View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by City Research Online

Hampton, J. A., Estes, Z. & Simmons, C. L. (2005). Comparison and contrast in perceptual categorization. Journal of Experimental Psychology: Learning, Memory and Cognition, 31(6), 1459 - 1476. doi: 10.1037/0278-7393.31.6.1459 <http://dx.doi.org/10.1037/0278-7393.31.6.1459>

CITY UNIVERSITY
LONDON

City Research Online

Original citation: Hampton, J. A., Estes, Z. & Simmons, C. L. (2005). Comparison and contrast in perceptual categorization. Journal of Experimental Psychology: Learning, Memory and Cognition, 31(6), 1459 - 1476. doi: 10.1037/0278-7393.31.6.1459 <http://dx.doi.org/10.1037/0278-7393.31.6.1459>

Permanent City Research Online URL: http://openaccess.city.ac.uk/1002/

Copyright & reuse

City University London has developed City Research Online so that its users may access the research outputs of City University London's staff. Copyright © and Moral Rights for this paper are retained by the individual author(s) and/ or other copyright holders. Users may download and/ or print one copy of any article(s) in City Research Online to facilitate their private study or for non-commercial research. Users may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. All material in City Research Online is checked for eligibility for copyright before being made available in the live archive. URLs from City Research Online may be freely distributed and linked to from other web pages.

Versions of research

The version in City Research Online may differ from the final published version. Users are advised to check the Permanent City Research Online URL above for the status of the paper.

Enquiries

If you have any enquiries about any aspect of City Research Online, or if you wish to make contact with the author(s) of this paper, please email the team at publications@city.ac.uk.

Running Head: PERCEPTUAL CATEGORIZATION

To appear in *Journal of Experimental Psychology: Learning Memory and Cognition, 2005*

Comparison and contrast in perceptual categorization

James A. Hampton, City University, London

Zachary Estes, University of Georgia

Claire L. Simmons, City University, London

Please address correspondence to:

James A. Hampton

Department of Psychology

City University

Northampton Square

London EC1V OHB

UK

Abstract

We report a series of five experiments in which people categorized pairs of perceptual stimuli varying in degree of category membership and in similarity to one another. Experiments 1-4 investigated color categorization. Experiment 1 required conjunctive or disjunctive category judgments, and Experiment 2 required simultaneous but separate categorization of each stimulus. In both experiments, categorization of one color was reliably contrasted from that of the other: As one color became more likely to be included in the category, the other color became more likely to be excluded. This similarity-based contrast effect occurred only when the context stimulus was relevant for the categorization of the target stimulus (Experiment 3). The effect was not simply owing to perceptual color contrast (Experiment 4), and it extended to pictures from common semantic categories as well (Experiment 5). Results were consistent with a sign-and-magnitude version of Stewart and Brown's (2005) SD-GCM model, in which both similarity to a target category and difference from a contrasting category inform categorization choices. The data are also modeled in terms of the balancing mechanism of Criterion Setting Theory (Treisman & Williams, 1984), whereby the decision criterion is systematically shifted toward the mean of the current stimuli.

The categorization of perceptual stimuli, such as colors and pictures, is one of the most basic functions of cognition. But how are such category decisions made? Traditional models of perceptual categorization posit that the target stimulus is initially compared to the category representations in long-term memory, and is subsequently judged to belong in the category to which it is most similar (for review see Murphy, 2002, Ch. 2). For example, Nosofsky's (1986) Generalized Context Model (GCM) first computes the similarity of a target to some set of stored exemplars from each candidate category, and then applies Luce's Choice Axiom to derive the probabilities that the target will be judged to belong in one category or another. Alternatively, in Ashby & Gott's (1988) Decision Bound model, a stimulus is deterministically allocated to a category depending on its position in stimulus space relative to some fixed decision criterion separating the candidate categories (see also Gardenfors, 2000). The location of a target in stimulus space is essentially its similarity to the other exemplars. Thus, in one form or another, most traditional models share the common assumption that categorization is similarity-based.

However, recent evidence suggests that these similarity-based models may be too simplistic. In particular, when a number of targets are categorized in turn, the similarity of successive targets may additionally influence categorization (DeCarlo, 2003; Petzold & Haubensak, 2001, 2004; Stewart & Brown, 2004; Stewart, Brown & Chater, 2002; Treisman & Williams, 1984). This inter-stimulus similarity can provide either a positive or a negative cue to categorization: If two instances are *similar*, they likely belong in the same category, and if two instances are *dissimilar*, they likely belong in different categories (Stewart & Brown, 2004). Long-term category representations with fixed decision criteria (e.g., the Decision Bound and GCM models) are unable to account for such effects of stimulus comparison, as described below. We report five experiments that investigate how the similarity or dissimilarity of two stimuli affects their categorization. The importance of such inter-stimulus effects lies in their implications for models of perceptual categorization.

Similarity-based Models of Categorization

In this section we consider the predictions of two well known similarity-based models of perceptual categorization (i.e., prototype and exemplar models), using color categorization as a paradigmatic example. According to prototype models, the color space is divided into (more or less) universal categories anchored around particular focal hues (Berlin & Kay, 1969; Kay & McDaniel, 1978; MacLaury, 1991; Rosch, 1975; Rosch Heider, 1972; but see Roberson, Davies & Davidoff, 2000). For example, there tends to be greater consensus about the best example of BLUE (i.e., focal blue) than about the boundary between BLUE and PURPLE. The categorization of any hue, then, is based on its similarity to the focal hues of the different color categories (Bornstein & Korda, 1984; Rosch, 1975; Rosch Heider, 1972). In order to decide if a given hue is BLUE or PURPLE, a decision criterion is defined on the color spectrum between those categories, and judgments are based on which side of this boundary the hue falls.

According to exemplar models, a category is represented by storing exemplars as they are experienced, and categorization is a function of the similarity of the target instance to some set of those stored exemplars that are retrieved for comparison (W. K. Estes, 1994; Lamberts, 1995, 2000; Medin & Schaffer, 1978; Nosofsky, 1986; Nosofsky & Palmeri, 1997). So continuing our example, the target hue is compared to a set of blue exemplars and a set of purple exemplars, and the target is included in the category to which its summed similarity is greatest. Thus, prototype and exemplar models both posit that categorization is a function of the similarity between the target instance and the category representation.

Ordinarily it is clear to which category a target instance is most similar, and hence categorization is simple and unambiguous. However, because category boundaries are not explicitly represented in either prototype or exemplar models, different individuals may well set the decision criterion at different points on the dimension, and even a single individual may vary the placement of the criterion across occasions. Thus, due to the continuous nature of many perceptual categories, vagueness arises in borderline regions between those categories. For example, the color spectrum gives rise to borderline stimuli when a target hue is equally similar to the BLUE and PURPLE categories. In such cases, judgments of the hue should be randomly divided between the two color categories (see Bornstein & Korda, 1984; Rosch, 1975; Rosch Heider, 1972).

Our experiments investigate the hypothesis that the categorization of borderline stimuli is not wholly based on similarity to stored category representations. Rather, some evidence indicates that participants tend to adopt a strategy of inter-stimulus comparison: When categorizing a borderline stimulus, the indecision produced by its equal similarity to the competing categories is resolved via comparison to other recent stimuli (Stewart et al., 2002). This inter-stimulus comparison results in systematic deviations in borderline categorization.

Notice that, because prototype and exemplar models were developed to explain the categorization of individual stimuli, neither model specifies any mechanism by which inter-stimulus comparison would affect category decisions. However, these similarity-based models can be adapted to accommodate the effects of recent stimuli. If the prototype is a running average of previously categorized stimuli, then one can assume that the category representation will be biased toward weighting recent stimulus values more heavily. And likewise in an exemplar model, recent exemplars may be more likely to be retrieved as representative of the category, or may be more heavily weighted in the computation of similarity (Nosofsky & Palmeri, 1997). The result of this overweighting of recent stimuli is that presentation of any non-focal hue will cause the category representation to shift in the direction of that hue. For example, presentation of a hue midway between focal blue and the blue-purple boundary (call it stimulus S_1) will draw the center of the BLUE category slightly toward the PURPLE category, since S_1 has been given greater weight than other stimuli. In fact, the magnitude of this category shift will depend on the extremity of the stimulus, with more distant stimuli causing greater shifts in the category representation.

