947	.940	.951
TC near	.918	.985	.909	.900	.905	.949	.928
TC far	.967	.995	.954	.939	.936	.974	.961
NC near	.939	.980	.922	.901	.888	.959	.932
NC far	.973	.989	.949	.880	.833	.983	.934
α near	.730	.822	.794	.735	.735	.680	.749
α far	.973	.992	.993	.961	.898	.984	.967
g near	.465	.334	.511	.370	.319	.492	.415
g far	.450	.165	.305	.210	.155	.402	.281
G^2	61.365	33.598	48.646	103.866	176.429	102.557	87.743
SSE	6.94E-04	4.17E-05	7.55E-04	2.12E-03	1.67E-03	7.28E-04	1.00E-03

Note. P = participant; *TL* = probability of perceiving the target letter; *TC* = probability of perceiving the target color; *NC* = probability of perceiving the nontarget color; α = probability of correctly binding colors and letters; *g* = probability of guessing two identical colors; *SSE* = sum of squared error.

seen in Table 2, there were very few responses in the categories IN– and many more in the categories CN–. Indeed, Donk's hypothesis that conjunction responses are not due to faulty binding, but rather reflect perceptual confusion, is refuted by all of the data reported in this article.

Model analysis. We start with a basic model based on Figure 3 in which there are five parameters adjusted to provide the best fit for 20 data points. We fit the model separately for the near and far conditions for each participant. Because we wanted to separately estimate the parameters for near and far conditions, we had 10 parameters (5 near and 5 far) and 40 data points. We obtained the fits using the method of gradient descent so as to minimize G^2 in the manner described by Dodson, Prinzmetal, and Shimamura (1998). To ensure that the fits did not represent local minima, we used at least 10 random starting values for each parameter. Table 3 shows the mean parameter values. In general, the fits were close to the observed data. The sum of the squared error averaged .002 for the no repeat group (range: .000 to .004) and .001 for the repeat group (range: .000 to .002).[4]

The model analyses for the nonrepeat group are moderately straightforward (see Table 3). To compare the two display configurations (near vs. far), we conducted pairwise *t* tests for each parameter.[5] The *TL* and *NC* parameters did not significantly differ across near and far conditions, $ts(5) = 0.40$ and 0.004, respectively. Averaging across the near and far conditions, the estimated probability of perceiving the target letter (*TL*) was approximately .95, whereas the probability of perceiving the nontarget color (*NC*) was approximately .89. There was a small but significant difference in the estimated probability of perceiving the target color (*TC*) as a function of near versus far. The *TC* parameter was higher in the far condition than the near condition (.97 vs. .93), $t(5) = 3.08$, $p < .01$ (two-tailed). We had not anticipated this small but reliable difference. The difference may have been due to lateral masking when the two colored items were adjacent. Alternatively, it could represent an inaccuracy within the model. Whatever the cause of the difference, as shown later, it did not affect the interpretation of the other parameters in this experiment.

[4] The values of G^2 were generally significant according to a chi-square test, suggesting that models should be rejected. However, the numbers of observations per observer for the nonrepeat and repeat groups were 2,304 and 1,728, respectively; hence, the power exceeded .9999 (Erdfelder, Faul, & Buchner, 1996). In these circumstances, it is not unexpected to obtain a significant G^2 (Batchelder & Riefer, 1994).

[5] An alternative method of determining whether parameters differ is to fit each observer's data with a reduced model. For example, a reduced model would assume that TL_{near} and TL_{far} are the same. If the reduced model is not significantly different from the full model, it is assumed that the extra parameter is not needed. In our experiments, examining individual data or testing for consistency across observers led to the same conclusions.

The critical parameters in this experiment were the feature binding parameter, α, and the guessing parameter, g. Recall that $\alpha = 1.0$ represents perfect binding, and $\alpha = .5$ is chance binding. If the model is correctly measuring feature integration errors, α should be less in the near condition than the far condition. For the far condition, α was .98, indicating nearly perfect feature binding. For the near condition, α was .73, indicating less than perfect but not random binding. The difference was reliable, $t(5) = 8.83$, $p < .01$.

Note that the differences in *TC* for the near and far conditions cannot account for the differences in α. The differences between *TC* for the near and far conditions would be expected to minimize differences in α. Failures in feature binding ($1 - \alpha$ branches in Figure 3) occur only when one or more of the color features are perceived. Thus, the differences in α between the near and far conditions may be underestimated.

In the current context, g refers to the probability of reporting the same color for both the target and nontarget letters. In the no repeat group, the two colors in the display were never the same. If observers are sensitive to this information, the probability of guessing that a nonperceived color is the same as a perceived color should be zero. In accord with this prediction, the parameter g averaged .01 and .00 for near and far conditions, respectively, $t(5) = 1.16$, *ns.*

In summary, the model for the no repeat group was very successful. As expected, the main difference between the near and far conditions was in the estimate of α, and the observers rarely reported seeing both letters in the same color, as reflected in the low estimates of g.

Turning to the repeat group, the results for the feature perception and integration parameters were quite similar to those reported for the no repeat group. The estimated probability of perceiving the target letter (*TL*) and the nontarget color (*NC*) did not significantly differ between the near and far conditions, $ts(5) = 0.15$ and 0.20, respectively. There was again a small but significant difference in the estimated probability of perceiving the target color (*TC*). *TC* was higher in the far condition than in the near condition (.96 vs. .93), $t(5) = 5.71$, $p < .01$ (two-tailed). The feature binding parameter, α, was higher for the far condition (.97) than for the near condition (.75). The difference was reliable, $t(5) = 10.17$, $p < .01$.

As predicted, the estimate of the parameter g was radically different for the repeat group than for the no repeat group. We had predicted that the value of g would be close to .25 given that there were four different colors, assuming pure guessing on trials in which only one color was perceived. The observed value for the far condition, .28, was close to this value. However, the value for the near condition was much higher than expected, .42. The difference between the two conditions was reliable, $t(5) = 4.72$, $p < .001$. We had not anticipated that this parameter would vary with display condition. A guessing parameter should reflect factors that affect guessing, not stimulus factors. The fact that g for the far condition averaged near optimal guessing (i.e., .25) was encouraging, but the fact that it varied between the near and far conditions was unexpected.

In retrospect, we are undecided about whether g should vary between display conditions. One of the authors believes that it is quite reasonable to assume that g would be higher in the near condition, whereas another believes that the fact that g varies between conditions is a severe problem that calls for an alternative model. (Reviewers were also split on this issue.) We present arguments for both positions and then suggest a pragmatic solution.

