A version of this paper is published in the *Journal of Experimental Psychology: Human Perception and Performance*, 251-269. The final version of this manuscript will differ from this version. This version is provided for personal use only.

A Measurement Theory Of Illusory Conjunctions

William Prinzmetal

Richard B. Ivry

University of California, Berkeley

Diane Beck

University College London

Naomi Shimizu

University of Hawaii at Manoa

John A. Burns School of Medicine

Illusory conjunctions are 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, and well as the incorrect combination of correctly perceived features. Recently several investigators have suggested using multinomial models as a tool for measuring feature integration. We test the adequacy of these models in two experiments by testing whether model parameters reflect changes in stimulus factors. In a third experiment, we use confidence ratings 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 reponse 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). They 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 will 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 even under extended viewing conditions. Prinzmetal, Henderson, and Ivry (1995) controlled eye movements and presented strings of colored letters in the periphery for as long as 2 seconds. Even without a diverting attention task, they obtained as many illusory conjunctions under these conditions as with brief exposures.

A variety of stimulus factors have been found to affect feature integration A plethora of data has demonstrated that illusory conjunctions are more likely between adjacent than distant items (e.g., Ashby, Prinzmetal, Maddox, & Ivry, 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. The effect of perceptual organization surprisingly extends to syllable or syllable-like units in printed words. When presented with a string of colored letters such as 'MAYBE,' observers will more likely report that the Y was the color of the M or A (within a syllable) than report that the Y was 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).

While the empirical findings reviewed above are impressive, most of the studies have lacked a theoretically justified, empirically verified method for measuring illusory conjunctions. This is a potentially serious limitation: As we discuss below, reporting 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 lead 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 the incorrect combination of features, and guesses that appear to be illusory conjunctions. We will refer to the objective reports in which features from different objects are combined together as “conjunction responses” (e.g., blue X in the example). We will use the term illusory conjunctions to refer to the subset of these responses in which the percept was 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, Prinzmetal et al., 1995; Donk, 1999) 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 the correction for guessing, consider the simple case of a student taking a 4- alternative multiple choice test. The probability correct, P(C), is the sum of items the student knew, P(K), and correct guesses as shown below:

$$P(C) = P(K) + (1/4 * (1 - P(K)) \qquad (1)$$

P(K), the proportion of items the student knew can be calculated algebraically. The model that we will present 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 above is a binomial problem, since there are only two outcomes on each item (correct and incorrect). In the illusory conjunction experiments of Ashby et al. (1996), there are many outcomes; hence it is a multinomial problem.

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 & Riefer (1990; Riefer & Batchelder, 1988, 1994; cf. Banks, in press). 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.

Insert Table 1 about here

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 above, 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 2-letter code. The first letter indicates whether the letter identity was <u>correct</u> (C) or <u>incorrect</u> (I). The second letter indicates the type of color response. The observer could respond with the <u>target</u> color (T, red in the example), the color of the <u>nontarget</u> letter (N, blue in the example), or some <u>other</u> 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 the correct letter but a color that was not part of the display, and so on. The descriptions listed in Table 1 have been adapted from Prinzmetal et al. (1995) and Ashby et al, (1996).

Insert Figure 1 about here

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
_ the probability of correctly binding the target color to the target letter
g a guessing parameter, described below

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 occur due to 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*, *NC*), respectively. Level IV expresses the probability of correctly binding the target color with the target letter (the parameter _).

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 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 * *CT* * *NC* * _ (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, IO) there is another color (designated Ø). The reason for two response categories is because 4 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 8 equally likely responses, one of which is a correct response (CT). The probability of a correct response is the sum of all 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 _, the probability of correctly joining features. This parameter can vary from 1.0, in which case binding is perfect, to 0.5, the situation in which the likelihood of binding the target features together is random, or at chance (Prinzmetal et al., 1995; Ashby et al, 1996). 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 are not the result of incorrect binding. An important aspect of our modeling approach is that estimates of the parameter _ provides a better measure of feature integration than estimates based directly on the ratio CN response 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). Further, suppose the observer correctly binds 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 since binding was correct. At the other extreme, observers would never guess the nontarget color (g = 0.0), that is they would apply an exclusionary strategy.

For each of the 6 response categories listed in Table 1, the probability is simply the sum of all the paths that lead to the particular response category. For example, the predicted probability correct is given by the following:

p(CT) = *TL* * *TC* * *NC* * _ + (3)
TL * (1-*NC*) * _ +
TL * *TC* * (1-*NC*) * (1-__ * *g* +
TL * (1-*TC*) * *NC* * _ * *g* +
TL * (1-*TC*) * (1-*NC*) +
(1-TL) * *TC* * *NC* * _ * 0.5 +
(1-TL) * (1-*NC*) * _ * 0.5 +
(1-TL) * *TC* * (1-*NC*) * (1-__ * *g* * 0.5 +
(1-TL) * (1-*TC*) * *NC* * _ * *g* * 0.5 +
(1-TL) * (1-*TC*) * (1-*NC*) * 0.5

In a similar manner, a formula is obtained for each of the other response categories. Unlike the simple example given in formula 1, there is no algebraic solution for the parameters (except for the parameter *TL*). Hence we adjust the parameters (i.e., *TL*, *TC*, etc.) 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:

$$G^2 = \sum \left[\left(2 * ObsFreq_i\right) * \ln\left(\frac{ObsProp_i}{PredProp_i}\right)\right] \tag{4}$$

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

In a recent review of over 80 applications of multinomial models, Batchelder and Riefer (Batchelder & Riefer, 1999) warn that "... a key question in the development of an MPT (multinomial processing tree) model is whether the model's parameters are, in fact, valid measure 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.

The multinomial models that we have used in our studies of illusory conjunctions (Ashby et al., 1996; Prinzmetal et al., 1995; also see Thompson, Hall, & Pressing, in press) encompass aspects of a threshold theory. In Figure 1, the parameters represent binary states--you either see the target letter or you don't. Such theories have been criticized as giving parameter estimates that are wrong or misleading (e.g., Kinchla, 1994). This problem is particular critical with guessing parameters. While 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 will be shown below, 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 see if the effects would be restricted to specific parameters in the model that we identified a priori. In Experiment 1, we manipulated the display configuration in a way that we predicted would uniquely affect the binding probability, _. 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 1was to demonstrate that the probability of correctly conjoining the 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 _ varies with the distance between the target and noise letters, with lower values occurring when the items were 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 inter-item distance is large.

In the current experiment, we again varied distance, but in addition, 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,b) or far apart ("far", Figures 2c,d).

Insert Figure 2 about here

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

The second independent variable was designed to evaluate the guessing parameter, *g*. This parameter comes into play when the observer has only perceived the color of one of the items. As noted above, *g* can vary from 0 to 1.0. A probability of 0 would correspond to the situation where 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 where 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 reflect the probability that colors are repeated in an experiment. In most illusory conjunction studies, the target color and nontarget colors are never the same (i.e., the colors are selected without replacement). Assuming the observers are informed of this constraint or become sensitive to it, *g* should be 0. The observer 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 the observer 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-group manipulation. For the nonrepeat group, the target and nontarget color were never the same. For the repeat group, that target and nontarget color were the same on 1/4 of the trials.