To consider the effect of inter-stimulus comparison, suppose that a hue on the category borderline between BLUE and PURPLE (stimulus S_2) is judged in the context of S_1. Because S_1 has drawn the BLUE category toward PURPLE, S_2 will now appear more similar to the BLUE category than it would have otherwise, and the categorization probability for S_2 therefore tips in favor of BLUE. Thus, in this case, similarity-based models would predict an *assimilation effect*, such that the target hue is more likely to be included in the category of the context hue.[1] Furthermore, given that the magnitude of the category shift is an increasing function of stimulus extremity, it follows that the likelihood of assimilating a borderline stimulus S_2 should increase as the context stimulus S_1 approaches that category boundary (providing of course that S_1 is itself still categorized in the target category). That is, if S_1 is a focal blue, it will not shift the category representation, and hence there should be no assimilation of S_2. But as S_1 approaches the BLUE-PURPLE boundary, the magnitude of the category shift increases, and therefore the likelihood of assimilating the borderline hue S_2 will also increase. To summarize, then, this adapted version of similarity-based models predicts an assimilation effect that increases as the context stimulus approaches the borderline stimulus.

The Similarity-Dissimilarity Model

An alternative account of inter-stimulus comparison has been proposed by Stewart and Brown (2005). Before describing their model, we first describe the *category contrast effect* that it was devised to explain. The basic finding is that the categorization of an atypical stimulus close to the category borderline is more accurate following a stimulus from the opposite category than following a stimulus from its own category. To demonstrate the phenomenon, Stewart et al. (2002) presented ten tones of varying pitch, with each tone perceptually equidistant from its neighbors, and they taught participants to categorize the five lowest (i.e., tones 1 to 5) into one category and the five highest (i.e., tones 6 to 10) into another. They found that tone 5 was categorized more accurately after tone 10 than after tone 1. And conversely, the categorization of tone 6 was more accurate following tone 1 than tone 10. In a second experiment they produced a similar result with visually-based categories of geometric stimuli. Thus, the categorization of a borderline stimulus was reliably contrasted from the category of the preceding stimulus. This is the opposite of what similarity-based models would predict.

Using a different perceptual domain, Friedenberg, Kanievsky and Kwasniak (2002) created ambiguous borderline face stimuli by morphing a male face and a female face, such that the morphed stimulus was a 50:50 male:female composite. Participants judged the apparent sex of those morphed faces, with judgment of each morphed face preceded immediately by either its original male face or its original female face. Friedenberg et al. also obtained a category contrast effect, such that the same morphed face was more likely to be judged FEMALE after presentation of its male face, and was more likely to be judged MALE after presentation of its female face.

To account for this category contrast effect, Stewart and Brown (2005) proposed an extension to the GCM – the Similarity-Dissimilarity GCM (SD-GCM). The novel aspect of the SD-GCM is its claim that the categorization of a stimulus is based not only on its similarity to the exemplars in category A, but also on its *dissimilarity* to the exemplars of category B. Effectively, when the target item is very similar to a category A exemplar, the similarity component will carry most weight, and one would expect assimilation (as in the similarity-based models). So if a context stimulus S_1 is very similar to the BLUE-PURPLE borderline stimulus S_2, and if S_1 is categorized as BLUE, then S_2 will also be called BLUE. However, if S_1 is quite different from S_2, then the dissimilarity component assumes more weight, and S_2 is more likely to be called PURPLE. Thus, the SD-GCM predicts assimilation when the stimuli are similar, and contrast when they are dissimilar. In intuitive terms, people may wish not to differentiate between very similar stimuli, and therefore will be biased to categorize them together. But when faced with a pair of stimuli that are quite different, people will be biased to place them into different categories.

In comparative judgments of this sort, the direction of difference (i.e., the sign) between S_1 and S_2 may constrain the response to S_2. To illustrate the concept of *sign*, consider three stimuli S_1, S_2 and S_3 that vary in order from least to most blue. In the context of S_1, S_2 has a positive sign, as S_2 is more blue than S_1. But in the context of S_3, S_2 has a negative sign, because S_2 is less blue than S_3. In certain cases, this sign information produces a *monotonicity constraint*: If S_1 is included in the target category, and if the sign of S_2 is positive, then S_2 should also be included in that category. For example, if S_1 is called BLUE, and S_2 is more blue than S_1, then by deduction S_2 must also be BLUE. In addition to this monotonicity constraint, Stewart et al. (2002) made further use of sign information. In fact, sign allows the SD-GCM to explain the contrast effect: If tone 1 belongs in category A, and tone 5 has a negative sign (i.e., is of higher pitch), then tone 5 is less likely to belong in category A.

Stewart and Brown (2004) evaluated two alternative versions of the SD-GCM. Both models use sign to predict category judgments. However, the models differ in their use of this sign information. In the sign-and-magnitude model, both the direction and the size of the difference between S_1 and S_2 affect the judgment of S_2. In Stewart et al.'s (2002) study, the borderline tone 5 has a negative sign in the context of both tones 1 and 3, but the magnitude of this difference is greater for tone 1 than for tone 3. Thus, by the sign-and-magnitude model, the judgment of tone 5 should exhibit a larger contrast effect when preceded by tone 1 than by tone 3. As an alternative, Stewart and Brown (2004; see also Laming, 1997) also tested a sign-only model, which incorporates the direction of difference between S_1 and S_2 but not the size of that difference. By the sign-only model, the judgment of tone 5 should exhibit a contrast effect of equivalent magnitude following tones 1 and 3, since it solely uses ordinal information to predict categorization.

The evidence for differentiating between the sign-and-magnitude and sign-only models is limited. Both Stewart et al. (2002) and Friedenberg et al. (2002) showed strong contrast effects between borderline stimuli and extreme category exemplars, but neither study manipulated the similarity of the context stimulus (S_1) to the borderline stimulus (S_2). For instance, Stewart and colleagues examined the judgment of tone 5 in the context of tone 1, but not in the context of tones 2, 3, or 4. As a result, neither study is capable of discriminating between the sign-and-magnitude and sign-only models. Stewart and Brown (2004) conducted a further experiment in which the borderline tone 5 was judged in the context of each of the other tones, and the category contrast effect was replicated. That is, the sign of the difference between stimuli affected the categorization of the borderline stimulus. However, Stewart and Brown found no evidence that the magnitude of the difference affected judgment—tones 1, 2, 3, and 4 produced equivalent contrast effects on the judgment of tone 5. Thus, the sign-only model received support. Despite this result, though, Stewart and Brown (2005) subsequently advanced the sign-and-magnitude model rather than the sign-only model. So it appears that before rejecting either model, further investigation is necessary.

The Present Research

We conducted a series of experiments that assessed the judgment of a borderline stimulus in the context of another stimulus of varying similarity either within or across the category boundary. The purpose of the experiments was to provide a more stringent and thorough test of the models of categorization described above. Our experiments differed in three important respects from the experiments of Stewart and Brown (2004; Stewart et al., 2002): We used familiar categories, without corrective feedback, and with simultaneous presentation of stimuli. These differences are discussed in turn below.

First, whereas Stewart and Brown (2004; Stewart et al., 2002) used arbitrary categories (i.e., tones and geometric shapes), we used familiar semantic categories (i.e., colors and animals). We suggest that the use of arbitrary categories makes inter-stimulus comparison more likely. Because participants have no prior experience with an arbitrary category, they are unable to resolve the category decisions by reference to representations in long-term memory, and hence they likely will be more reliant on comparison to other experimental stimuli. In the case of well learned semantic categories, though, long-term category representations may be used to resolve the category decision. For example, when judging whether a given hue is BLUE, the participant can easily access a stable, long-term representation of the category, and therefore need not compare that hue to other contextual stimuli. But when judging whether a particular tone belongs in category A, there is no long-term category representation, and hence a context stimulus is likely to be used as a cue to categorization. Thus, inter-stimulus comparison would seem less likely with familiar semantic categories than with arbitrary categories. In order to provide a more stringent test of inter-stimulus comparison effects, then, we used perceptual stimuli from common semantic categories.

Second, whereas Stewart and Brown (2004; Stewart et al., 2002) provided corrective feedback throughout the categorization task, our procedure did not include feedback. It seems likely that when the feedback for S_1 immediately precedes the presentation of S_2, that feedback will remain highly salient during the judgment of S_2. Thus, much like the use of arbitrary categories, the use of corrective feedback could also encourage inter-stimulus comparison. So again, we sought to generalize inter-stimulus effects by using a categorization task without feedback.

Third and perhaps most importantly, we used a method of simultaneous presentation, whereas Stewart and Brown (2004; Stewart et al., 2002; see also Friedenberg et al., 2002) used sequential presentation. This methodological difference is theoretically critical, as several investigations have shown that contrast is more likely to occur with sequential presentation than with simultaneous presentation of stimuli (Geiselman, Haight & Kimata, 1984; Wedell, Parducci & Geiselman, 1987; but see Wedell, 1995). For example, Wedell et al. (1987) found that a given face (S_2) was judged reliably more attractive when presented after an unattractive face (S_1). But when S_1 and S_2 were presented simultaneously, S_2 was then judged less attractive. Thus, Wedell et al. found successive contrast and simultaneous assimilation. To extrapolate, then, the category contrast effect obtained with sequential presentation could potentially be reversed with simultaneous presentation. Such a result would suggest an important modification of the SD-GCM. We therefore used simultaneous presentation in our experiments.