We can envision two reasons g should be higher in the near condition. First, suppose that the participant knows (explicitly or implicitly) that the visual world contains correlations between local feature values. For example, if one patch of a lawn is green, the adjacent patch is most likely green. A distant patch of lawn may be another color (e.g., yellow or brown). If one patch of sky near the setting sun is vivid red, an adjacent patch should also be vivid red, but a distant patch of sky in the east is less likely to be the same color. Hence, if one perceives that the nontarget is green, a reasonable guess for an adjacent letter would also be green.

Second, suppose that the participant perceives only one color and two letters in the near condition. The participant is confronted with a situation in which he or she saw two letters and only a single color in the immediate vicinity of both objects. The participant perceived quite a bit from this location and thus might assume that he or she perceived all of the information in the area. Hence, both letters must be the same color. In the far condition, this color is assigned to the close letter. The participant interprets the failure to perceive a second color as simply a perceptual failure; thus, he or she guesses randomly.

On the other hand, one might suppose that a guessing parameter should not be affected by display conditions. For example, in signal detection theory, one generally expects that display conditions will affect the parameter d' but not the guessing parameter, β. When one finds that display conditions affect β, this is often taken as evidence that an alternative model is required (e.g., an unequal variance model). In this spirit, we propose an alternative model for the repeat group.

Alternative model. We have investigated various alternative models and present here the one that we believe provides the most parsimonious account of the data. In our basic model (Figure 3), failures of feature binding ($1 - \alpha$) result in a transposition of the two colors. That is, the target and nontarget colors switch positions (or if only one color was perceived, it switches to the other letter). A color might also migrate to a new letter, resulting in an illusory conjunction, but still remain associated with its actual letter. This latter possibility would be akin to color spreading, a phenomenon that has been described in the literature (Prinzmetal, 1981; see also Prinzmetal & Keysar, 1989; Prinzmetal & Millis-Wright, 1984).

This idea was investigated by Treisman and Schmidt (1982) in their seminal article. They concluded that features only switch positions; they found no evidence of spreading. However, Prinzmetal (1981) asserted that features occasionally spread over several display items. In piloting the no repeat condition, we occasionally perceived two letters of the same color, even though we knew this was not possible. Perhaps errors in feature integration reflect a composite of trials in which feature switches occur as well as trials in which one color spreads over both letters. Spreading would also be expected to occur more often in the near condition. Because there is no parameter to measure this in the models in Figures 1 and 3, the excess responses of two identical letters can only inflate the parameter g. This post hoc hypothesis provides an account of the inflated values for g in the near condition.

Given this, we might assume that the estimate of g in the far condition represents the actual propensity to guess two identical colors. The mean value of .28 for g is close to the actual color probabilities (.25 of the trials had a repeated color). In contrast, g in the near condition is a mixture of guessing and trials on which a color spreads over both target and nontarget letters. Thus, the difference between g in the near and far conditions represents this spreading parameter, which we call s. This interpretation suggests that colors spread over letters with a probability of .14 (g near minus g far).

We tried several models that included s. However, it soon became apparent that, with our task, g (guessing that two colors were identical) and s (a color spreading over two letters) cannot be separately estimated; models with both s and g are mathematically equivalent. Hence, we cannot test between the first interpretation (g should vary with display distance) and the alternative model.

Until this issue is resolved, the safe solution is to use nonrepeat colors. Under these conditions, observers do not respond with repeated colors, and their guessing behavior is appropriate in that g is near zero.

Discussion

The modeling results from Experiment 1 are, for the most part, in accord with the predictions we derived on the basis of an analysis of the psychological status of the various parameters. The first goal was to compare situations in which observers would make many feature integration errors with a situation in which there would be few feature integration errors. In the near condition, binding errors, as indexed by the parameter α, were fairly common. In the far condition, feature binding was nearly perfect, as indicated by α values close to 1.0 for all observers. Note that although we use the terms *near* and *far*, there are differences between the two conditions in addition to simply the distance between the items. Our focus here is not on why binding failures do not occur under some conditions (e.g., distance or grouping) but, rather, on how the parameters would be influenced when the probability of correct binding varies. In terms of the other parameters, we found that the probabilities of perceiving the target letter and nontarget color were nearly identical in the near and far conditions. However, contrary to our expectations, observers appear to be slightly more accurate at identifying the target color in the far condition than in the near condition. This difference could only serve to minimize the differences in the binding parameter, α.

The second goal of the experiment was to determine whether observers would be sensitive to the probability of repeated colors in the display and whether this would influence their guessing behavior. For one group of observers, the displays never contained two identical colors (no repeat group); for the other group, a single color was used for both the target and nontarget letters on one fourth of the trials (repeat group). As expected, observers in the no repeat condition rarely reported the same color for both the target and the nontarget letter, and the estimated probability of guessing the same color, g, was near zero. Ashby et al. (1996) had assumed that observers use an exclusionary strategy when feature information is incomplete (e.g., only one color is perceived). For example, having perceived red for the nontarget item, observers would not guess red for the target item. The results for the no repeat group support this assumption.

In the repeat group, observers were sensitive to the probability of repeated colors. The average probabilities of guessing a repeated color, g, averaged .42 for the near condition and .28 for the far condition. Thus, as reflected in the model fits, observers were sensitive to the fact that colors repeated in the display, and the model fits reflected this. It may be that g should vary with distance. When one does not know the color of an object, a reasonable guess is the color of an adjacent object. However, we also proposed an alternative model in which colors will sometimes spread from one letter to the other without g changing.

At this point, the data do not allow us to adjudicate between the original model and the color-spreading model. Because of this uncertainty, only the no repeat condition was used in the subsequent experiments. We can safely assume that, under such conditions, observers will use an exclusionary strategy when they perceive the color of only one of the display items.

Although the model is clear in indicating that the observers in the no repeat group do not guess a color twice (i.e., $g \sim 0$), the model is mute on why observers constrain their responses in this manner. There are at least two possibilities. First, it might be that observers consciously inhibit responding with the same color twice. According to this explanation, they might perceive two red letters on a trial but refrain from reporting this because they believe it would not correspond to the actual display. Second, observers might unconsciously learn stimulus constraints and never (or rarely) perceive two identical colors. The fact that the parameter g is formally a guessing parameter does not mean that it is not perceptual in the sense of reflecting what observers consciously perceive.

These two possibilities are not mutually exclusive. We have some evidence that g can reflect phenomenal experience. Observers were asked after the last session whether they perceived displays in which the same color was repeated and, if so, to estimate the probability of this occurrence. As expected, all of the observers in the repeat condition reported that colors repeated, and their estimates of how often this occurred varied from 20% to 45% of the trials. Only one observer in the no repeat group indicated seeing a color repeated, and she reported that this occurred on 10% of the trials (Participant 2). If these retrospective reports reflect what observers perceive, it implies that observers quickly learn constraints about the displays, and these constraints can affect perception.