In summary, this experiment is a 2 x 2 design manipulating display configuration and repetition. The display configuration manipulation is expected to affect the probability of correctly joining features, _. Given that we don't expect this manipulation to have any effect on feature perception, we do 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 _. Manipulating these variables in the same experiment has one added advantage. The ideal situation 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 _ are not mutually constrained: The same fit can be obtained for different pairs of the values *g* and _. To overcome this problem, we used a more complex experiment and model. In addition to reporting the identity and color of the target letter, the observers were also required to report the color of the nontarget color. The model for this more complex experiment is illustrated in Figures 3. Only the *TL* branch is illustrated, as the 1-*TL* branch is identical except that the observer must guess the target letter and will only be correct on 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, it is the probability that the same color is guessed for both letters. By requiring the observers 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 now 20 response categories instead of six. For each of the six response categories shown in Table 1, there are three 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 <u>nontarget</u> color was correct. The letter T indicates that the <u>target</u> color was reported as the nontarget color. The letter O indicates that a color not in the display was reported as the nontarget color (<u>other</u>). Since there were 4 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.

Insert Figure 3 about here

A second advantage of the expanded report requirement is related to the parameter *NC*. Ashby et al. (1996) found that the parameter *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 is to test the model, it would seem important to have as accurate an estimate of all of parameters as is possible. Including a report of the nontarget color has been shown to have 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 was collected (see below) the exposure duration averaged 96 milliseconds (5.78 computer refreshes at 60 hz), with a standard deviation of 20.79 milliseconds). Following the stimulus, the white making rectangle reappeared.

Insert Figure 4 about here

In the center of the post-stimulus rectangle was a 3 x 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 response palette. The observers were required to first indicate the target letter and color by selecting one of the 8 buttons on the top two rows of the response palette. They would click on the top row for a T response (buttons with the symbol '-'), and the second row for an X response ((buttons with the symbol '∞'). 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, they were to use the button that corresponded to the target color. Next, the observer would indicate the color of the nontarget letter, the colored O by selecting a button on the bottom row (button marked *). On trials in which the observer was 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 4 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. With the no repeat group, the target and nontarget colors were never the same. With 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., Figure 2a and 2c vs. Figure 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 block were random. Following each block, observers were told their overall percent correct. Each participant was tested on five 1-hour 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.

Following the last session, observers were asked the following two questions: (1) Did you see any displays with repeating color (i.e. two colors the same) and if so, on what percent of trials did this occur? (2) Did you see any displays with two target letters, and if so, on what percent 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 in (approximately 28 pixels per cm). The letters were created with a custom font such that the height and width of each letter was equal. Each 4-letter string subtended approximately 5 degrees 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 4 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); yellow (x=.40, y=.48). In the Macintosh computer code, the color values were as follows for the 4 colors: red (r = $FF00, g = $2C00, b = $2C00); green (r = $2A00, g = $F200, b = $2A00); blue (r= $2C00, g= $2C00, b= $FC00); 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), 79.0 cd/m^2 (yellow).

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

Observers. Twelve observers, 4 male and 8 female, were recruited at the University of California, Berkeley. Their ages ranged 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 six were assigned to the No Repeat condition.

Results

Raw response data. The proportion of responses for each of the 20 response categories, averaged over observers, is given in Table 2. Because we wanted the analysis for the Repeat and Norepeat groups to be identical, we do not include the data from the Repeat group on trials in which the target and nontarget colors were identical.

Insert Table 2 about here

Performance in the Far condition was excellent, with the proportion of correct responses (CTN) averaging approximately 87% for both the Repeat and Norepeat 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., CNT, CNI) would be conjunction responses. According to Donk, these occur because observers confuse the colored nontarget letter (the letter O) with one of the target letters (T or X). However, this explanation can not account for the numerous responses in the CN-categories, especially the high number of CNT responses (see Prinzmetal, Diedrichsen, & Ivry, 2000). Our observers reported the wrong letter on less than 2.5% of the trials (all 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 are instances where 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 seen in the table, 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 paper.

Model Analysis. We start with a basic model based on Figure 3 in which there are five parameters that are adjusted to provide the best fit of 20 data points. We fit the model separately for the near and far conditions for each subject. 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. The fits were obtained using the method of gradient decent so as to minimize G^2 in the manner described by Dodson, et al. (1998). To ensure that the fits did not represent local minimums, at least 10 random starting values for each parameter were used. Table 3 gives the mean parameter values. In general, the fits were close to the observed data. The sum of the squared error (SSE) averaged .002 for the Norepeat group (range .000 to .004) and .001 for the Repeat group (range .000 to .002).[4]

Insert Table 3 about here

The model analysis 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, $t(5) = .40$ and .004 for *TL* and *NC* parameters, respectively. 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 vs. far. The parameter *TC* 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 with the model. Whatever the cause of the difference, as we will show below, it does not affect the interpretation of the other parameters in this experiment.

The critical parameters in this experiment were the feature binding parameter, _, and the guessing parameters, *g*. Recall that _ = 1.0 represents perfect binding, and _ = 0.5 is chance binding. If the model is correctly measuring feature integration errors, _ should be less in the near condition than the far condition. For the far condition, _ = 0.98, indicating nearly perfect feature binding. For the near condition, _ = 0.73, indicating less than perfect, but not random binding. The difference was reliable, $t(5) = 8.83$, p<.001.

Note that the differences in TC for near and far conditions can not account for the difference in _. The difference between *TC* for the near and far conditions would be expected to minimize differences in ____ Failures in feature binding (1 - _ branches in Figure 3) only occur when one or more of the color features are perceived. Thus, the difference in _ 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 Norepeat 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 was the same as a perceived color should be 0.0. In accord with this prediction, the parameter *g* averaged 0.01 and 0.00 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 ___ and the observers rarely reported seeing both letters in the same color as reflected in the low estimates of the parameter *g*.

Turning to the Repeat group, the results for the feature perception and integration parameters were quite similar as those reported for the Norepeat 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, $t(5) = .15$ and .20 for *TL* and *NC* parameters, 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 the near condition, .96 vs. .93, $t(5) = 5.71$, p < .01 (two tailed). The feature binding parameter, _, was higher for the far condition, _ = 0.97, than for the near condition, _ = 0.75. The difference was reliable, $t(5) = 10.17$, p<.001.

As predicted, the estimate of the parameter g was radically different for the Repeat group compared to the Norepeat 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, $g = .28$, was close to this value. However, the value for the near condition was much higher than expected, $g = .42$. 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, *g*, should reflect factors that affect guessing, not stimulus factors. The fact that *g* for the far condition averages near optimal guessing (i.e., 0.25) is encouraging, but the fact that it varies between the near and far conditions indicates a problem with the model.

Alternative Model for Repeat Group. We have investigated various alternative models and present below the one that we believe provides the most parsimonious account of the data. In our basic model (Figure 3), failures of feature binding (1 - _), result in a transposition of the two colors. That is, the target and nontarget color switch positions (or if only one color was perceived, it switches to the other letter). It is also possible that a color might 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 phenomena 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 paper. 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 Norepeat 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. Since 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 will call *s*. This interpretation suggests that colors spread over letter with a probability of .14 (*g* near - *g* far).