In summary, sequential presentation of stimuli from arbitrary perceptual categories with corrective feedback may be the optimal methodology for obtaining a contrast effect. This methodology was ideal for Stewart and Brown (2004; Stewart et al., 2002), as their intention was to demonstrate a novel phenomenon that cannot be explained by purely similarity-based models of categorization. The purpose of our experiments, though, was to provide a more challenging test of the SD-GCM in general, and also to discriminate between the sign-and-magnitude and the sign-only versions of that model. We therefore used simultaneous presentation of stimuli from familiar semantic categories without corrective feedback. If the contrast effect were to obtain under these drastically different experimental conditions, that would provide even stronger support for the use of dissimilarity in perceptual categorization. Friedenberg et al. (2002) also used common semantic categories (i.e., MALE and FEMALE), however they used the same sequential presentation paradigm as Stewart and Brown (2004). Moreover, because Friedenberg and colleagues did not manipulate the similarity between S_1 and S_2, their data cannot discriminate between the sign-and-magnitude and the sign-only versions of the SD-GCM.

Experiments 1 through 3 shared a common methodology, with only minor modifications. A calibration phase was followed by an experimental phase. In the calibration phase, hue patches were presented individually, and participants provided binary color judgments of the presented patch (e.g. “Is this square blue?”). Using data from the initial phase, the experimental phase was calibrated so as to center on each participant’s borderline hue. This calibration resulted in a set of hues ranging from clearly blue to clearly purple for that individual. The experimental phase consisted of simultaneous presentation of two hue patches, with participants again providing color judgments. All possible pairwise combinations of hues were presented several times each in the experimental phase.

We initially elicited disjunctive or conjunctive judgments of the hues. For instance, we asked participants whether *either* stimulus was blue (Experiment 1a), or whether *both* stimuli were purple (Experiment 1b). We then allowed independent assessment of the two hues by broadening the range of response options in Experiment 2: Participants were asked whether the left stimulus only, the right stimulus only, both stimuli, or neither stimulus was blue. Next, by introducing an expanded range of hues (BLUE-PURPLE-RED), Experiment 3 tested whether the inter-stimulus effect depends on the context item's degree of membership in the category, or on its similarity to the borderline stimulus. In Experiment 4 we tested whether there could have been a low-level perceptual color contrast effect operating in Experiments 1-3. Finally, Experiment 5 extended our results to other familiar semantic categories (i.e., CATS and DOGS).

Experiments 1a and 1b

Pairs of hues ranging from clearly blue to clearly purple were presented simultaneously, and participants judged whether *either* of the hues was BLUE (Experiment 1a), or whether *both* of the hues were PURPLE (Experiment 1b). Categorization of the borderline hue was the critical measure, with probability of a positive response serving as the dependent variable. Experiment 1b was a conceptual replication of Experiment 1a; judging whether either stimulus is blue is logically complementary to judging whether both stimuli are purple. Thus, the results of Experiment 1b should mirror those of Experiment 1a.

In the shorthand notation used to denote the various hues, we have centered on the borderline hue, designated *b*. The other six experimental hues radiate outward from the borderline, such that hues from the blue end of the spectrum are negative in denotation (relative to the borderline hue) and hues from the purple end of the spectrum are positive in denotation. Thus, the hue just inside the blue region of the borderline is designated hue *b-1*, while the hue just outside the blue region is designated hue *b+1*. And similarly, the bluest of the three blue hues is *b-3*, whereas the purplest hue is *b+3*. So the hues ranged from clearly blue (*b-3*) through the borderline (*b*) to clearly purple (*b+3*), with hue labels indicating distance and direction from the boundary between the BLUE and PURPLE categories.

If color categories are represented as prototypes (e.g., Rosch, 1975) or as collections of exemplars (e.g., Nosofsky, 1986), then judgment of a borderline hue should be assimilated to the category of a context hue (cf. Nosofsky & Palmeri, 1997), and this assimilation effect should increase as the context hue approaches the borderline hue. Alternatively, the SD-GCM (Stewart & Brown, 2005) predicts a contrast effect when the context hue is relatively dissimilar from the borderline hue. More specifically, the sign-and-magnitude model predicts greater contrast as the context and borderline hues become more dissimilar, whereas the sign-only model predicts a contrast effect that is impervious to the similarity of the context and borderline hues.

Method

Participants. All participants in each of the experiments reported below had normal color vision. Seventeen students of City University London participated in Experiment 1a, and 16 participated in Experiment 1b. All were paid £3 for participation.

Materials. Stimuli consisted of nine hues that systematically varied from clearly blue to clearly purple, as determined by the probability of categorization as BLUE in pilot testing (N = 32). The hue range was centered on the modal borderline hue from the pilot test. The Red-Green-Blue (henceforth "RGB") values for the hues, from blue to purple, were 15-5-55, 17-5-53, 19-5-51, 21-5-49, 23-5-47, 25-5-45, 27-5-43, 29-5-41, and 31-5-39 in Microsoft QuickBasic 4 programmed under DOS on an IBM-compatible PC with VGA graphics, where values for each scale range from 0 to 63. The stimuli were presented on a standard 8.5" x 11" color monitor with 640 x 480 pixels.[2]

Procedure. Because pilot testing revealed considerable variability across participants in the location of the blue-purple borderline, we used a two-stage procedure. In an initial calibration phase, participants were presented individual 4.45 cm (100 pixel) square patches of hue. Each hue patch appeared in either a left or a right location just above the vertical midpoint on the computer screen, alternating left and right locations on each successive trial. The left and right squares were separated by a gap of 8.9 cm (200 pixels). Viewing distance was not strictly controlled, but was approximately 50-60 cm, giving a horizontal visual angle of about 5 degrees for each square patch and 10 degrees for the gap between them. The background was dark gray (RGB = 10-10-10), and the target question (e.g., "Is the square BLUE?") appeared in white letters (RGB = 63-63-63) in the lower half of the screen. The target question was "Is the square BLUE?" for Experiment 1a and "Is the square PURPLE?" for Experiment 1b. The inter-trial interval consisted of a blank gray screen for a 1 sec. duration, and there was a further 1.5 sec. delay between stimulus onset and presentation of the target question. Responses were not accepted prior to onset of the target question, which required pressing the plus (+) key for “Yes” or the minus (-) key for “No”, followed by the return/enter key. The color patch and question remained on the screen until a valid response was made.

Each of the nine hues was singly and randomly presented eight times in the calibration phase, for a total of 72 calibration trials. The sequence was constrained so that the same hue was never presented on consecutive trials. At the end of this phase, the program used a regression algorithm to determine, for that participant, the hue for which probability of positive categorization was nearest to .50. In other words, the program estimated each participant’s borderline hue, or closest approximation to a borderline within the range of presented hues. This calibration procedure allowed us to focus on the judgment of each participant’s individuated borderline hue.

Seven of the nine hues from the calibration phase were automatically retained for use in a subsequent experimental phase, while two were excluded. The seven stimuli retained for each participant were that participant’s borderline hue, the three hues from the blue range closest to the borderline, and the three hues from the purple range closest to the borderline.[3] Once a participant’s range of seven experimental hues was determined, the experimental phase began. In this phase two hue patches were presented simultaneously on each trial. The hues occupied both the left and the right locations used in the calibration phase.

The only difference between Experiments 1a and 1b in the second phase was that in Experiment 1a participants were required to press the plus (+) key if either square was blue, or the minus (-) key if neither was blue, whereas in Experiment 1b participants were required to press the plus (+) key if both squares were purple, or the minus (-) key if either was not purple. After the (+) or (-) choice was accepted by pressing the return key, the next trial began.

There were four blocks of trials, each consisting of two trials of each possible pairwise combination of hues, such that in each block each possible pair of hues appeared once with one of the hue patches in the left location and once with that hue patch in the right location. Trials on which both patches were the same hue were also presented twice per block. Thus, each of the four blocks contained 56 trials, for a total of 224 experimental trials. Trials within each block appeared in random order, subject to the constraint that the same hue never appeared in two successive trials. The task lasted 20 to 30 minutes, with a self-paced break halfway through the experimental phase.