Experiment 2

In Experiment 1, the independent variables were predicted to influence the parameters corresponding to feature binding and guessing. In Experiment 2, we turned our attention to the parameters associated with perception of the display colors, *TC* and *NC*. To investigate these parameters, we varied the saturation of the colors using two levels (high and low) that varied independently for the target and nontarget letters. Thus, there were four saturation conditions: target and nontarget saturated, target saturated and nontarget unsaturated, target unsaturated and nontarget saturated, and target and nontarget unsaturated. We expected that the variation in saturation would primarily influence the probability that the target and nontarget colors were perceived. It was unclear whether saturation levels would influence feature binding. The model sketched in Figure 3 does not assume any effect. However, our informal observation in preparing illusory conjunction experiments has been that conjunction responses are more likely to occur with unsaturated colors (e.g., Prinzmetal & Millis-Wright, 1984). We included only the no repeat condition of Experiment 1; thus, we expected the guessing parameter to be zero.

Method

Procedure. The procedure for each trial was essentially identical to that of Experiment 1. The only difference was that there were three color choices (red, green, and blue). We did not include yellow, because we were unable to generate a satisfactory version of an unsaturated yellow. Exposure duration was adjusted as in Experiment 1. During the test blocks, the exposure duration averaged 67 ms (four computer refreshes at 60 Hz).

There were 96 trials in a block. These trials were evenly divided between the near and far conditions (see Figure 2) and among the four saturation conditions. Two different colors were randomly chosen on each trial, as well as the target letter (*X* or *T*). Each observer was tested in 1-hr sessions of six blocks each for 9 days, with the first session used as practice. Each observer completed 4,608 trials.

Stimuli. The same stimulus configurations as in Experiment 1 were used. There were three possible target and nontarget colors. The saturated colors were the same as the red, green, and blue in Experiment 1. The unsaturated colors were about half as saturated as the saturated colors. In the Macintosh computer code, the color values were as follows for the three unsaturated colors: red, r = $FF00, g = $9580, b = $9580; green, r = $7F9F, g = $F200, b = $7F9F; and blue, r = $9400, g = $9400, b = $FC00. The luminance values of the unsaturated colors were 60.0 cd/m^2 (red), 70.0 cd/m^2 (green), and 53.0 cd/m^2 (blue). The luminance values of the unsaturated colors were slightly lower than those of the saturated colors. However, as indicated later, this did not affect the probability of perceiving the target letter (*TL*).

Observers. Six observers, selected from the same population as in Experiment 1, were recruited at the University of California, Berkeley. Observers were paid $5 per hour.

Results

Raw response data. The proportions of responses for the 18 response categories are shown in Table 4, listed separately for the four saturation conditions and two configuration (near–far) conditions. Note that there were only 18 response categories per condition, because there were only three possible colors. For all conditions, observers reported all three features and showed correct binding of the target features (CTN) on at least a majority of the trials. Such reports were greatest when both target and nontarget were saturated and lowest when they were both unsaturated. For the near condition, the second highest category of responses was the conjunction response category, CNT, which varied from .105 to .242. In the far condition, the average CNT response varied only from .021 to .047. Thus, as in Experiment 1, the raw data indicate that many more conjunction responses occur in the near condition.

Model analysis. We fit the data for each observer as before, using at least 10 starting values to avoid local minima. Given the results of Experiment 1, we fixed the value of g at zero, reducing the number of free parameters to four. Separate fits were obtained for each of the four saturation conditions under both near and far conditions. Table 5 presents the mean value of each of these parameters, along with goodness of fit measures. Because of the large number of conditions (eight fits per observer), we do not show the results for individual observers. However, to evaluate the patterns in the parameter estimates, we conducted a repeated measures analysis of variance (ANOVA) on each of the parameters. Variables in this analysis were distance, target saturation, and nontarget saturation.

The parameter *TL* did not significantly vary as a function of any of the independent variables. The effects of distance, target saturation, and nontarget saturation were $F(1, 5) = 4.59$, $F(1, 5) = 3.78$, and $F(1, 5) = 4.32$, respectively (all ps $> .05$). None of the interactions approached significance. This finding may indicate that the perception of the target letter is independent of the saturation level of this letter. Or it may reflect a ceiling effect in that the estimated values of *TL* were quite high.

The results for *TC* were more complex. First, as predicted, the largest effect was caused by target saturation (see Table 5). *TC* was dramatically higher when the target color was saturated than when it was not saturated (.990 vs. .766), $F(1, 5) = 73.97$, $p < .01$. This variable did not significantly interact with distance (near vs. far configurations). There were a number of smaller effects and interactions with both the parameters *TC* and *NC* (described subsequently) that were unanticipated. However, these effects seemed to follow a simple pattern. First, the estimate of *TC* varied as a function of the saturation level of the nontarget color. Collapsing over the other factors, the mean value of *TC* when the nontarget was saturated was .85; when the nontarget was unsaturated, the mean rose to .90, $F(1, 5) = 67.57$, $p < .01$. This effect can be described as a trade-off: When the nontarget is saturated, it may draw attention away from the target letter. This trade-off was greater when the target was not saturated than when it was saturated, as reflected in a significant interaction between the saturation levels for the target and nontarget items, $F(1, 5) = 171.10$, $p < .01$. The trade-off was also greater in the near condition than in the far condition, resulting in a significant interaction between distance and the nontarget color, $F(1, 5) = 9.27$, $p < .05$.

The results in the analysis of *NC* as the dependent variable followed a pattern that was similar to the analysis of *TC*. The most dramatic effect on *NC* was the level of nontarget saturation. *NC* averaged .97 when the nontarget color was saturated but only .73 when the nontarget color was not saturated, $F(1, 5) = 97.36$, $p < .01$. *NC* was also larger in the near condition than in the far condition (.89 vs. .81), $F(1, 5) = 97.36$, $p < .01$, a result that was unexpected. One possibility is that, as attention shifts toward the target letter, the probability of detecting the nontarget color increases when it is close to the target. However, we did not obtain a similar result in Experiment 1.

Although not reliable, a similar trade-off was observed with *NC* as that described with *TC*. *NC* was slightly lower when the target color was saturated than when it was not saturated (.84 vs. .86). This difference was reliably greater when the nontarget color was not saturated than when it was, $F(1, 5) = 10.92$, $p < .05$. The trade-off was also larger in the near condition than in the far condition, but again the interaction was not significant.