We tried several models that included a parameter, *s*. However, it soon became apparent that with our task, *g* (guessing two colors were identical) and *s* (a color spreading over two letters) can not be separately estimated; models with both *s* and *g* are mathematically equivalent. To understand this restriction, consider a stimulus that contains a red T and a green O. A response of "red T, red O" could arise because the color red spread over both letters, or because the observer perceived the color of the T and guessed the color of the O was the same. We can not distinguish between these two scenarios since they both involve reporting the same features. Nonetheless, two considerations favor the spreading hypothesis. First, we have had the phenomenal experience of seeing both letters in the same color even when practicing under no repeat conditions. Second, it seems reasonable that a guessing parameter should not vary between stimulus conditions. The need for an additional parameter is only apparent in the Repeat condition.

Discussion

The modeling results from Experiment 1 are, for the most part, in accord with the predictions we derived based 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 to one in which there would be few feature integration errors. In the near condition, binding errors, as indexed by the parameter _, were fairly common. In the far condition, feature binding was nearly perfect as indicated by _ values close to 1.0 for all observers. Note that, while we use the terms "near" and "far", there are differences between the two conditions in addition to just the distance between the items. Our focus here is not on why binding failures do not occur under some conditions (e.g., distance, 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 probability 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, _.

The second goal of the experiment was to see if observers would be sensitive to the probability of repeated colors in the display, and whether this would this influence their guessing behavior. For one group of observers, the displays never contained two identical colors (Norepeat group); for the other group, a single color was used for both the target and nontarget letter on one fourth of the trials (Repeat group). As expected, observers in the Norepeat condition rarely reported the same color for both the target and nontarget letter, and the estimated probability of guessing the same color, *g*, was near 0. Ashby et al. (1996) had assumed observers use an exclusionary strategy when the feature information is incomplete (e.g., only one color perceived). For example, having perceived red for the nontarget item, observers would not guess red for the target item. The results for the Norepeat group support this assumption.

For the Repeat group, observers were sensitive to the probability of repeated colors. The average probability 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 fact. However, we had not anticipated that the parameter *g* would vary as a function of stimulus conditions. We had predicted that it would be the same for the near and far conditions. This finding has led us to consider an alternative model in which colors will sometimes spread from one letter to the other, an event that is more likely to occur when the distance between the target and nontarget is small.

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 is 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 were very clear in indicating that the observers in the Norepeat 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. Because 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 it occurred varied from 20 to 45% of the trials. Only one observer in the Norepeat group reported seeing a color repeated, and she reported that this occurred on 10 percent of the trials (subject 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 turn out attention to the parameters associated with the 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 were varied independently for the target and nontarget letters. Thus, there were four saturation conditions (i.e. target and nontarget saturated, target saturated/nontarget unsaturated, target unsaturated/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 the feature binding. The model sketched in Figure 3 does not assume any affect. However, our informal observations 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 only included 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 used in Experiment 1. The only difference was that there were three color choices (red, green, and blue). We did not include yellow since we were unable to generate a satisfactory version of an unsaturated yellow. The exposure duration was adjusted as in Experiment 1. During the test blocks, the exposure duration averaged 67 milliseconds (4 computer refreshes at 60 hz.)

There were 96 trials in a block. These were evenly divided between the near and far conditions (see Figure 2) and 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 for 6 blocks per 1-hour session for 9 days with the first session used as practice. A total of 4,608 trials per observer were used in the analyses.

Stimuli. The same stimulus configurations as in Experiment 1 were used. There were 3 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); 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 less than the saturated colors. However, as indicated below, this did not affect the probability of perceiving the target letter (*TL*).

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 proportion 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 are only 18 response categories per condition because there were only 3 possible colors. For all conditions, the observers report all three features and correctly bind the target features (CTN) on at least a majority of the trials. Such reports are greatest when both target and nontarget are saturated, and lowest when they are both unsaturated. For the near condition, the second highest category of responses is the conjunction response category, CNT, which varies from .105 to .242. In the far condition, the average CNT response is only from .021 to .047. Thus, as in Experiment 1, the raw data indicate that many more conjunction responses occur in the near condition.

Insert Table 4 about here

Model Analysis. We fitted 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 0, 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 (8 fits/observer), we do not show the results for individual observers. However, to evaluate the patterns in the parameter estimates, we conducted a repeated measures analsysis of variance on each of the parameters. Factors in this analysis were distance, target saturation, and nontarget saturation.

Insert Table 5 about here

The parameter *TL* did not significantly vary as a function of any of the independent variables. The effect of distance, target saturation, and nontarget saturation were $F(1,5) = 4.59$, 3.78, and 4.32, respectively. 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 factor 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* (below) that were unanticipated. However, these effects seem 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 0.85; 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 is 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 compared to 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 the parameter *NC* was the level of nontarget saturation. *NC* averaged 0.97 when the nontarget color was saturated, but only 0.73 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, 0.89 vs. 0.81, $F(1,5) = 97.36$, $p < .01$, a result that was unexpected. One possibility is that, as attention shifts towards 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 on *NC* as was described with *TC*. *NC* was slightly lower when the target color was saturated than not saturated, 0.84 vs 0.86. 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.

Insert Figure 5 about here

Similar to Experiment 1, the distance manipulation had a marked effect on _, the feature binding parameter. In the far condition, feature binding was nearly perfect, 0.96; in the near condition, it was 0.77, $F(1,5) = 20.97$, $p < .01$. More interesting, the saturation manipulation produced some interesting effects on feature binding (Figure 5). As noted above, we have previously informally observed 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 clear that feature binding errors were more common when the target was not saturated (see Figure 5). The parameter _ was .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 on _, $F(1,5) = 13.02$, $p < .05$. When the target is not saturated, feature integration is more accurate (i.e., higher _) when the nontarget is saturated than when it is unsaturated. However, when the target is saturated, the opposite is true: correct feature integration is greater with the nontarget is saturated than when it is saturated.

Discussion

The purpose of Experiment 2 was to test the hypothesis that the estimates of the parameters TC and NC will be sensitive to the salience of the target and nontarget colors. We varied salience by varying the 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, _, is subject to distance constraints.

The modeling work revealed a number of unexpected results. These unexpected finding 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 we will 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 observed interactions may reveal important insights into the binding process.

Consider the finding that the parameter *TC* varied as a function of the nontarget color saturation and, similarly, the parameter *NC* varied as a function of the target color saturation. *TC* was higher when the nontarget color was saturated compared to when the nontarget was not saturated. Moreover, this effect was greater when the target color was not saturated. This pattern suggests that the 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, _, would only be affected by the distance manipulations. However, 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 the 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. First, there is the interaction involvine the target and nontarget saturations on 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. Second, there is the interaction of salience on 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 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 _are likely to be higher on trials in which the observers were confident, compared to trials in which their confidence was low. A priori, we do not know if some of the parameters will 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 when they comment "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 only had two levels of confidence ("sure" and "think") so that 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 only required to report the target color and target letter. Subsequent to this, they were required to gave a confidence rating (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), a colored nontarget letter (the letter O). The observer responded with the color and identity of the target letter by clicking the appropriate button on 3 x 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 9 buttons with 1 for least confident and 9 for most confident. Thus, as in experiments 1 and 2, observers made two "clicks" each trial. All other aspects of the procedure were as in the previous experiment. Over all the test sessions, the exposure duration averaged 107 ms.