Results and Discussion

Full results of Experiments 1a and 1b are presented in Tables 1 and 2, respectively. In the tables, the target is the hue for which data are presented, and the context is the other hue that was present during the trial. So for instance, when hue *b* was presented alongside hue *b+3*, the proportion of trials on which either hue was judged BLUE is listed in Table 1 at the intersection of target *b* and context *b+3*. If context had no effect, then the choice proportions in each column would be randomly distributed around the same value. This pattern was not found, however. For the column headed *b* – the borderline target hue – in Table 1, the probabilities increased as context hues became increasingly purple, and those probabilities decreased in the corresponding column of Table 2. In interpreting these probabilities, we assumed monotonicity of judgments. That is, we assumed that a positive response in Experiment 1a (is either blue?) always indicated that (at least) the bluer of the two hues was blue, and correspondingly that a negative response in Experiment 1b (are both purple?) implied that (at least) the bluer of the two hues was not purple. To explain with the example from above, hue *b* is bluer than hue *b+3*. Therefore, a positive response in Experiment 1a (“Is either blue”) is informative of hue *b* (hence the term “target”), but nothing can be inferred about the judgment of hue *b+3* (hence the term "context"). For this reason the upper right half of each table is incomplete. (This assumption of monotonicity was supported in Experiments 2 and 3, where it could be tested directly.)

The results of interest are presented in Figure 1, which shows the proportions of positive categorization responses for the borderline hue in Experiments 1a and 1b. Three findings are apparent in the Figure. Notice first that the data from Experiment 1b mirrored those of Experiment 1a, indicating that results of 1b successfully replicated those of 1a. Second, note that both lines have a nonzero slope between context hues *b+1* and *b+3*. Thus, the presence of a context hue reliably affected categorization of the borderline hue. For context hues more than one step away from the borderline target, there was an increasing contrast effect, such that the more purple the context hue became, the more likely was the borderline hue to be categorized as blue (Experiment 1a) or as not purple (Experiment 1b). This contrast effect held regardless of the joint judgment (i.e., *either*, *both*) and regardless of the color category (i.e., BLUE, PURPLE). Finally, the *b+1* context hue had a slight tendency to assimilate the borderline target towards the purple category, so that the lines are not straight from *b* through *b+3*.

The positive categorization proportions for the borderline hue in Experiments 1a and 1b were initially submitted to two separate one-way repeated measures ANOVAs, each with four levels of Context Hue (i.e., *b*, *b+1*, *b+2*, *b+3*). Both analyses showed a significant main effect of Context Hue [F (3, 48) = 11.78, p < .001 in Experiment 1a, and F (3, 45) = 3.96, p = .01 in Experiment 1b]. Subsequently, the data from Experiments 1a and 1b were analyzed together. The data from Experiment 1b were reverse coded to determine their fit with the data from Experiment 1a. A 2 (Experiment: 1a, 1b) x 4 (Context Hue: *b*, *b+1*, *b+2*, *b+3*) repeated measures ANOVA, with Experiment as a between-participants factor, revealed that neither the main effect of Experiment nor the Experiment x Context Hue interaction was significant (both F < .5). Only the main effect of Context Hue was reliable, F (3, 93) = 13.69, p < .001. The data of Experiment 1a therefore essentially mirrored those of Experiment 1b. The main effect of Context Hue indicates that the presence of a context hue reliably affected the judgment of the borderline hue. A polynomial set of contrasts revealed significant linearity, F (1, 31) = 20.30, p < .001. The quadratic and cubic trends were also reliable, F(1, 31) = 6.35, and F(1,31) = 6.46, both *p* < .05, reflecting the changes in slope from *b* through *b+3*, as evident in Figure 1.

One further question of interest is whether assimilation occurred when the borderline was presented in the context of the *b+1* hue, relative to *b* paired with itself. Recall that the SD-GCM model predicts assimilation for very similar pairs. Figure 1 suggests that this might have been the case. The difference between [*b*|*b*] and [*b*|*b+1*] did not reach significance in either experiment alone (both paired samples t < 1.10), nor when the data from Experiments 1a and 1b were pooled (t (32) = 1.46). However examination of Table 1 shows that in fact for targets other than *b*, there was also a systematic tendency for assimilation when the context hue was just one step away. A t-test comparing the sum of positive responses to all target stimuli, either in the context of an identical hue, or in the context of a hue that was one step more purple, gave a significant assimilation effect for Experiment 1a (t(16) = 2.33, p = .03), but not for Experiment 1b (t < .5), nor when they were pooled.

To summarize, inter-stimulus comparison was manifest as a contrast effect: The probability of positive categorization of the borderline hue varied systematically as a function of its similarity to the context hue. The more purple the context hue, the more likely was the borderline to be judged BLUE (Experiment 1a), and the less likely was the borderline to be judged PURPLE (Experiment 1b). The size of this effect was quite substantial, as the probabilities of positive categorization ranged across different context conditions from .28 to .54 in Experiment 1a and from .30 to .51 in Experiment 1b. This effect suggests that people are influenced by the difference between the hues when categorizing colors (Stewart & Brown, 2005). Moreover, the magnitude of this difference also appears to affect judgment, with greater contrast as the dissimilarity between the context and borderline hues increased. Thus, unlike the experiment by Stewart and Brown (2004), the present experiment supported the sign-and-magnitude version of the SD-GCM.

Experiment 2

Experiment 1 demonstrated that the dissimilarity between two hues affects their categorization in a continuous (i.e., sign-and-magnitude) manner. However, because that experiment elicited joint judgments of the two hues (i.e., Is *either* square blue? Are *both* squares purple?), monotonicity of responses had to be assumed. In order to corroborate the support for the sign-and-magnitude model, while also directly testing for monotonicity, in the present experiment we asked participants to indicate whether the *left square only*, the *right square only*, *both squares,* or *neither square* was blue. Thus, each judgment was informative of both hues per trial, and monotonicity could be observed rather than assumed. Notice also that, although the two judgments must be combined to provide a single response, context and target hues not need be judged together in this procedure. Hence the finding of a dissimilarity-based contrast effect here would provide stronger evidence for the SD-GCM.

Method

Sixteen students of City University London were paid £3 for participation. Materials were identical to those of Experiment 1. The procedure (calibration phase followed by experimental phase) was also identical to Experiment 1, with the exception that in the experimental phase participants were required to press the left arrow (←) key if only the left square was blue, the right arrow (→) key if only the right square was blue, the up arrow (↑) key if both squares were blue, or the down arrow (↓) key if neither square was blue. These four response options were displayed schematically on the screen as follows:

BOTH

LEFT RIGHT

NEITHER

Pressing one of the four arrow keys highlighted its corresponding response on the screen (i.e., Left, Right, Both, Neither). After the highlighted choice was accepted by pressing the return key, the next trial began. All other procedures were identical to Experiment 1.

Results and Discussion

Each response was analyzed to provide independent categorization responses for the left and right stimulus respectively. For example, the stimulus displayed on the left was scored as “blue” if either the LEFT or the BOTH response option was chosen, and as “not blue” otherwise. Categorization of each stimulus was then scored as a function of the other stimulus with which it was presented. This allowed examination of the monotonicity constraint. Of the 2,688 trials in the experiment (across participants), only four responses violated the monotonicity constraint. That is, only very rarely did participants judge the bluer of two hues as PURPLE and the purpler hue as BLUE. Clearly participants were extremely adept at using monotonicity to provide consistent responses on each trial of the experiment. In fact, the overwhelming monotonicity of responses suggests that inter-stimulus comparison again occurred in this categorization task. This finding also supports our assumption that responses in Experiment 1 were largely monotonic as well.[4]

Figure 2 plots the observed probability of the borderline hue being categorized as BLUE in the context of the other hues. As in Experiment 1a, the probability of categorizing the borderline hue as BLUE increased as the context hue varied from blue to purple. Thus, the contrast effect was not limited to conjunctive or disjunctive response combinations, and the sign-and-magnitude model was once again supported.

Full results are presented in Table 3. In the table, the target is the hue for which data are presented, and the context is the other hue that was present during the trial. For example, when hue *b+3* was presented with hue *b*, the probability that hue *b* was judged BLUE is listed in Table 3 at the intersection of target *b* and context *b+3*, while the probability that hue *b+3* was judged BLUE is listed at the intersection of target *b+3* and context *b*.

A one-way repeated measures ANOVA was performed on the probabilities of positive judgment of the borderline hue. There were seven levels in the analysis, one level for each of the seven possible context hues with which the borderline was presented (including itself). A reliable effect was detected, F (6, 90) = 10.63, p < .001. The effect was captured by a significant linear trend in a subsequent polynomial contrast, F (1, 15) = 16.77, p < .001. There was also a significant 5th order trend, F (1, 15) = 5.72, p < .05, reflecting the four points of inflection in the curve, at hues *b-2*, *b-1*, *b+1* and *b+2*. This significant fit to a 5th order polynomial function neatly captures the boundary conditions of the contrast effect (see Figure 2). It suggests that the contrast effect asymptoted from *b-3* to *b-2* and from *b+2* to *b+3*, and that, as in Experiment 1, hues *b-1* and *b+1* produced only minimal changes in categorization of the borderline hue.