Similar to Experiment 1, the distance manipulation had a marked effect on α, the feature binding parameter. In the far condition, feature binding was nearly perfect, .96; in the near condition, it was .77, $F(1, 5) = 20.97$, $p < .01$. More interesting, the saturation manipulation produced some interesting effects on feature binding (Figure 5). As noted earlier, we have observed informally that illusory conjunctions are more likely with unsaturated colors (e.g., Prinzmetal & Millis-Wright, 1984). However, this observation has remained part of “lab lore” and not been subject to experimental investigation. The present results make it

Table 4
Results for Experiment 2 Broken Down Into Response Type and Color Condition

Response type	Target saturated, nontarget saturated	Target unsaturated, nontarget saturated	Target saturated, nontarget unsaturated	Target unsaturated, nontarget unsaturated
		Near condition		
CNO	.002	.065	.001	.051
CNN	.001	.001	.000	.000
CNT	.174	.229	.105	.242
CON	.003	.088	.002	.039
COO	.000	.000	.000	.000
COT	.000	.001	.011	.019
CTO	.006	.008	.114	.063
CTN	.801	.584	.759	.567
CTT	.000	.000	.000	.000
INO	.000	.001	.000	.003
INN	.000	.000	.000	.000
INT	.004	.009	.004	.005
ION	.000	.003	.000	.001
IOO	.000	.000	.000	.000
IOT	.000	.000	.000	.001
ITO	.000	.000	.001	.001
ITN	.008	.010	.002	.008
ITT	.000	.000	.000	.000
		Far condition		
CNO	.000	.010	.001	.023
CNN	.001	.001	.001	.001
CNT	.021	.027	.030	.047
CON	.003	.109	.002	.065
COO	.000	.000	.001	.000
COT	.001	.007	.011	.023
CTO	.014	.017	.167	.121
CTN	.951	.806	.780	.707
CTT	.000	.001	.000	.000
INO	.000	.002	.000	.002
INN	.000	.000	.000	.000
INT	.003	.004	.001	.002
ION	.000	.005	.001	.002
IOO	.000	.000	.000	.000
IOT	.000	.000	.000	.000
ITO	.000	.001	.001	.001
ITN	.006	.011	.004	.007
ITT	.000	.000	.000	.000

Note. The first letter of the response code indicates whether the letter was correct (C) or incorrect (I). The second letter indicates the type of color response for the target: the target color (T), the nontarget color (N), or one of two other colors that were not part of the display (O or Ø). The third letter indicates the type of color response for the nontarget: T, N, O, or Ø.

clear that feature binding errors were more common when the target was not saturated (see Figure 5). Values for α were .91 and .82 when the target was saturated and unsaturated, respectively, $F(1, 5) = 38.61$, $p < .01$. Moreover, there was a small but significant interaction between the target–nontarget and saturation variables for α, $F(1, 5) = 13.02$, $p < .05$. When the target was not saturated, feature integration was more accurate (i.e., higher α) when the nontarget was saturated than when it was unsaturated. However, when the target was saturated, the opposite was true: Correct feature integration was greater when the nontarget was unsaturated than when it was saturated.

Discussion

The purpose of Experiment 2 was to test the hypothesis that estimates of the parameters *TC* and *NC* will be sensitive to the salience of the target and nontarget colors. We varied salience by varying target and nontarget saturation. As expected, these parameter estimates were strongly influenced by the saturation manipulation. In addition, the near–far manipulation in Experiment 2 provides further evidence that the feature binding parameter, α, is subject to distance constraints.

The modeling work revealed a number of unexpected results. These unexpected findings were generally small in magnitude. Nonetheless, as shown by the various interactions, the effects were consistent. There are two general interpretations that one might give to these effects. First, they might simply reflect failures of the model. In fact, in the General Discussion section, we suggest how the model could be changed to account for some of these findings. Second, they might reflect true aspects of feature integration and the task demands of the experiment. We propose that some of the

Table 5
Parameters and Goodness of Fit for Experiment 2

Condition	*TL*	*TC*	*NC*	α	G^2	*SSE*
Target near	.973	.988	.986	.819	27.083	1.69E-04
Nontarget far	.981	.990	.969	.975		
Target near	.954	.685	.979	.695	52.619	2.95E-03
Nontarget far	.954	.742	.945	.958		
Target near	.986	.992	.747	.879	31.055	4.12E-04
Nontarget far	.987	.988	.637	.959		
Target near	.961	.817	.832	.686	87.745	2.64E-03
Nontarget far	.973	.819	.702	.929		

Note. For the target and nontarget parameters, italics indicate saturated colors, and plain text indicates unsaturated colors. *TL* = probability of perceiving the target letter; *TC* = probability of perceiving the target color; *NC* = probability of perceiving the nontarget color; α = probability of correctly binding colors and letters; *SSE* = sum of squared error.

observed interactions may reveal important insights into the binding process.

Consider the finding that the parameter *TC* varied as a function of nontarget color saturation, and, similarly, the parameter *NC* varied as a function of target color saturation. *TC* was higher when the nontarget color was saturated than when it was not saturated. Moreover, this effect was greater when the target color was not saturated. This pattern suggests that observers' attention may be drawn to the more saturated color. Thus, when the target color is not saturated and the nontarget color is saturated, observers tend to attend to the nontarget letter, and *NC* is boosted. Similarly, when the nontarget color is not saturated and the target color is saturated, observers tend to attend to the target letter, boosting *TC*. This explanation accounts for many of the findings, but it is admittedly post hoc.

In keeping with the model depicted in Figure 3, we had also predicted that the binding parameter, α, would be affected only by the distance manipulations. However, the finding that α was less for unsaturated colors is consistent with the findings of Ivry and Prinzmetal (1991). In that study, illusory conjunction responses were greater when the target and nontarget colors were similar in hue or shape. In color space, unsaturated colors are closer together (i.e., more similar) than saturated colors. Hence, the effect of saturation on α could be considered an effect of similarity. Note that Ivry and Prinzmetal found that feature similarity could have an effect on feature integration independent of its effect on feature identification, which was not the case in the present experiment. For years we have deliberately chosen nonsaturated colors for our illusory conjunction experiments because we have continually failed to obtain a large number of conjunction errors when using highly saturated colors (e.g., Prinzmetal & Millis-Wright, 1984). Experiment 2 provided verification of this observation.

In summary, as predicted, varying the saturation of colors had its primary effect on the estimates of the two parameters associated with perceiving the target and nontarget colors. The modeling work, though, did reveal two types of interactions that require further study. The first is the interaction involving the target and nontarget saturations for the parameter reflecting the other color; varying the salience of the color of one object appears to have an effect on the likelihood that another object's color will be perceived. The second is the interaction of salience with binding itself: Unsaturated target colors lead to more binding failures.