Each observer was tested for five 1-hour 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 2880 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 3a and 3b. 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.

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

Results

Raw response data. The mean number of responses for each of the 6 response categories and 9 confidence ratings, averaged over observers, is given in Table 6. The most frequent response category corresponds to correct responses, CT, and the highest error category is the conjunction responses, CN. As in the other two experiments, conjunction responses are much more likely when the target letter is identified, CN, compared to when the wrong letter is 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, Diedrichsen, & Ivry, 2000).

Insert Figure 6 and Table 6 about here

The raw data give a clear picture of observers' confidence when they make a conjunction response (CN). Figure 6 plots the proportion 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 proportion of correct responses were given the highest confidence level. When observers made conjunctions responses they were, on average, less confident than when they were correct. Nevertheless, confidence for correct responses and conjunction responses form overlapping distributions so that there are a substantial number of trials on which correct responses have 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 can not 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 $g = 0$, eliminating one free parameter. Second, since there were only 3 colors in the experiment, the response categories CØ and IØ do not exist. Even though there were 2880 responses per observer, when the data was broken down into 9 response categories, there were many empty cells. 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, C8-C9). Thus for each of the seven observers we solved for *TL*, *TC*, *NC*, and _ at four confidence levels. The fits, averaged over the 7 observers, are given in Table 7. We conducted a one-way repeated ANOVA on each of the parameters as a function of confidence.

Insert Table 7 about here

The parameter *TL* significantly differed for the four different confidence categories. For the lowest confidence level it averaged .787 and for the highest confidence level it averaged .996, $F(3,18) = 18.42$, $p < .01$. The parameters that represent the perception of the target and nontarget color, *TC* and *NC* also rose as confidence increased. However, the estimates of *TC* and *NC* did not significantly increase with confidence, F(3,18) = 1.5 and1.93 for *TC* and *NC*, respectively, likely due to a ceiling effect. As can be seen in Table 6, responses that were classified as color feature errors (CO or IO) only 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 _ varied from .559 for the least confident category to .940 for the most confident category. The difference between _ 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 only reliable for TL and _.

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 _, the likelihood of making an illusory conjunction increases as confidence decreases. However, the results are in accord with the claim of Triesman 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 subjects in many illusory conjunction experiments and continue to be amazed that a very clear perception of, for example, a red T and blue O is wrong, that the display actually contained a blue T and 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 isn't obvious that these sorts of effects would influence binding. Binding requires that at least one shape and one color are perceived, and the _ parameter describes the likelihood that the features will be bound correctly.

Why would binding become less accurate as confidence goes down (or, alternatively, why does confidence go down as binding becomes less accurate)? To account for this, it is important to consider the underlying mechanism or mechanisms that might cause variation in _. Ashby et al. (1996) argue 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 spatial representation of the red stimulus compared to the blue stimulus. While the models described by Ashby et al. are similar to that depicted in Figure 3, _ was not computed separately for each distance as in the present experiments, but rather 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., Robertson, & Treisman, 1995;.Logan, 1996; Prinzmetal & Keysar, 1989). When seen this way, 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.

General Discussion

The goal of the experiments reported here was to test the adequacy of using multinomial models for investigating feature integration. We used two strategies. In Experiments 1 and 2, stimulus factors were varied to test specific predictions derived from assumptions 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 inter-object distance and grouping. The feature binding parameter, _, was significantly affected by these manipulations in both Experiments 1 and 2. In Experiment 1, we also compared conditions in which the colors were selected on each trial without replacement to 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 and nontarget color, *TC* and *NC*, were higher with saturated colors than with nonstaturated colors.

In the third experiment, we took a novel approach for testing the adequacy of the multinomial approach. We obtained confidence ratings after each trial. We assumed that on trials with low confidence, observers would be guessing more often 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, due to a ceiling effect on color accuracy, this effect was only significant for the parameters TL and _.

While the results were generally in accord with our predictions, 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 when the two items were close to one another. One hypothesis is that when items are close together, observers are more willing to guess that they were 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 since 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 (explictly or implicitly) that the two colors are always different. Hence the parameter g is essentially 0 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 affect on the binding parameter, _. Feature integration errors occurred at a higher rate when the target color was less saturated. 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 paper are not process models; rather they are intended as measurement tools. Thus while discovering 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, it is important to consider models that assume that the representation (neural or psychological) of a colored stimulus can be described as a distribution of activation over space (e.g., Ashby, et al., 1996; Logan, 1996). It may be that an unsaturated color produces a weaker activation and thus the representation of its location is more susceptible to noise. Attention may also be drawn to saturated colors. Note that _ was affected mostly be the target color saturation, not the nontarget saturation.

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, like signal detection theory (SDT) 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 4 people: 2 of the sources were male and 2 were female. At test, participants were presented a list of words and they had to determine whether the words were old or new. For items judged old, observers also had to indicate the source. The data was 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: know that the color was a "warm color" (red or orange), even if they 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 which specific color it was.

A comparison of how continuous-state models (e.g., SDT) 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 on the nature of that partial information. The multinomial approach does not have infinite levels of partial information, but is also more constrained since it forces the investigator to precisely characterize the information (e.g., the discrimination of warm vs. cool colors).

Another criticism of the 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 where judgements about color and shape are not independent (e.g., Bonnel & Prinzmetal, 1998). There almost surely is a non-zero 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 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 where one suspects that the probability of perceiving the target color and target letter are highly correlated. Working from the model in Figure 1, the correlated version would begin in the same manner at Level I, using a single parameter for TL. At Level II, instead of one parameter, TC, there are two parameters: TC given the target letter was perceived ($TC \mid TL$) and TC given the target letter was not perceived ($TC \mid \sim TL$). In this manner, the correlation can be tested. If the parameters ($TC \mid TL$) and ($TC \mid \sim TL$) are not significantly different, then it is reasonable to simplify the situation and just use a single TC parameter. Indeed, in the studies presented here, it was not necessary to use one of these models.

In conclusion, the multinomial approach provides a rigorous yet flexible tool for the study of feature integration. While 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. (in press). Recognition and source memory as multivariate decision processes. Psychological Science,

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

Fischer, e. H., & Laming, D. (Eds.). (1994). Analysis of a model for source monitoring. New York : Springer-Verlag:

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. Peycholonomic 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 contrain 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. Psychologicical 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.

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

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.

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, 1999, 1207-1253.

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). 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 its 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., & Millis-Wright, M. (1984). Cognitive and linguistic factors affect visual feature integration. Cognitve Psychology, 16, 305-340.

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

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

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., 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., 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., & Wilson, A. (1997). The Effect of Attention on Phenomenal Length. Perception, 26, 193-205.