In corroboration of Experiment 1, the results of Experiment 2 support the sign-and-magnitude SD-GCM (Stewart & Brown, 2005), which predicts increasing contrast as the context and target stimuli become more dissimilar. No assimilation effect was apparent, although the slope of the context effect did appear to flatten when the stimuli were only one step apart. The observation of an asymptote for the contrast effect at the two ends of the scale also makes good sense in terms of the SD-GCM. In the model, dissimilarity is just (1-similarity), where similarity decreases exponentially with distance. Hence once would expect dissimilarity to asymptote after a certain point, and further increases in distance between context and target should have little further effect.

Experiment 3

We have argued that the context hue is used as an anchor such that relative distance or closeness to the context influences categorization (Stewart & Brown, 2005). However all of our results are also consistent with a simpler hypothesis, namely that people have a bias toward placing simultaneously presented hues in different categories unless they are very similar to each other. This bias could result from a pragmatic strategy to use names in a distinctive manner, and may not reflect specific comparative judgment effects on the categorization decision. One simple test of this account would be to provide context items that are as likely to be in the same or different categories as before, but where the similarity between context and target is low. This manipulation can be achieved by using context hues that approach the same category (PURPLE) as the target, but from the other side of the color spectrum. Thus a blue-purple borderline hue can be presented with a hue that is on the red-purple portion of the color dimension. If there is a pragmatic bias to respond with different category names, then as the red-purple context item becomes purpler, so the likelihood of calling the blue-purple borderline target item PURPLE should decrease. If on the other hand the contrast effect is owing to comparison processes that only affect items within the same borderline range, then this manipulation should have no effect.

The current experiment included a larger range of stimuli spanning not just the blue-purple continuum, but also the purple-red hue continuum. Here, as in Experiment 1b, the target category was PURPLE. The expanded range of hues provided two borderline hues (blue-purple and purple-red), and permitted evaluation of the effect of presentation of one hue range as context on categorization of the other as target. If the contrast effect were sensitive to the relevance of the context stimulus for judgment of the target stimulus (Petzold & Haubensak, 2004; see also LeBoeuf & Estes, 2004; Mussweiler, 2003; Stapel & Winkielman, 1998), then context hues from the *purple-red* range should *not* produce the similarity-based contrast effect on the *blue-purple* borderline hue, and vice versa, since different reddish hues will be too far removed from the target to provide any useful information for disambiguating the category boundary between blue and purple. The SD-GCM (Stewart & Brown, 2005) would not predict any additional effect of a context hue from the other side of the target category, since beyond a certain distance, dissimilarity will asymptote, and further increases in the distance of the context item will have no effect. So the dissimilarity-based contrast effect should be replicated when the context and target hues are from the same hue range, but when the context and target are from opposite hue ranges there should be no effect.

Method

Twenty students of City University-London were paid £3 for participation. The materials were augmented with a range of nine hues from the purple-red continuum. Their RGB values, from purple to red, were 34-5-36, 37-5-33, 40-5-30, 43-5-27, 46-5-24, 49-5-21, 52-5-18, 55-5-15 and 60-5-5, respectively. The procedure of the calibration remained the same, except that its length was doubled due to the addition of the nine new hues. Thus, each of the four blocks included each of the 18 hues twice (once in each location), for a total of 144 trials in the calibration phase. At the conclusion of the calibration phase, one blue-purple borderline hue and one purple-red borderline hue were determined for each participant. In order to limit the duration of the experiment, five (rather than seven) hues were selected from each hue continuum, for a total of 10 experimental hues. This limitation was expected to effectively remove the apparent asymptotes observed in Experiment 2, while preserving the linear and/or cubic aspect of the contrast effect. Also for the sake of time considerations, the inter-trial interval was reduced from 1 sec. to 800 msec., and the delay between stimulus onset and onset of the target question (e.g., which of these are PURPLE?) was reduced from 1.5 to 1 sec.

In the experimental phase, all possible pairs of hues were presented twice in each of the four blocks (with location counterbalanced) in a random order. This resulted in 110 trials per block, 440 trials total. In all other respects, the procedure was identical to Experiment 2. With the changes introduced in the present experiment, its duration was approximately the same as the preceding experiments.

Results and Discussion

As in Experiment 2, the monotonicity constraint was overwhelming observed in the present experiment as well. Of 3,200 trials in the experiment, only four violated monotonicity. This provides preliminary evidence that inter-stimulus comparison occurred.

Proportions of PURPLE category judgments are displayed in Table 4, which presents judgments of the two hue ranges separately. To summarize, the dissimilarity-based contrast effect was found for both borderline hues when paired with context hues from within the same hue range, but was *not* found when paired with context hues from the other hue range. For instance, judgment of the blue-purple borderline hue exhibited a contrast effect when that borderline was presented in the context of hues from the blue-purple range. But categorization of that blue-purple borderline hue was unaffected by context hues from the purple-red range (see Figure 4). Thus, the contrast effect appears to be sensitive to the relevance of the context stimulus for disambiguating the borderline stimulus (LeBoeuf & Estes, 2004; Mussweiler, 2003; Petzold & Haubensak, 2004; Stapel & Winkielman, 1998). When the context hue was from a different range, distance between target and context was uniformly large, and so bringing the reddish context hue closer to purple had no observable effect on classification of the blue-purple borderline color.

The borderline categorization data were tested with a 2 (Borderline: blue, red) x 2 (Context Range: same, different) x 5 (Context Hue: *b-2, b-1, b, b+1, b+2*) repeated measures ANOVA. To begin with, the main effect of Borderline was nonsignificant, p > .75, indicating that judgments of the blue borderline and the red borderline did not differ overall. The effect of Context Hue was significant, F (4, 76) = 8.55, p < .01, as it was in the previous experiments. The effect of Context Range was also significant, F (1, 19) = 15.74, p < .01. These main effects, however, were overshadowed by a significant interaction of Context Range and Context Hue, F (4, 76) = 6.95, p < .01. As apparent in Figure 3, the two hue ranges (i.e., same or different range) exerted differential effects on the judgment of the borderline hues. In particular, context hues from the same range as the target hue produced the comparison-based contrast effect found in Experiments 1 and 2, but context hues from the different range did not produce the comparison-based contrast effect.[5]

Context Range also interacted with Borderline, F (1, 19) = 9.86, p < .01. Close inspection of Table 4 shows that the blue-purple borderline and the red-purple borderline were affected differently when paired with hues from the other hue range. Specifically, targets on the blue-purple borderline were rendered less purple by presentation of any of the context hues from the red-purple range, whereas categorization of targets on the red-purple borderline was unaffected by presentation of context hues from the blue-purple set. It is possible that the interaction reflects the different clarity of blue and red hues as displayed on our equipment. We can offer no other explanation for this difference. Neither the interaction of Borderline and Hue (p > .20) nor the three-way interaction (p > .50) was reliable.

In order to explicate the interaction of Context Range and Context Hue more clearly, two separate ANOVAs were conducted—one on the data from the different range only, and one on the data from the same range. First, a 2 (Borderline: blue, red) x 5 (Context Hue: *b-2, b-1, b, b+1, b+2*) ANOVA was performed on probabilities of positive categorization of the borderline hue when paired with context hues from the *different* range. This corresponds to the dashed line in Figure 4. The effect of Context Hue did not approach significance, F (4, 76) = .90, p = .47. The 2 x 5 ANOVA on data from the *same* hue range (i.e., the solid line in Figure 4), however, did verify a significant effect of Context Hue, F (4, 76) = 12.69, p < .01. A polynomial contrast revealed a significant linear trend, F (1, 19) = 20.66, p < .01. There was also a significant fit to a cubic function, $F(1, 19) = 7.69$, $p = .01$, indicating that hues just inside or outside the category boundary had little effect on categorization of the borderline hue, just as in the previous experiments. These results again support the sign-and-magnitude model, which predicts a linear contrast effect when the hues are from the same hue range, but predicts an asymptote when they are from the opposite hue range.

Experiment 4

The preceding experiments showed a marked contrast effect whereby the categorization of a borderline hue was shifted away from the category in which the context hue was located. We have argued that this dissimilarity-based contrast effect occurs at a decision-making stage and results from the influence of the context stimulus on categorization of the target stimulus (Stewart and Brown, 2004). However, the observed contrast effect could instead involve more low-level perceptual mechanisms, reflecting the well established color contrast effect (see e.g., Abramov & Gordon, 1994; Jameson & Hurvich, 1989 for review). Thus, the purpose of Experiment 4 was to test whether the effect obtained in the preceding experiments may be attributable to a low-level perceptual contrast. To this end, we devised a paradigm to replicate the standard perceptual contrast effect, and simultaneously observe whether it can explain the result of Experiments 1 through 3.