Experiment 3

Experiment 3 had two goals. First, we tested the adequacy of the multinomial approach for studying feature integration in a novel manner. In the preceding experiments, we varied stimulus factors and tested predictions based on how these factors would influence specific parameters. In Experiment 3, we did not vary the stimulus. Instead, on each trial the observers gave a confidence rating on a scale ranging from 1 (*guessing*) to 9 (*very confident*). We assumed that the confidence ratings would be based on the observers' assessment of how well they had perceived the stimulus features. Correspondingly, we predicted that the parameter estimates should correlate with the confidence ratings. *TL, TC, NC,* and α are likely to be higher on trials in which observers are confident than on trials in which their confidence is low. A priori, we did not know whether some of the parameters would be more sensitive to the confidence ratings than others.

The second goal of this experiment was to determine the phenomenal reality of the "illusion" in illusory conjunctions. At one extreme, illusory conjunctions may be just as phenomenally real as correct perceptions. At the other extreme, illusory conjunctions may never seem like real perceptions but always result from guesses. An intermediate position is that the trials on which conjunction responses are made (CN; see Table 1) will have, on average, lower confidence rating than correct responses (CT), but there will be considerable overlap in the distributions so that some conjunction responses will have higher confidence ratings than correct responses. Treisman and Schmidt took a middle position: "At least some conjunction errors are consciously and confidently experienced as perceived physical objects rather than reflecting simply guessing in the absence of information" (1982, p. 138). In one experiment, Treisman and Schmidt took confidence ratings, but unfortunately they had only two levels of confidence (*sure* and *think*), so it is difficult to evaluate the relationship between the different types of responses and the distribution of the confidence ratings.

In Experiment 3, we adopted a procedure that was expected to lead to a large proportion of conjunction errors. The displays were similar to the no repeat–near condition of Experiments 1 and 2. The observers were required only to report the target color and target letter. Subsequent to this, they were required to give a rating

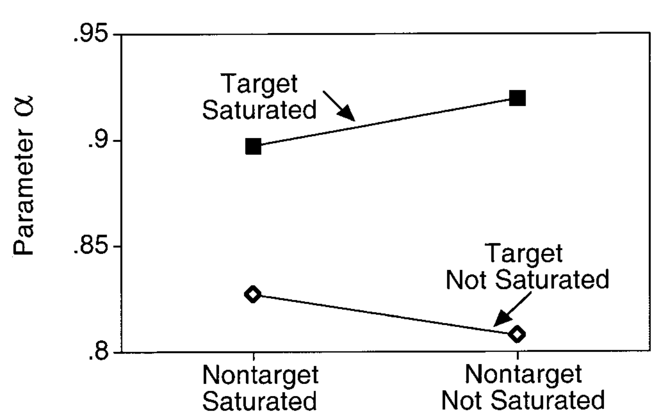

Figure 5. Experiment 2: Feature integration parameter, α, as a function of target and nontarget saturation.

(1 to 9) to indicate their confidence for that trial. Because observers had to make confidence ratings, we eliminated the nontarget color responses to keep the load in this experiment similar to the other experiments.

Method

Procedure. On each trial, observers were briefly presented a stimulus that contained a colored target letter (*X* or *T*) and a colored nontarget letter (*O*). The observer responded with the color and identity of the target letter by clicking the appropriate button on a 3 × 2 response palette. If the target letter was perceived to be *T*, the click was directed to the appropriate box on the top row; if the target letter was perceived to be *X*, responses were made on the bottom row. Just below the palette was a row of nine buttons labeled 1 to 9 from left to right. After indicating the color and identity of the target letter, observers indicated their confidence by clicking on one of the nine buttons (1 = *least confident*, 9 = *most confident*). Thus, as in Experiments 1 and 2, observers made two clicks on each trial. The first click indicated the target letter and the target color. The second click indicated the participant's overall confidence for both components of the response (target identity and color). All other aspects of the procedure were as in the previous experiments. Over all test sessions, the exposure duration averaged 107 ms.

Each observer was tested in five 1-hr sessions, with the first session used for practice. Test sessions began with a minimum of 16 warm-up trials, followed by six blocks of 96 trials each, yielding a total of 2,880 observations per observer.

Stimuli. The stimulus in Experiment 3 consisted of a single row of four letters. The dimensions were identical to the horizontal row of letters shown in Figures 2A and 2B. The two colored letters were always adjacent to each other. Whether the target was on the left or right of the colored *O* was randomly determined on each trial. In each block of trials, the row of letters appeared equally often in each of the four quadrants. The two target letters (*T* and *X*) occurred equally often in each of three colors (red, green, and blue). The target and nontarget colors were never the same.

Observers. Seven observers, selected as before, participated in the experiment.

Results

Raw response data. Mean numbers of responses for each of the six response categories and nine confidence ratings, averaged over observers, are presented in Table 6. The most frequent response category corresponded to correct responses (CT), and the highest error category involved conjunction responses (CN). As in the other two experiments, conjunction responses were much more likely when the target letter was identified (CN) than when the wrong letter was reported (IN). This finding challenges the model proposed by Donk (1999) and indicates that at least some of the errors were due to feature migration (see Prinzmetal et al., 2001).

The raw data give a clear picture of observers' confidence in making a conjunction response (CN). Figure 6 plots the proportions of conjunction responses (CN), correct responses (CT), and all other categories of responses that fall into each confidence bin. The sum of each of the curves is 1.0. The highest proportions of correct responses were given the highest confidence level. When observers made conjunction responses, they were, on average, less confident than when they were correct. Nevertheless, confidence for correct responses and conjunction responses formed overlapping distributions, so there were a substantial number of trials on which correct responses had lower ratings than conjunction responses. As shown in Figure 6, confidence was much lower on trials in which errors other than conjunction responses occurred. One problem with an analysis based on the raw data is that we cannot separate correct responses that resulted from veridical perception and those that resulted from guesses that were lucky.

Model analysis. We fit the data with a model identical to that shown in Figure 1, with two exceptions. First, because we did not repeat colors, we fixed g at zero, eliminating one free parameter. Second, because only three colors were used in the experiment, the response categories CØ and IØ were not included. Even though there were 2,880 responses per observer, there were many empty cells when the data were broken down into the nine confidence ratings. Thus, we created four data sets for each observer by combining across the lowest three ratings (C1–C3) and then grouping the pairs at higher confidence levels (C4–C5, C6–C7, and C8–C9). As a result, for each of the 7 observers, we solved for *TL*, *TC*, *NC*, and α at four confidence levels. The fits, averaged over the 7 observers, are presented in Table 7. We conducted a one-way repeated ANOVA on each of the parameters as a function of confidence.