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. (2000). Illusory Conjunctions are Alive and Well: A Reply to Donk (1999). Journal of Experimental Psychology: Human Perception and Performance, submitted for publlication,

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? file: 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. (in press). Pitch and duration in unfamiliar tone sequences. Journal of Experimental Psychology: Human Perception & Performance,

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

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

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

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

Table 1

Stimulus: X_{red} O_{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 Conj. Error
T_{green}	IO	Color Letter Feature Error

Table 1. Sample stimulus display and response categories in simple feature integration experiment.

Table 2

Nonrepeat Group

	Target Letter Correct			Target Letter Wrong	
Response type	Near	Far		Near	Far
CTT	0.0006	0.0003	ITT	0.0000	0.0000
CTN	0.6421	0.8782	ITN	0.0056	0.0146
CTO	0.0580	0.0638	ITO	0.0025	0.0027
CNT	0.2325	0.0148	INT	0.0081	0.0019
CNN	0.0006	0.0000	INN	0.0001	0.0003
CNO	0.0233	0.0039	INO	0.0025	0.0012
COT	0.0056	0.0025	IOT	0.0012	0.0006
CON	0.0139	0.0120	ION	0.0007	0.0023
COØ	0.0020	0.0009	IOØ	0.0006	0.0001
COO	0.0001	0.0000	IOO	0.0000	0.0000

Table 2

		Repeat Group			
	Target Letter Correct			Target Letter Wrong	
Response type	Near	Far	Response type	Near	Far
CTT	0.0199	0.0110	ITT	0.0014	0.0014
CTN	0.6578	0.8818	ITN	0.0104	0.0139
CTO	0.0272	0.0313	ITO	0.0010	0.0012
CNT	0.2149	0.0262	INT	0.0075	0.0029
CNN	0.0318	0.0079	INN	0.0015	0.0037
CNO	0.0116	0.0025	INO	0.0008	0.0006
COT	0.0027	0.0014	IOT	0.0002	0.0000
CON	0.0093	0.0131	ION	0.0010	0.0000
COØ	0.0000	0.0002	IOØ	0.0002	0.0000
COO	0.0008	0.0002	IOO	0.0002	0.0010

Table 2. Proportion of each response category for Nonrepeat and Repeat groups. Note for repeat group, proportions are based on 2880 trials per observer and for the nonrepeat group, proportions are based on 2160 trials per observer. Trials on which a color repeated are not included in the table or analysis.

Table 3
Nonrepeat Group

Subject:	S1	S2	S3	S4	S5	S6	Mean
TL near	0.951	0.972	0.984	0.943	0.910	0.984	0.957
TL far	0.967	0.983	0.981	0.878	0.920	0.986	0.953
TC near	0.922	0.959	0.950	0.923	0.872	0.980	0.934
TC far	0.955	0.984	0.981	0.953	0.955	0.981	0.968
NC near	0.917	0.956	0.937	0.893	0.650	0.987	0.890
NC far	0.971	0.946	0.913	0.863	0.684	0.963	0.890
_ near	0.737	0.689	0.713	0.760	0.828	0.650	0.729
_ far	0.984	0.978	0.989	0.966	0.968	0.988	0.979
g near	0.000	0.008	0.014	0.000	0.000	0.080	0.017
g far	0.000	0.010	0.000	0.000	0.002	0.014	0.004
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

Table 3
Repeat Group

Subject:	S1	S2	S3	S4	S5	S6	Mean
TL near	0.937	1.000	0.924	0.935	0.958	0.963	0.953
TL far	0.935	0.988	0.984	0.912	0.947	0.940	0.951
TC near	0.918	0.985	0.909	0.900	0.905	0.949	0.928
TC far	0.967	0.995	0.954	0.939	0.936	0.974	0.961
NC near	0.939	0.980	0.922	0.901	0.888	0.959	0.932
NC far	0.973	0.989	0.949	0.880	0.833	0.983	0.934
_ near	0.730	0.822	0.794	0.735	0.735	0.680	0.749
_ far	0.973	0.992	0.993	0.961	0.898	0.984	0.967
g near	0.465	0.334	0.511	0.370	0.319	0.492	0.415
g far	0.450	0.165	0.305	0.210	0.155	0.402	0.281
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

Table 3. Fits for Experiment 1. The parameters are as follows: *TL* (probability of perceving the target letter, *TC* (probability of perceiving the target color), *NC* (probability of perceiving the nontarget color); _ (the probability of correctly binding colors and letters), *g* (the probability of guessing two identical colors).

Table 4
Near Condition

RESPONSE	**Target Saturated; Nontarget Saturated**	Target Unsaturated; **Nontarget Saturated**	**Target Satrated;** Nontarget Unsaturated	Target Unsaturated; Nontarget Unsturated
CNO	0.002	0.065	0.001	0.051
CNN	0.001	0.001	0.000	0.000
CNT	0.174	0.229	0.105	0.242
CON	0.003	0.088	0.002	0.039
COO	0.000	0.000	0.000	0.000
COT	0.000	0.001	0.011	0.019
CTO	0.006	0.008	0.114	0.063
CTN	0.801	0.584	0.759	0.567
CTT	0.000	0.000	0.000	0.000
INO	0.000	0.001	0.000	0.003
INN	0.000	0.000	0.000	0.000
INT	0.004	0.009	0.004	0.005
ION	0.000	0.003	0.000	0.001
IOO	0.000	0.000	0.000	0.000
IOT	0.000	0.000	0.000	0.001
ITO	0.000	0.000	0.001	0.001
ITN	0.008	0.010	0.002	0.008
ITT	0.000	0.000	0.000	0.000

Table 4
Far Condition

RESPONSE	**Target Saturated; Nontarget Saturated**	Target Unsaturated; **Nontarget Saturated**	**Target Satrated;** Nontarget Unsaturated	Target Unsaturated; Nontarget Unsturated
CNO	0.000	0.010	0.001	0.023
CNN	0.001	0.001	0.001	0.001
CNT	0.021	0.027	0.030	0.047
CON	0.003	0.109	0.002	0.065
COO	0.000	0.000	0.001	0.000
COT	0.001	0.007	0.011	0.023
CTO	0.014	0.017	0.167	0.121
CTN	0.951	0.806	0.780	0.707
CTT	0.000	0.001	0.000	0.000
INO	0.000	0.002	0.000	0.002
INN	0.000	0.000	0.000	0.000
INT	0.003	0.004	0.001	0.002
ION	0.000	0.005	0.001	0.002
IOO	0.000	0.000	0.000	0.000
IOT	0.000	0.000	0.000	0.000
ITO	0.000	0.001	0.001	0.001
ITN	0.006	0.011	0.004	0.007
ITT	0.000	0.000	0.000	0.000

Table 4. Results for Experiment 2 broken down into response type and color condition.

Table 5

Condition		*TL*	*TC*	*NC*	_	G^2	SSE
Target	Near	0.973	0.988	0.986	0.819	27.083	1.69E-04
Nontarget	Far	0.981	0.990	0.969	0.975		
Target	Near	0.954	0.685	0.979	0.695	52.619	2.95E-03
Nontarget	Far	0.954	0.742	0.945	0.958		
Target	Near	0.986	0.992	0.747	0.879	31.055	4.12E-04
Nontarget	Far	0.987	0.988	0.637	0.959		
Target	Near	0.961	0.817	0.832	0.686	87.745	2.64E-03
Nontarget	Far	0.973	0.819	0.702	0.929		

Table 5. Parameters and goodness of fit for Experiement 2. For Target and Nontarget, bold indicates saturated colors, plain text indicates unsaturated colors.