In a typical demonstration of perceptual color contrast, a smaller target patch is embedded within a larger context patch (e.g., Jameson & Hurvich, 1959). Judgment of the target patch contrasts from the context patch. In the present case, for instance, a blue-purple borderline target would appear more blue when surrounded by a purple context, and would appear more purple in a blue context. Such perceptual contrast might account for the effect obtained in the preceding experiments. However, an important difference between this perceptual contrast paradigm and that used in the previous experiments is the presentation, or relative location, of the context and target hues. Specifically, in the preceding experiments we intentionally presented relatively small hue patches with a substantial separation between them, so as to minimize the likelihood of perceptual contrast effects. Nonetheless, the possibility of perceptual contrast remains, even with the separated hue presentation used in those previous experiments.

In Experiment 4 we used both an embedded presentation (cf. Jameson & Hurvich, 1959) and a separated presentation (as in Experiments 1 – 3) of context and target hues. We sought to replicate the standard perceptual contrast effect in the embedded condition, and also to observe whether the separated presentation of Experiments 1 through 3 is susceptible to the same perceptual contrast effect. If so, then the results of the preceding experiments may simply demonstrate perceptual contrast.

To demonstrate perceptual contrast, we used a change detection task in which participants had to detect whether the target patch changed in hue or remained constant. While observing the target patch for a hue change, the context patch also changed step by step across the full hue range from blue to purple. That is, the context patch changed hue on every trial (from clearly blue to clearly purple, or vice versa), the target patch changed hue on half the trials (from slightly blue to slightly purple, or vice versa), and participants indicated whether or not the target patch had changed colors. Critically, on trials in which the target patch did change hues, that change was very small, and the target hue change was either in the *same direction* as the context hue change (i.e., both from blue to purple, or both from purple to blue), or in the *opposite direction* to the context hue change (i.e., one from blue to purple, and the other from purple to blue). Thus, the experiment employed a 2 (Presentation: separated, embedded) x 2 (Hue Change Direction: same, different) design.

Note that this change detection task does not require participants to classify the target hue. In this task, perceptual contrast would be manifest as more accurate detection of a change in target hue when the context hue changed in the opposite direction, as the opposing change in the context hue would make the target hue change more apparent (see Abramov & Gordon, 1994; Jameson & Hurvich, 1989). Thus, if the result of the preceding experiments were attributable to perceptual contrast, then Hue Change Direction should affect accuracy in both the embedded and the separated presentation conditions (i.e., a main effect). Alternatively, we predicted an interaction, such that Hue Change Direction should affect accuracy in the embedded condition, but not in the separated condition. Failure to obtain contrast in the separated presentation condition would strongly suggest that perceptual contrast was not the principle explanation of the results of the preceding experiments.

Method

Participants. Sixteen undergraduates at City University-London participated for course credit.

Design. The factors of interest were Presentation (separated, embedded; within-participants) and Hue Change Direction (same, different; within-participants). In the separated presentation condition, the context and target hues were presented exactly as in Experiments 1 through 3, whereas in the embedded presentation condition, the target hue was embedded in a larger, fully surrounding, context hue (cf. Jameson & Hurvich, 1959). This factor was blocked and counterbalanced, with half the participants completing the separated presentation condition prior to the embedded condition, and half completing the two conditions in the reverse order. There were three blocks of 16 trials for each presentation condition, for a total of 96 trials. Within each block of 16 trials there were four same direction trials, four different direction trials, and eight constant (no change) trials. Order of trials within blocks was randomized for each participant. Additional control factors, both counterbalanced within blocks, were the direction of the context hue change (blue to purple or purple to blue) and, in the separated presentation condition, the location of the target (left or right position).

Stimuli. The experiment was conducted on the same hardware and software used in the previous experiments. The separated presentation condition used the same on-screen display as in Experiments 1 – 3, within a 640 x 480 pixel display. The background was a dark gray (10-10-10 RGB). The embedded presentation condition used a larger rectangle for the context hue, with height 250 pixels and width 300 pixels, situated with its lower left corner 200 pixels up and 230 pixels to the right from the bottom corner of the screen. The target hue was displayed in a 100-pixel square exactly centered within this larger rectangle. A dark gray border, one pixel in width, surrounded the target square to separate it from the context hue.

Fifteen hues (labeled #1 to #15) were constructed with RGB values starting at 16-5-54 and progressing by single unit steps of increasing red and decreasing blue to 30-5-40. In the constant (no change) filler trials, the target hue was always in the center of the range (#8, 23-5-47 RGB), presumably on or near the borderline between BLUE and PURPLE. In the experimental trials, five successive presentations of hue #7, five presentations of #8, and five presentations of #9 occurred in either blocked ascending order (i.e., 77777 → 88888 → 99999) for the blue-to-purple targets, or blocked descending order (i.e., 99999 → 88888 → 77777) for the purple-to-blue targets.

Procedure. For those participants who completed the separated presentation condition first, the following instructions were displayed in white font on a black background: “Welcome to the experiment. On each trial of the experiment you will first see a cue of some dotted lines on the left or the right of the screen. Press a key to see how this will look…[demonstration of cue]…Shortly afterwards you will see two coloured squares on the screen. Watch the one that is on the side that has just been cued. The squares will start to flash and the one you are not watching will start to change colour – from blue to purple or from purple to blue. Watch the cued square and see if you can tell whether it also changes colour from start to finish. When it changes it is only a small change. If it changes it gets slightly more blue or slightly more purple. After a short while, a question will appear on the screen, asking you whether the square you were watching changed colour or not. When you have answered, then another trial will start." A practice trial followed, and if the participant had no further questions, the experiment proper began.

For the embedded condition, the instructions were duly modified to refer to a small target square inside a larger one. Participants were told that the larger one would change color and that they should watch the central square to see if it also changed color or not. Appropriate instructions were given at the start of each half of the experiment.

On each trial of the separated presentation condition, a cue (i.e., "== Watch this one ==") was displayed either to left or right of center. The cue was presented in white on black, and was located immediately below the position where the target color patch was to appear. The cue remained on the screen for three seconds. The two color patches then appeared, one on each side of the screen, exactly as in the previous experiments. A sequence of 15 pairs of patches was presented, with each pair appearing for 300 ms, separated by a blank screen displayed for 30 ms. Across the sequence, the context hue changed in equal RGB increments from blue to purple (#1 to #15) or from purple to blue (#15 to #1), according to the trial. Simultaneously, the target hue either remained constant (50% of trials), changed slightly in the same direction as the context (25% of trials), or changed slightly in the opposite direction from the context (25% of trials). At the end of the trial, the display was blanked, and a question box appeared at the bottom of the display asking, "Did the target square change colour at all?" Participants pressed the <+> key for yes or the < – > key for no, and the words "Yes" or "No" (respectively) appeared in the box. Participants then pressed <Return> to confirm their choice. The procedure of the embedded presentation condition was identical, except that the target location cue (left or right) did not occur, since the target always appeared embedded within the context hue.

Results and Discussion

Percentage of correct responses served as the dependent measure. Results are illustrated in Figure 4. To summarize, the embedded presentation condition produced a perceptual color contrast effect, which was evidenced by greater detection of a target hue change when the context hue simultaneously changed in the opposite direction than when the context hue changed in the same direction as the target. Now, if the contrast effect observed in Experiments 1 through 3 were also attributable to perceptual contrast, then the separated presentation condition should also exhibit greater change detection in the Different Direction condition than in the Same Direction condition. However, this did not obtain; rather, the direction of change of the context hue exerted no effect on detection of target hue change in the separated presentation condition. Thus, we replicated perceptual contrast in the embedded condition, and demonstrated that such perceptual contrast provides an insufficient explanation of results from the separated presentation. The similarity-based contrast effect obtained in Experiments 1 through 3 appears to be attributable to a higher level cognitive process (i.e., context-target comparison) rather than a lower level perceptual process.

The constant (no change) filler trials allow an initial evaluation of relative task difficulty. Unsurprisingly, change detection was less accurate in the embedded condition (M = 51%, SE = 4%) than in the separated condition (M = 81%, SE = 4%), t (15) = 9.46, p < .001. In fact, accuracy did not differ from chance (50%) in the embedded condition for these constant trials, p = .89. The experimental conditions are analyzed below.