Table 6
Number of Responses as a Function of Response Type and Confidence Ratings in Experiment 3, Averaged Over Observers

	Confidence rating									
Response type	1	2	3	4	5	6	7	8	9	Total
CN	67.7	65.7	95.3	89.9	83.7	64.7	48.9	26.9	13.6	556.3
CO	8.1	3.4	4.1	2.4	0.9	1.0	0.6	0.3	0.4	21.3
CT	78.9	104.3	172.6	223.9	269.1	306.4	364.7	305.9	367.4	2193.2
IN	27.4	13.1	12.6	7.7	6.4	2.7	2.7	1.0	0.3	74.0
IO	2.4	0.0	0.4	0.1	0.1	0.3	0.1	0.1	0.0	3.7
IT	15.1	3.3	5.0	2.3	1.4	1.6	2.0	0.3	0.6	31.6
Total	199.7	189.9	290.0	326.3	361.7	376.7	419.0	334.4	382.3	2,880

Note. The first letter of the response code indicates whether the letter was correct (C) or incorrect (I). The second letter indicates whether the color response was the target color (T), nontarget color (N), or the other color that was not part of the display (O).

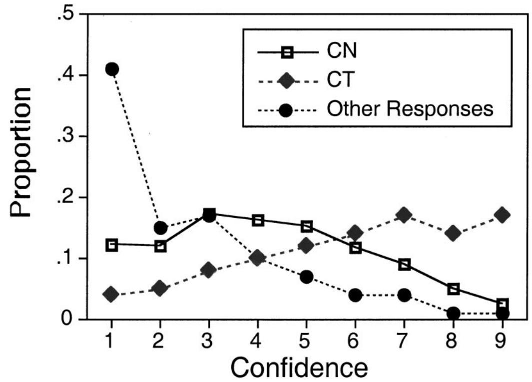

Figure 6. Experiment 3: Proportions of CN (conjunction responses), CT (correct responses), and all other responses as a function of confidence. Note that each function sums to 1.0.

The parameter *TL* significantly differed for the four different confidence categories. It averaged .787 for the lowest confidence level and .996 for the highest confidence level, $F(3, 18) = 18.42$, $p < .01$. The parameters representing the perception of the target color and nontarget color, *TC* and *NC,* also rose as confidence increased. However, estimates of *TC* and *NC* did not significantly increase with confidence, $Fs(3, 18) = 1.50$ and 1.93, respectively, probably as a result of a ceiling effect. As can be seen in Table 6, responses that were classified as color feature errors (CO or IO) occurred on less than 1% of the trials. We assume that the colors were missed on some other trials but the participant guessed a color that had been part of the display. Nonetheless, color perception was clearly quite good in this experiment.

Most interesting, the parameter α varied from .559 for the least confident category to .940 for the most confident category. The difference over the four confidence categories was reliable, $F(3, 18) = 114.11$, $p < .01$. Thus, when feature binding is near chance, observers have little confidence in their responses. When feature binding is near perfect, observers are very confident in their responses.

Discussion

In this experiment, we tested the adequacy of multinomial modeling in a unique way. Instead of changing stimulus conditions and tracking changes in parameter values, we kept the stimulus conditions constant and obtained confidence ratings. We assumed that low confidence corresponds to perceptual uncertainty. Thus, the parameter values should vary with confidence ratings. Indeed, all four parameters varied with confidence, although the statistics were reliable only for *TL* and α.

This experiment affords us the best view yet of the phenomenal reality of feature integration errors. It is not surprising that, on average, observers are more confident when they make a correct response than when they make an error. Similarly, as reflected in the estimate of α, the likelihood of making an illusory conjunction increases as confidence decreases. However, the results are in accord with the claim of Treisman and Schmidt (1982) that there are some trials in which feature integration errors appear to be as phenomenally real as correct responses. Our own experience is similar. We have served as participants in many illusory conjunction experiments and continue to be amazed that, for example, a very clear perception of a red *T* and a blue *O* is wrong, that the display actually contains a blue *T* and a red *O.* In the current experiment, 5% of the trials in the highest confidence group (C8–C9) resulted in conjunction errors.

At first glance, it seems surprising that the binding parameter is sensitive to confidence. It is intuitively reasonable to expect that the values of the parameters representing the features would be correlated with confidence. On some trials, the observers may have not gotten a good look at the briefly presented stimuli; perhaps they blinked at the wrong time or were momentarily distracted. But it is not obvious that these sorts of effects would influence binding. Binding requires that at least one shape and one color are perceived, and the α parameter describes the likelihood that the features will be bound correctly.

Why would binding become less accurate as confidence decreases (or, alternatively, why would confidence decrease as binding becomes less accurate)? To account for this, it is important to consider the underlying mechanism or mechanisms that might cause variation in α. Ashby et al. (1996) argued that true binding errors are the result of variability in the perceived location of features. If the display contains a blue *T* and red *O,* we may report a red *T* if the location of the *T* is perceived as closer to the spatial representation of the red stimulus than to the blue stimulus. Although the models described by Ashby et al. are similar to that depicted in Figure 3, α was not computed separately for each distance, as in the present experiments; rather, it was computed as a function of bivariate distributions of perceived locations. In essence, Ashby et al. proposed that feature integration errors were due to imprecise location information (e.g., Friedman-Hill et al., 1995; Logan, 1996; Prinzmetal & Keysar, 1989). When viewed in this manner, it becomes clear that, similar to the way feature perception may be more fuzzy on trials in which confidence is reported to be low, perceived location may also be more variable on such trials (see also Prinzmetal, Amiri, Allen, & Edwards, 1998; Prinzmetal, Nwachuku, Bodanski, Blumenfeld, & Shimizu, 1997; Prinzmetal & Wilson, 1997).