Table 6. Responses in Experiment 3.

Confidence	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.
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	2880

Table 6. Number of responses as a function of response type and confidence ratings in Experiment 3 averaged over observer.

Table 7

Confidence	C1-C3	C4-C5	C6-C7	C8-C9
TL	0.787	0.944	0.977	0.996
TC	0.973	0.992	1.000	1.000
NC	0.929	0.985	0.971	0.976
–	0.559	0.729	0.848	0.940
G^2	29.529	23.650	15.730	10.645
sse	4.09E-03	1.25E-03	2.01E-04	3.05E-05

Table 7. Average fits for observers in Experiment 7 with confidence rating as the independent variable.

Author Notes

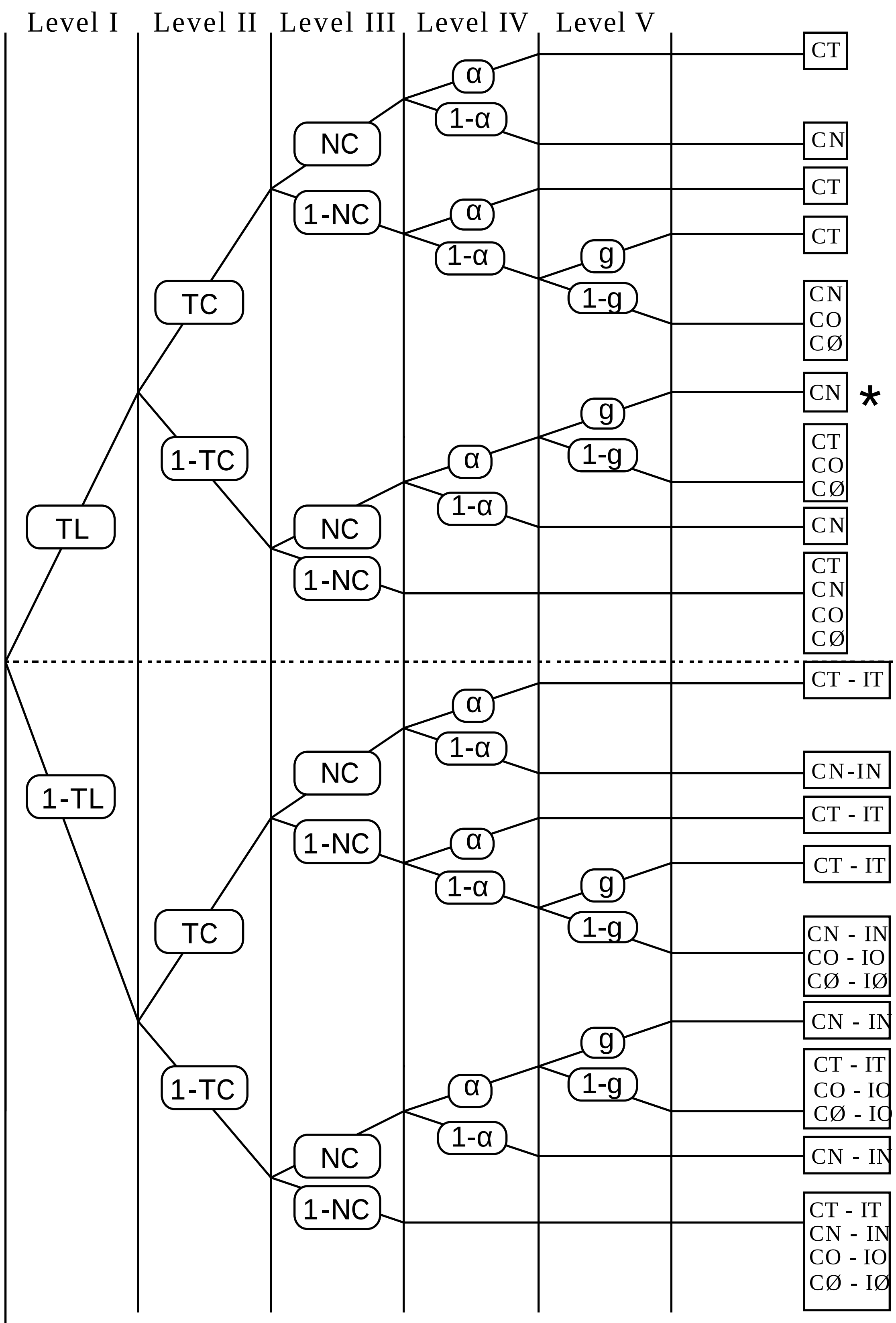
This research was supported by National Science Foundation grant SBR-9319103. Reprint requests should be sent to W. Prinzmetal, Department of Psychology, University of California, Berkeley, CA 94720-1650 or to wprinz@socrates.berkeley.edu.

Figure Captions

Figure 1. Tree diagram for simple illusory conjunction experiment.
Figure 2. Sample stimuli from Experiment 1. Panels a and b are examples of the near condition, c and d are example of the far condition. The + marks the center of the monitor.
Figure 3. Model for illusory conjunctions when observers respond to the target letter, target color and nontarget color. Only trials in which the observer correctly identified the target letter are included in this panel.
Figure 4. Response palette used in Experiments 1 and 2. The observer responded by "clicking" the mouse on buttons located on the screen.
Figure 5. Feature integration parameter, _, as a function of target and nontarget saturation.
Figure 6. Proportion of CN, CT, and all other responses as a function of confidence. Note that each function sums to 1.0.

Footnotes

1. With 6 response categories and 5 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 that the number of parameters is always much less that the number of data points. Furthermore, even in the simplest case, there are internal constraints in the model so that a perfect fit is not guaranteed (see Riefer, Hu, & Batchelder (1994).
2. This treatment of perceiving the target letter is slightly different from 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. By using the same structure 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 et al. (1998) for a tutorial on a simple method for fitting multinomial data using the Solver function in Excel™. This method has the advantage in 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/.
4. The values of G^2 were generally significant by a chi square test suggesting that models should be rejected. However, the number of observations per observer for the Nonrepeat and Repeat groups was 2304 and 1728, respectively, hence the power exceeds .9999 (Buchner, Faul, & Erdfelder, 1996). In these circumstances, it is not unexpected to get a significant G^2 (Batchelder & Riefer, 1994).
5. An alternative method to determine whether parameters differ is to fit each observers data with a reduced model. For example, a reduced model would assume that TL_{near} and Tl_{far} the same. If the reduced model is not significantly different than the full model then 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.