A 2 (Presentation: separated, embedded) x 2 (Hue Change Direction: same, different) repeated measures ANOVA compared change detection accuracy in the four experimental conditions. Both main effects and their interaction were reliable. The main effect of Presentation, F (1, 15) = 9.15, p < .01, indicates once again that the separated condition yielded better accuracy overall than the embedded condition. This result corroborates that found in the no-change filler trials reported above. The main effect of Hue Change Direction, F (1, 15) = 22.33, $p < .001$, confirms that the target hue change was more likely to be detected if the context hue changed in the opposite direction than in the same direction. This result indicates reliable perceptual contrast (cf. Abramov & Gordon, 1994; Jameson & Hurvich, 1989). Most importantly, the interaction [F (1, 15) = 15.48, $p < .001$] reveals that the direction of change of the context hue affected detection of the target hue change in the embedded condition, but not in the separated condition. Planned comparisons supported this conclusion: Accuracy for embedded targets differed between the same direction (M = 45%, SE = 6%) and different direction conditions (M = 78%, SE = 4%), t (15) = 4.75, $p < .001$, whereas accuracy for separated targets was identical across directions of change (both M = 72%, SE = 4%). There was thus no evidence that perceptual contrast, demonstrated here in the embedded condition, occurred with the separated presentation of Experiments 1-3.

There are of course other differences in procedure between Experiment 4 and the preceding experiments. In particular, whereas in the earlier experiments the participants' eye gaze would be directed at each of the squares in turn, in Experiment 4 it would most likely be focused on the target square only. Determining the degree to which one might expect low-level color contrast to act in our stimulus display is itself problematic, given the general complexity of the phenomenon (Whittle, 2003). Experiment 4 thus serves as a useful control, although it cannot be said to definitively rule out effects of retinal contrast coding. In order to confirm the generality of the contrast effect observed in the preceding experiments, and in view of the possibly unknown effects of contrast coding of colors, we therefore chose to switch to a different stimulus domain for our final experiment.

Experiment 5

If the dissimilarity-based contrast effect is not a low-level effect of color vision, as suggested by Experiment 4, then it ought to obtain with other stimulus domains as well. Thus, in Experiment 5 we replicated Experiment 2, but with CAT and DOG instead of BLUE and PURPLE. That is, we presented photos of a cat, a dog, and several intermediate morphed images of that cat and that dog (see Figure 5). As in Experiment 2, stimuli were presented in pairs, and participants judged whether the left image only, the right image only, both, or neither was a member of the target category (CATS). The present experiment therefore served the dual purposes of (1) confirming that the contrast effect is not due to perceptual color contrast, and (2) extending the contrast effect to a different semantic domain.

Method

Twenty-four undergraduates at the University of Georgia participated for partial course credit. One participant gave “no” responses to every pair of stimuli, and was excluded from the analysis. Stimuli consisted of seven images: One photo of a cat (*CD1*), one perceptually similar photo of a dog (*CD7*), and five intermediate morphed images approximately equidistant across the morphing process (*CD2 – CD6*). Each of the seven images was presented with every other image (including itself) twice, with location (left, right) counterbalanced, for a total of 56 trials. Stimuli were presented as in Experiment 2. The procedure was also the same as that of Experiment 2, except that there was no calibration phase in the present experiment.

Results and Discussion

Full results are reported in Table 5 and illustrated in Figure 6. Although the overall rate of monotonicity violations (1.3%) was higher than in Experiments 2 and 3, participants again consistently observed the monotonicity constraint in their responding. Only two of the 23 participants violated monotonicity—that is, gave a response in which the more cat-like picture was called a DOG, and the more dog-like picture was called a CAT—on multiple trials.

The dissimilarity-based contrast effect that was obtained with color categories in Experiments 1 through 3 was replicated in Experiment 5. The finding that this contrast effect extends to other conceptual domains is strong evidence that the results of Experiments 1 – 3 were not purely attributable to perceptual color contrast. Instead, the contrast effect appears to be a general phenomenon of comparative perceptual judgment, regardless of category domain (see also Friedenberg et al., 2002; Stewart et al., 2002).

Because the present experiment lacked a calibration procedure to determine each participant's borderline image, there was no *a priori* borderline stimulus, nor any consensual borderline. That is, different images emerged as the borderline for different participants. For instance, although *CD3* appeared to be the borderline stimulus overall (with a .51 response rate), only 9 of the 24 participants had *CD3* as the closest to the borderline. Data were therefore analyzed first by group, collapsing across individual differences in the location of the category borderline, and then by individuals, adjusting for differences in the location of the borderline.

Group analysis. Figure 6 illustrates a vertical spread of the seven target images, indicating that the different targets exhibited different probabilities of positive categorization. More importantly, the context images did indeed tend to affect judgment of the target images, as the slopes are generally positive. For each target image, we conducted a one-way ANOVA with seven levels (i.e., *CD1, CD2, CD3, CD4, CD5, CD6, CD7*). The effect of Context was reliable for targets *CD1* [$F(6, 132) = 3.21$, $p = .006$], *CD2* [$F(6, 132) = 3.04$, $p = .008$], *CD3* [$F(6, 132) = 4.39$, $p < .001$], *CD4* [$F(6, 132) = 3.27$, $p = .005$], *CD5* [$F(6, 132) = 2.59$, $p = .021$], and *CD6* [$F(6, 132) = 3.06$, $p = .008$], but not *CD7* [$F(6, 132) = .34$, $p = .912$]. Furthermore, consistent with the previous experiments, the effect of the context images was significantly linear for most of the targets (i.e., *CD1* through *CD5*, all $p < .01$ by polynomial contrast). That is, as the context image became more dog-like (i.e., *CD1* → *CD7*), the target image was more likely to be positively categorized as a CAT.

Individual analysis. In the following analysis, we located for each individual the image that was closest to having 50% positive category judgments, and then examined the effect of the context images on the categorization of that borderline image (cf. Experiments 1 through 3). To locate each participant's borderline, we calculated the mean probability of positive categorization for each target across all seven contexts. The target for which that mean probability was closest to .50 was defined as that participant's borderline (*b*). Once each participant's borderline was determined, we then examined the probability of positive categorization of that borderline target in the context of the two nearest within-category images (*b-2* and *b-1*), in the context of itself (*b*), and in the context of the two nearest images outside the category boundary (*b+1* and *b+2*).[6] This procedure therefore adjusted for individual differences in the location of the category borderline, and was essentially a *post hoc* analog of the calibration procedure of the preceding experiments.

Data were submitted to a one-way ANOVA with five levels, one for each context image (i.e., *b-2, b-1, b, b+1, b+2*). The effect of Context was significant, F (4, 88) = 3.02, $p < .05$, and a polynomial set of contrasts revealed a reliable cubic trend, F (1, 22) = 4.59, $p < .05$. As in the earlier experiments, there was no contrast effect when the context picture was only one step removed from the target, but increased in strength from one step onwards. Individual correlations were also calculated between categorization probability and context stimulus number for each target and each individual, where the data permitted. Of the 23 participants, 21 showed positive contrast effects, with the categorization probability of targets as CATS increasing on average with the dogginess of the context stimulus.

In summary, the group analysis and the individual analysis converged on the conclusion that the dissimilarity-based contrast effect extends to perceptual stimuli from familiar semantic categories. More specifically, the results support the sign-and-magnitude model (Stewart & Brown, 2005), as the size of the contrast effect was an increasing function of the dissimilarity between the context and target stimuli.

General Discussion

The experiments reported above demonstrate that the judgment of a borderline stimulus is systematically influenced by its similarity (or dissimilarity) to a simultaneously presented context stimulus. Specifically, as the context and target stimuli became less similar, the probability of including the borderline stimulus in the category of the context stimulus decreased. This effect was remarkably large and robust. It was obtained with colors (BLUE and PURPLE) and animals (CATS and DOGS), and it occurred only when the context stimulus was relevant for the disambiguation of the borderline stimulus. Results supported the SD-GCM in general, and the sign-and-magnitude model in particular (Stewart & Brown, 2005).

This contrast effect has previously been investigated under different conditions. Most notably, Stewart and Brown (2004; Stewart et al., 2002) illustrated the contrast effect using arbitrary categories (i.e., tones and geometric figures) with sequential presentation (see also Friedenberg et al., 2002). However, in the introduction we suggested that the use of arbitrary categories is likely to induce inter-stimulus comparison, and furthermore, that contrast effects are more likely to occur with sequential presentation than with simultaneous presentation. Thus, the prior studies have used a methodology that is relatively likely to produce a contrast effect. In order to provide a more stringent test of the contrast effect, then, we instead used familiar semantic categories with simultaneous presentation. Familiar categories presumably have long-term representations that can be accessed to resolve the category decision (e.g., Nosofsky, 1986; Rosch, 1975), and therefore inter-stimulus comparison should be less likely. Moreover, assimilation effects are more often observed with simultaneous presentation than with sequential presentation (Geiselman et al., 1984; Wedell et al., 1987). Thus, the present demonstration of contrast with simultaneous presentation of familiar categories may provide the strongest support to date for Stewart and Brown’s (2005) SD-GCM.