General Discussion

The goal of the experiments reported here was to test the adequacy of using multinomial models for investigating feature

Table 7
Average Fits for Observers in Experiment 3, With Confidence Rating as the Independent Variable

	Confidence rating			
Parameter	C1–C3	C4–C5	C6–C7	C8–C9
TL	.787	.944	.977	.996
TC	.973	.992	1.000	1.000
NC	.929	.985	.971	.976
α	.559	.729	.848	.940
G^2	29.529	23.650	15.730	10.645
SSE	4.09E-03	1.25E-03	2.01E-04	3.05E-05

Note. *TL* = probability of perceiving the target letter; *TC* = probability of perceiving the target color; *NC* = probability of perceiving the nontarget color; α = probability of correctly binding colors and letters; *SSE* = sum of squared error.

integration. We used two strategies. In Experiments 1 and 2, stimulus factors were varied to test specific predictions concerning the psychological status of the model's parameters. For example, we varied a host of factors that have been shown to affect feature integration, such as interobject distance and grouping. The feature binding parameter, α, was significantly affected by these manipulations in both Experiment 1 and Experiment 2. In Experiment 1, we also compared conditions in which the colors were selected on each trial without replacement with conditions in which the colors were selected with replacement. Thus, in the second condition, the same color could be associated with both the target and nontarget letters. As predicted, this manipulation affected the guessing parameter, g. In Experiment 2, we varied the saturation level of the colors. The parameters that reflect the probability of perceiving the target color and nontarget color, TC and NC, were higher with saturated colors than with nonsaturated colors.

In the third experiment, we took a novel approach to testing the adequacy of the multinomial approach. We obtained confidence ratings after each trial. We assumed that observers would be guessing more often on trials with low confidence than on trials with high confidence. Hence, one or more of the model parameters should be lower on trials with low confidence. All of the parameters were lower on trials that received low confidence. However, as a result of a ceiling effect on color accuracy, this effect was significant only for the parameters TL and α.

The results were generally in accord with our predictions, but there were two notable exceptions. First, in Experiment 1, the likelihood that an observer would guess that the same color was used for both the target and nontarget was higher than expected from base rate probabilities in the near condition. One hypothesis is that when items are close together, observers are more willing to guess that they are the same. Another hypothesis is that colors sometimes spread over adjacent items (or locations). Color spreading has not been incorporated into multinomial models of feature integration, and the present design precluded adding this parameter, because such a model would be mathematically equivalent to the original. For this reason, we only included the no repeat color condition in Experiments 2 and 3.

The results of Experiment 1 convincingly demonstrate that when colors do not repeat in a display, observers do not guess the same color twice. This does not mean that spreading no longer occurs. Rather, it may be that observers adopt an exclusionary strategy when they recognize (explicitly or implicitly) that the two colors are always different. Hence, the parameter g is essentially zero and can be dropped from the model. We recommend this simplification to other investigators.

The second unexpected finding was that, in Experiment 2, the color saturation manipulation not only affected the likelihood of perceiving the color features (and minimally at that), but it also had an effect on the binding parameter, α. Feature integration errors occurred at a higher rate when the target color was less saturated. Although we did not predict this result, it is consistent with the results of Ivry and Prinzmetal (1991), who found that conjunction responses were more likely when the display contained similar than dissimilar features. Unsaturated colors are closer to each other in color space than saturated colors, and hence they are more similar. We have informally observed this effect before, leading us to choose pastel colors when conducting illusory conjunction experiments. There is, of course, a danger in using colors that are too unsaturated: The parameters that reflect the probability of perceiving the color will also become low.

As noted in the introduction, the multinomial models described in this article are not process models; rather, they are intended as measurement tools. Thus, although they indicate that feature integration errors are more likely with unsaturated colors, they do not tell us why this is so. To understand why unsaturated colors might lead to more feature integration errors, consider models that bind features that are near each other in physical space (e.g., Ashby et al., 1996; Logan, 1996). It may be that features in color space work in a manner analogous to features in physical space: Features from items in space that are near each other are more likely to lead to binding errors (see Ivry & Prinzmetal, 1991).

Multinomial models have been criticized because information is represented in an all-or-none fashion. In our context, it is assumed that, on each trial, the observer either knows the target color or has no information about this color. The models do not seem to allow for partial information. This approach can be contrasted with a continuous state theory, such as signal detection theory, that allows for partial information.

There are ways in which partial information can be represented in multinomial models. Consider a source memory experiment reported by Dodson, Holland, and Shimamura (1998). A list of words was read to each participant by one of four people; two of the sources were male, and two were female. At test, participants were presented a list of words and had to determine whether the words were old or new. For items judged old, observers also had to indicate the source. The data were best fit by a model that included a parameter for remembering the gender of the source, but not which specific male or female read the item. Knowledge of gender in this context constitutes partial source information.

This approach could easily be extended to models of feature integration. Suppose the possible target colors were red, orange, blue, and cyan. An observer might have partial information: He or she might know that the color was a warm color (red or orange) even if he or she could not identify the exact color. One could include an additional parameter to indicate whether a color was perceived as a warm color or cool color and another parameter to indicate the specific color.

A comparison of how continuous state models (e.g., signal detection theory) and multinomial models represent partial information is revealing. The continuous state theory is designed to represent many states of partial information. However, it is vague in regard to the nature of that partial information. The multinomial approach does not have infinite levels of partial information, and it is more constrained because it forces the investigator to precisely characterize the information (e.g., the discrimination of warm vs. cool colors).

Another criticism of multinomial models is that the parameters are generally assumed to be independent. Thus, the probability of perceiving the target letter identity (TL) is formally independent of the probability of perceiving its color (TC). There are clearly situations in which judgments about color and shape are not independent (e.g., Bonnel & Prinzmetal, 1998). There almost surely is a nonzero correlation between, for example, the perception of the target letter and its color. For instance, on some trials, observers will blink, not be attending, or not be appropriately fixated. On these trials, observers will be forced to guess on both the target letter and target color. Similarly, in considering the interactions observed in the saturation study, we proposed that a trade-off might exist as attention is drawn to the more salient stimulus.

In modeling various data sets, we have found that one can easily include correlated parameters. Consider a situation in which one suspects that the probabilities of perceiving the target color and the target letter are highly correlated. Working from the model in Figure 1, the correlated version would begin in the same manner at Level I, with a single parameter for *TL*. At Level II, instead of one parameter, *TC*, there are two parameters: *TC* given that the target letter was perceived ($TC|TL$) and *TC* given that the target letter was not perceived ($TC|\sim TL$). In this manner, the correlation can be tested. If the parameters ($TC|TL$) and ($TC|\sim TL$) are not significantly different, then it is reasonable to simplify the situation and use only a single *TC* parameter. Indeed, in the studies presented here, it was not necessary to use one of these models.

It should be clear that we do not consider any specific model, such as the one presented in Figure 1, as the ultimate and correct model and other models as wrong. Models are representations of the world that are more or less useful, accurate, and parsimonious. Some aspects of the data were not completely captured by the specific model we tested. In arriving at the model we tested here, we tried and rejected several alternative multinomial models (e.g., Ashby et al., 1996; Prinzmetal et al., 1995). Thus, we view the present effort as one stop in a journey toward a better model, not the ultimate destination.