Prinzmetal, et al., Figure 1

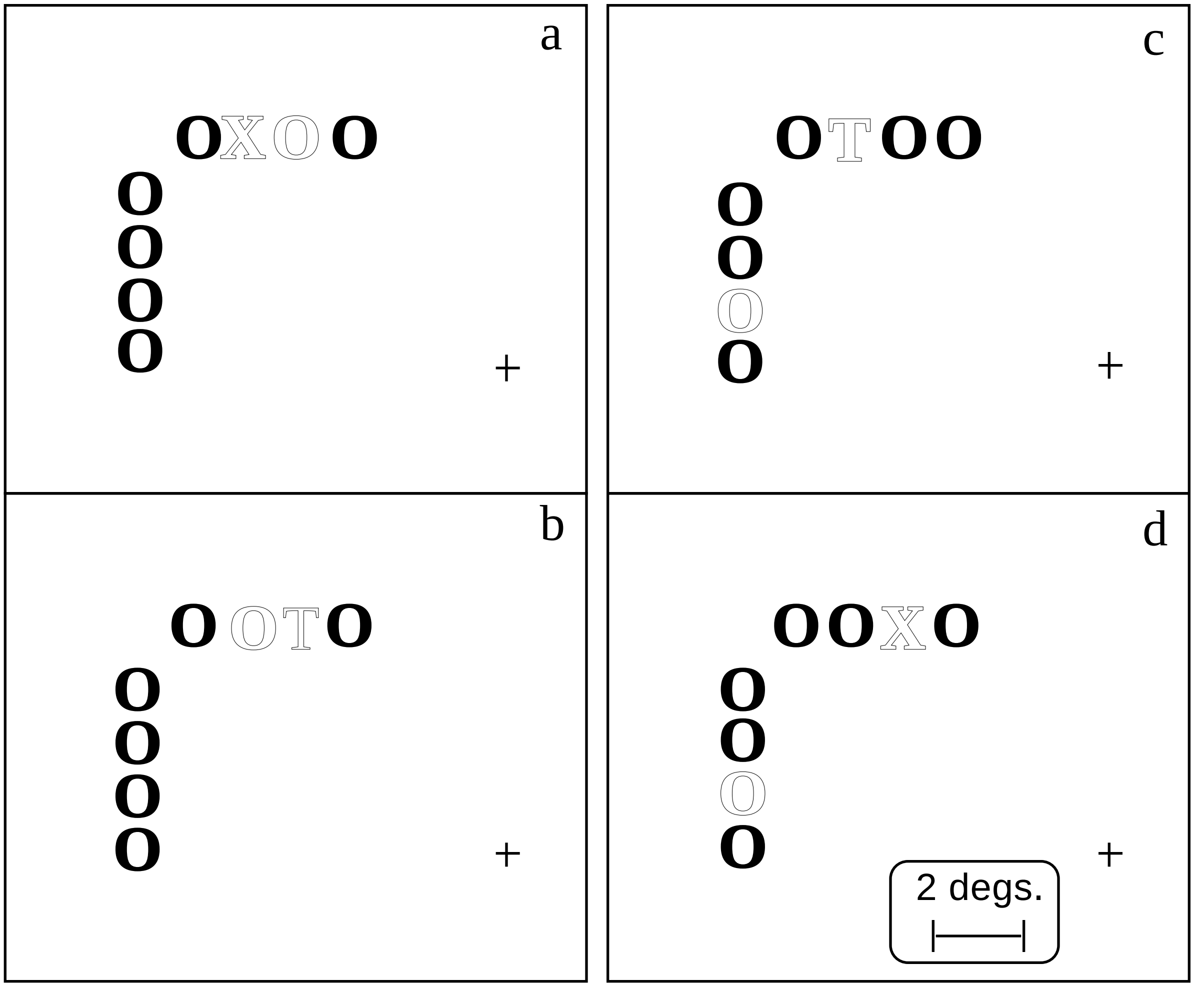


Figure 2

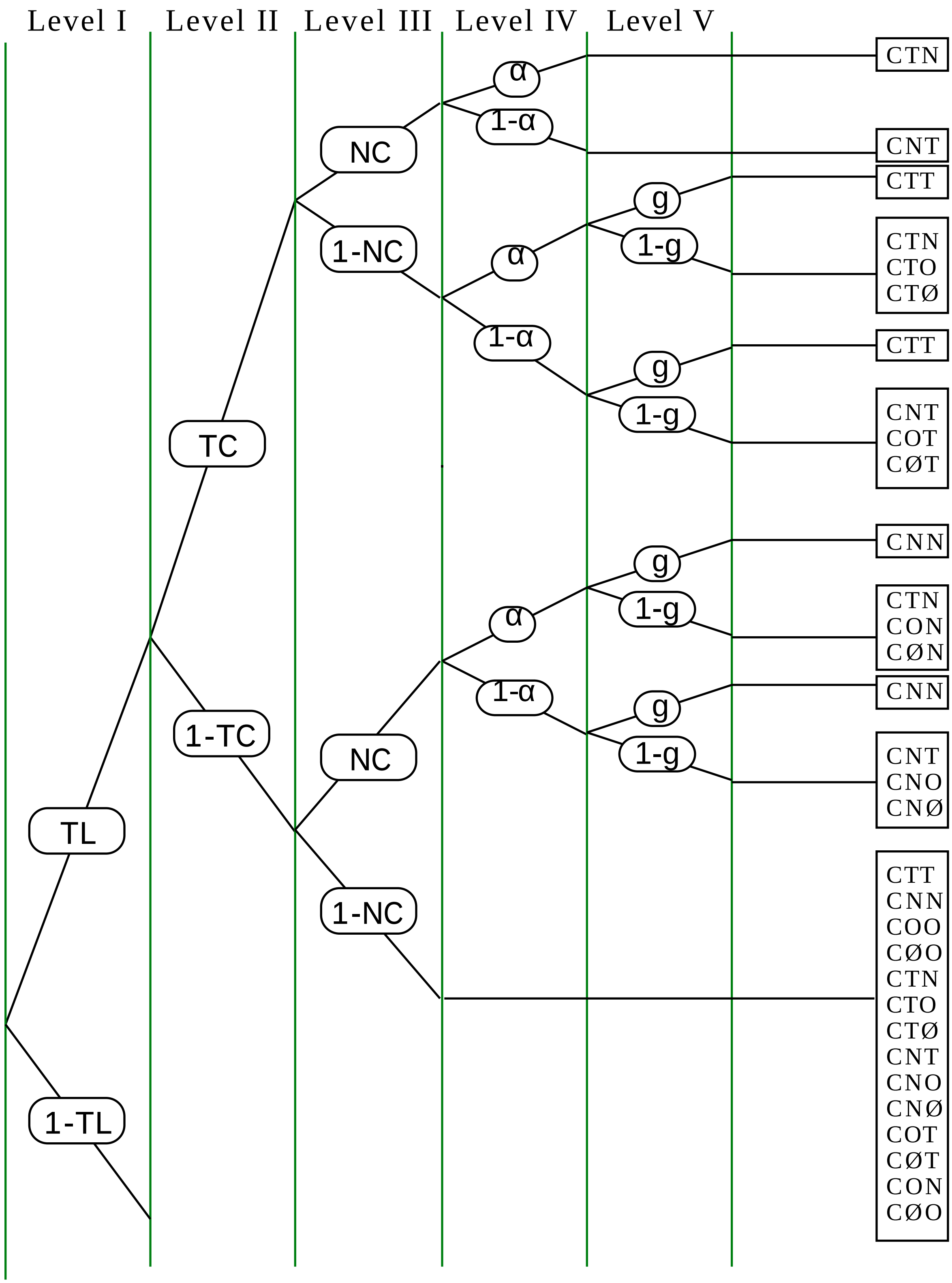

Level I
Level II
Level III
Level IV
Level V
TL
1-TL
TC
1-TC
NC
1-NC
NC
1-NC
α
1-α
α
1-α
α
1-α
g
1-g
g
1-g
g
1-g
g
1-g
CTN
CNT
CTT
CTN
CTO
CTØ
CTT
CNT
COT
CØT
CNN
CTN
CON
CØN
CNN
CNT
CNO
CNØ
CTT
CNN
COO
CØO
CTN
CTO
CTØ
CNT
CNO
CNØ
COT
CØT
CON
CØO