In conclusion, the multinomial approach provides a rigorous yet flexible tool for the study of feature integration. Although we have focused on the most common classes of these models, ones that assume no correlation between the different parameters and assume all-or-none states, it should be straightforward to extend this approach to cases with partial information and correlated parameters. In promoting the multinomial approach, we do not intend to denigrate the utility of continuous state models. Indeed, we have hypothesized a hybrid model with both multinomial and continuous state components (Ashby et al., 1996). However, it remains to be seen if continuous state models can be validated and applied as easily as the current multinomial approach. Until that time, our current multinomial approach provides a valid measurement tool for studying feature integration.

References

Arguin, M., Cavanagh, P., & Joanette, Y. (1994). Visual feature integration with an attentional deficit. *Brain and Cognition, 24,* 44–56.

Ashby, F. G., Prinzmetal, W., Ivry, R., & Maddox, T. (1996). A formal theory of illusory conjunctions. *Psychological Review, 103,* 165–192.

Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science, 11,* 267–273.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing model of source monitoring. *Psychological Review, 97,* 548–564.

Batchelder, W. H., & Riefer, D. M. (1994). Analysis of a model for source monitoring. In E. H. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 51–65). New York: Springer-Verlag.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review, 6,* 57–66.

Bayen, U., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 1–19.

Baylis, G. C., Driver, J., & McLeod, P. (1992). Movement and proximity constrain miscombinations of colour and form. *Perception, 21,* 201–218.

Bonnel, A.-M., & Prinzmetal, W. (1998). Dividing attention between the color and the shape of objects. *Perception & Psychophysics, 60,* 113–124.

Butler, B. E., Mewhort, D. J. K., & Browse, R. A. (1991). When do letter features migrate? A boundary condition for feature-integration theory. *Perception & Psychophysics, 49,* 91–99.

Chastain, G. (1982). Feature mislocalizations and misjudgments of intercharacter distance. *Psychological Research, 44,* 51–66.

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.

Dixon, P. (1986). Attention and interference in the perception of brief visual displays. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 133–148.

Dodson, C. S., Holland, P. W., & Shimamura, A. P. (1998). On the recollection of specific and partial source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1121–1136.

Dodson, C. S., Prinzmetal, W., & Shimamura, A. (1998). Using Excel to estimate parameters from observed data: An example from source memory data. *Behavior Research Methods, Instruments & Computers, 30,* 517–526.

Donk, M. (1999). Illusory conjunctions are an illusion: The effects of target–nontarget similarity on conjunction and feature errors. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1207–1233.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments & Computers, 28,* 1–11.

Friedman-Hill, S. R., Robertson, L. C., & Treisman, A. (1995, August 11). Parietal contributions to visual feature binding: Evidence from a patient with bilateral lesions. *Science, 269,* 853–855.

Gallant, J. L., & Garner, W. R. (1988). Some effects of distance and structure on conjunction errors. *Bulletin of the Psychonomic Society, 26,* 323–326.

Hazeltine, E., Prinzmetal, W., & Elliot, K. (1997). If it's not there, where is it: Locating illusory conjunctions. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 263–277.

Ivry, R. B., & Prinzmetal, W. (1991). Effect of feature similarity on illusory conjunctions. *Perception & Psychophysics, 49,* 105–116.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67–85.

Keele, S. W., Cohen, A., Ivry, R., Liotti, M., & Yee, P. (1988). Tests of a temporal theory of attentional binding. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 444–452.

Khurana, B. (1998). Visual structure and the integration of form and color. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1766–1785.

Kinchla, R. A. (1994). Comments on Batchelder and Riefer's multinomial model for source monitoring. *Psychological Review, 101,* 166–171.

Lasaga, M. I., & Hecht, H. (1991). Integration of local features as a function of global goodness. *Perception & Psychophysics, 49,* 201–211.

Logan, G. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review, 103,* 603–649.

Prinzmetal, W. (1981). Principles of feature integration in visual perception. *Perception & Psychophysics, 30,* 330–340.

Prinzmetal, W. (1990). Neon colors illuminate reading units. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 584–597.

Prinzmetal, W. (1995). Visual feature integration in a world of objects. *Current Directions in Psychological Science, 4,* 90–94.

Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). The phenomenology of attention; Part 1. Color, location, orientation, and "clarity." *Journal of Experimental Psychology: Human Perception and Performance, 24,* 261–282.

Prinzmetal, W., Diedrichsen, J., & Ivry, R. B. (2001). Illusory conjunctions are alive and well: A reply to Donk (1999). *Journal of Experimental Psychology: Human Perception and Performance, 27,* 538–541.

Prinzmetal, W., Henderson, D., & Ivry, R. (1995). Loosening the constraints on illusory conjunctions: The role of exposure duration and attention. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1362–1375.

Prinzmetal, W., Hoffman, H., & Vest, K. (1991). Automatic processes in word perception: An analysis from illusory conjunctions. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 902–923.

Prinzmetal, W., & Keysar, B. (1989). A functional theory of illusory conjunctions and neon colors. *Journal of Experimental Psychology: General, 118,* 165–190.

Prinzmetal, W., & Millis-Wright, M. (1984). Cognitive and linguistic factors affect visual feature integration. *Cognitive Psychology, 16,* 305–340.

Prinzmetal, W., Nwachuku, I., Bodanski, L., Blumenfeld, L., & Shimizu, N. (1997). The phenomenology of attention: Part 2. Brightness and contrast. *Consciousness and Cognition, 6,* 372–412.

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.

Prinzmetal, W., Treiman, R., & Rho, S. H. (1986). How to see a reading unit. *Journal of Memory and Language, 25,* 461–475.

Prinzmetal, W., & Wilson, A. (1997). The effect of attention on phenomenal length. *Perception, 26,* 193–205.

Rapp, B. C. (1992). The nature of sublexical orthographic organization: The bigram trough hypothesis examined. *Journal of Memory and Language, 31,* 33–53.

Riefer, D. M., & Batchelder, D. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95,* 318–339.

Riefer, D. M., Hu, X., & Batchelder, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 680–693.

Seidenberg, M. (1987). *Sublexical structures in visual word recognition: Access units or orthographic redundancy?* Hillsdale, NJ: Erlbaum.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Application to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34–50.

Thompson, W. F., Hall, M. D., & Pressing, J. (2001). Pitch and duration in unfamiliar tone sequences. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 128–140.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Treisman, A., & Paterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 12–31.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14,* 107–141.

Wolford, G., & Shum, K. H. (1980). Evidence for feature perturbations. *Perception & Psychophysics, 27,* 409–420.

Received May 25, 2000
Revision received June 7, 2001
Accepted June 22, 2001 ■