Figure 3

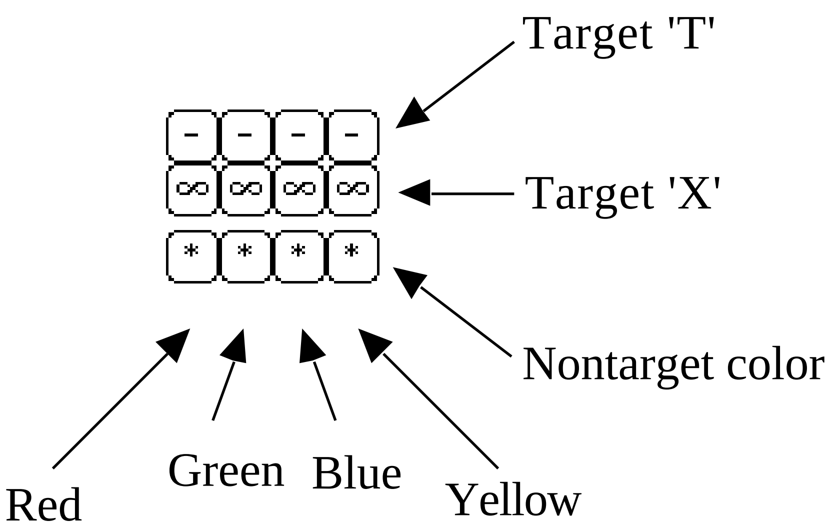


Figure 4

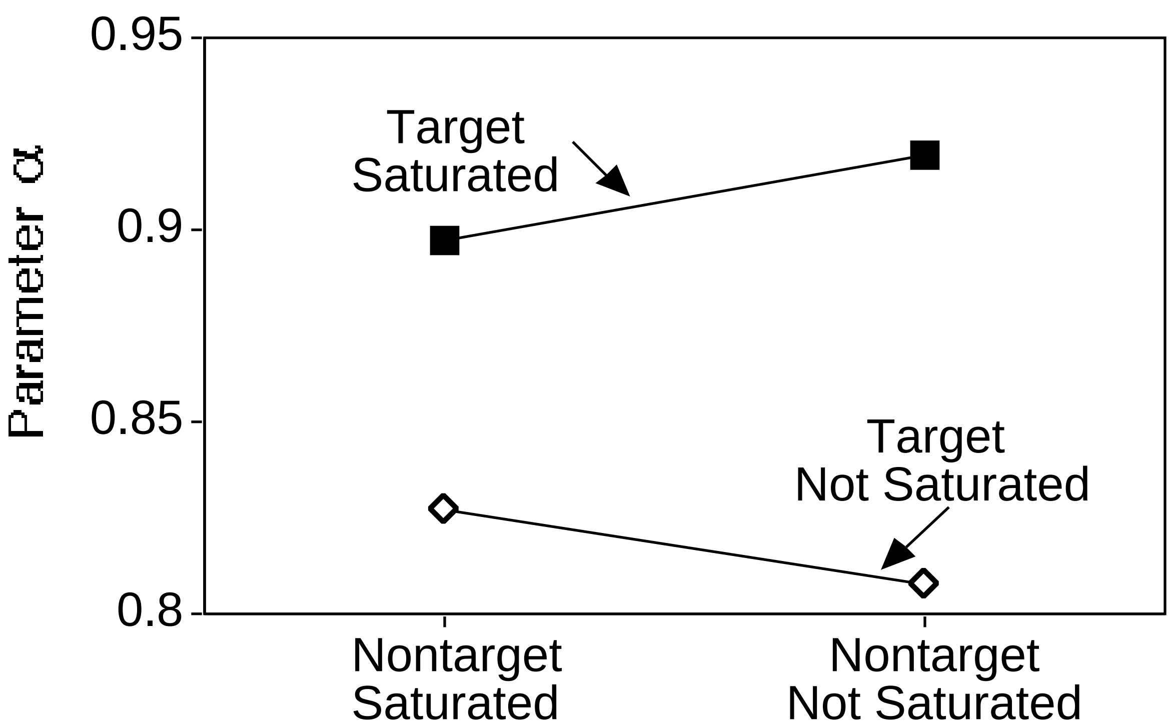
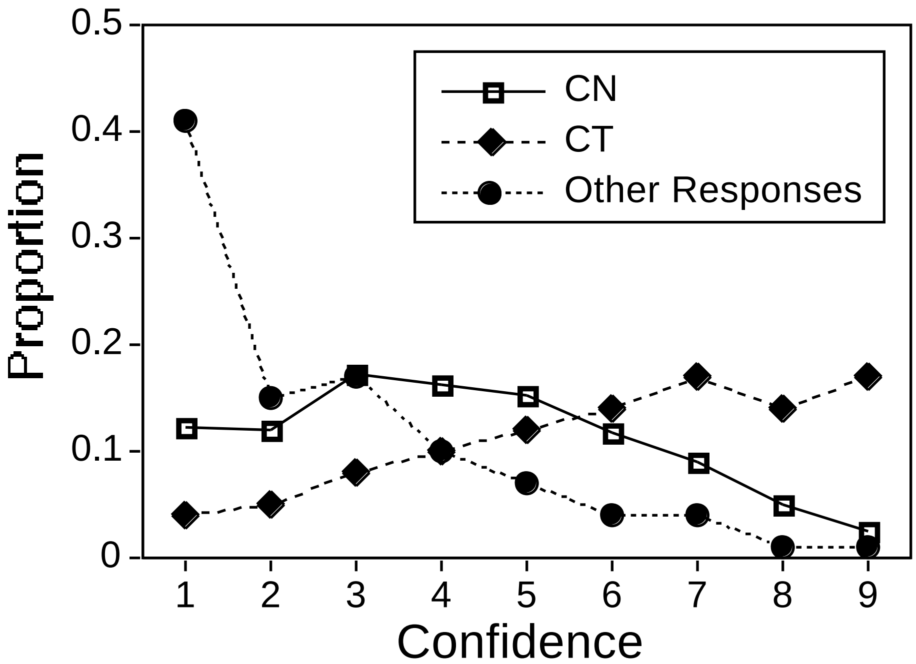


Figure 5



0.5
0.4
0.3
0.2
0.1
0
Proportion
CN
CT
Other Responses
1
2
3
4
5
6
7
8
9
Confidence

Figure 6