The present experiments also provide a more precise investigation of the relationship between inter-stimulus similarity and contrast. Because Friedenberg et al. (2002) and Stewart et al. (2002) did not manipulate the similarity of the context and target stimuli, those studies could not discriminate between the sign-only and the sign-and-magnitude models of the SD-GCM. Stewart and Brown (2004) did manipulate inter-stimulus similarity, and they found that the magnitude of the dissimilarity between context and target stimuli had no effect. That is, the size of the contrast effect was the same regardless of whether the context and target stimuli were extremely dissimilar or only slightly dissimilar. Thus Stewart and Brown (2004; see also Laming, 1997) supported the sign-only model. But in the present experiments, we repeatedly found that the size of the contrast effect was an increasing function of the dissimilarity between context and target hues (see Figures 1, 2, 3, and 6). This consistent linearity supports the sign-and-magnitude model instead.

Our finding of a linear effect of context-target dissimilarity, together with Stewart and Brown's (2004) lack of linearity, suggests that the use or disuse of relative magnitude information in categorization tasks may be contingent upon stimulus presentation conditions. It could simply be that magnitude information is more likely to be used with simultaneous presentation: When both stimuli are present simultaneously, the magnitude of their difference is readily available (see Wedell, 1995). But when stimuli are presented sequentially, their magnitude of difference cannot be directly compared, and hence it may be less available. So extrapolating from the present results, any general model of perceptual categorization must incorporate the magnitude of the inter-stimulus dissimilarity.

The present experiments provided very little evidence of assimilation. In the introduction we considered an extension of traditional prototype and exemplar models, neither of which was originally intended to account for the judgment of paired stimuli. In order to accommodate inter-stimulus comparison effects, we assumed that the category representation (whether a prototype or a sample of exemplars) could be shifted in the direction of the context stimulus (Nosofsky & Palmeri, 1997). So if a context stimulus were relatively similar to the borderline stimulus, the category representation would shift toward that context stimulus, and hence the borderline stimulus should assimilate to the category (Stewart & Brown, 2004, 2005; Stewart et al., 2002). Despite its intuitive appeal, this extension of the similarity-based models received no support. Rather than finding assimilation that increased with similarity between a pair of stimuli, instead we found contrast that increased with the difference between them. It therefore seems unlikely that category representations are shifted to any significant extent by the presentation of atypical context stimuli, at least not in the case of familiar semantic categories.

The SD-GCM (Stewart & Brown, 2005) also predicts an assimilation effect. Specifically, when the context stimulus is near the category boundary, that context stimulus will be very similar to the borderline stimulus, and hence there will be a bias for the borderline stimulus to be included in the same category as the context stimulus. So when context and target stimuli are highly similar, the SD-GCM and the similarity-based models make the same prediction, but for different reasons. As described above, the similarity-based models predict assimilation as a consequence of a representational shift. The SD-GCM, on the other hand, does not even assume that there *is* any long-term category representation (Stewart & Brown, 2004; Stewart et al., 2002). Rather, the SD-GCM predicts assimilation solely as a function of inter-stimulus similarity. In the present study, only Experiment 1a showed a significant assimilation effect. And in fact, the assimilative aspect of the SD-GCM has proven elusive for Stewart and Brown (2004) as well. Thus, an important endeavor for future investigation will be to directly test for assimilation in this sort of categorization task. If assimilation cannot be obtained, then the SD-GCM would likely require modification.

However, the SD-GCM could potentially account for the present lack of assimilation in the following way. The model includes a parameter that effectively represents the magnitude of similarity necessary to produce an assimilative category judgment. If one simply assumes that this parameter is set to some value that is greater than the largest value observed in the given stimulus set, then no stimulus in that set will produce an assimilative judgment. To illustrate, assume that the similarity between two stimuli must exceed some value *V* in order for assimilation to occur. If the similarity between hues *b* and *b+1* fails to exceed *V*, then the similarity between *b* and *b+2* must also fail to exceed *V*, and so on. Thus, if no two stimuli in the set have a similarity value that is greater than the value of the similarity parameter itself, then no assimilation would be predicted.

Although theoretically tenable, we find this explanation implausible on empirical grounds. First of all, the difference between neighboring stimuli in our experiments was actually quite small (roughly akin to Just-Noticeable-Differences). For instance, when one compares the neighboring stimuli in Figure 5, their differences are not immediately apparent. It therefore seems unlikely that an even greater similarity between stimuli should be able to induce assimilation. A second reason to doubt this explanation is that Stewart and Brown (2004, p. 423) used a larger difference between stimuli than that used by Stewart et al. (2002), yet this difference between stimuli appeared to have no effect on categorization.

We speculate that the difficulty in obtaining assimilation may instead be attributable to the monotonicity constraint described in the introduction. When judging a pair of stimuli, it almost never occurred that a person would call the bluer stimulus PURPLE, and the purpler stimulus BLUE. In Experiments 2 and 3, only eight such non-monotonic responses occurred, out of almost 6000 opportunities. In Experiment 5 with cats and dogs, the rate was higher (1.3%) but still extremely low. The relative position of the two stimuli on the dimension thus provided a strong constraint on the responses. A simple way to see the effect of monotonicity on category judgments is to consider a borderline target *b* presented in the context of a stimulus *b-1* that itself has a .84 probability of being BLUE (see for example Table 3). Clearly any assimilation effect (to make *b* more blue) will be mitigated by the fact that on 16% of trials the context *b-1* is considered PURPLE, and thus by monotonicity the target *b* (which is less blue than *b*-1) *has to be* categorized as PURPLE. Thus, the logical effect of this monotonicity constraint is to introduce a bias toward contrasting judgments. As a result, the presumed tendency of borderline stimuli to assimilate toward nearby context hues may be counteracted by an opposite bias for contrast introduced by the monotonicity constraint. Note that this source of contrast operates only where the context stimulus has a categorization probability that is not at floor or ceiling. If the context item had a 1.0 probability of being called BLUE, then monotonicity would *not* constrain categorization of the borderline target. So the fact that contrast *increased* with the distance between stimuli indicates that monotonicity does not explain the main contrast effect that we observed. Rather, monotonicity could only explain the lack of assimilation from context hues that were near the category boundary.

Modeling the data – a Criterion Setting account

In this section we discuss a statistical model of our data, and then we sketch an alternative account of the observed contrast effect. For the statistical model we used the data of Experiments 2, 3 and 5, where full data were obtained for each stimulus in the context of all others. We adopted the standard assumptions of statistical decision theory that (1) categorization is determined by the position of the perceived stimulus relative to a decision criterion, and (2) noise in the distance of the perceived stimulus from the decision criterion is normally distributed. Accordingly, the dependent variable was transformed from probability estimates to standardized normal deviates. Thus probabilities of .5 became z scores of 0, those below .5 were negative z scores, and those above were positive.

The resulting z scores based on the data in Tables 3, 4 (same-range only) and 5 were submitted to separate linear regressions, with target scale position (i.e., relative hue labels), and context scale position as independent predictors. Where floor and ceiling effects were apparent, those individual targets were excluded from analyses. However, those stimuli were included in the analyses as context stimuli. Table 6 summarizes the results of the regressions, and also gives details about the range of target stimuli used. It is clear from Table 6 that a simple additive regression model provided an excellent fit to the cell means, accounting for 96% to 98% of the variance in each experiment. In each of the regressions an interaction term failed to be included as a significant predictor in a second step. Regression coefficients indicated that for colors the effect of changing the target stimulus by one step on the scale produced about nine times the change in z score as did changing the context by the same amount. For cats and dogs the context effect was stronger than for colors, with target shifts producing four times the effect of context shifts. For both stimulus domains, it can be concluded that a simple additive effects model, assuming equally spaced steps along the dimension, provides an accurate account of the mean data. One must obviously be cautious of interpreting choice data that have been averaged across participants, but it is interesting to consider what type of process might yield this pattern of results. One plausible alternative to the SD-GCM is a criterion setting model.

Treisman and Williams (1984) proposed that three components are involved in the setting of a criterion for perceptual judgment. Two in particular are of interest here. In a detection task, they suggested that there is a *tracking* mechanism which sensitizes the decision mechanism to stimuli of the class just seen, on the ecological grounds that stability in the world leads to positive autocorrelation between successive trials. More broadly however, a second *balancing* component “works to center the positions of decision criteria in relation to the prevailing flux of relevant sensory inputs” by moving the criterion towards a running average of the stimuli presented. This second component, similar to Helson’s (1964) adaptation-level theory, introduces contrast between successive stimuli. As the criterion is drawn away from a given target stimulus and toward the position of the mean of both stimuli on the scale, so the probability of categorization of the target (be it a positive or a negative categorization) tends to become more extreme.

In order to arrive at the statistical pattern observed in our data, it is sufficient to assume that the criterion (for example between BLUE and PURPLE) starts at some initial position C_0 and then shifts toward the mean of the two stimuli presented, in some constant proportion to the distance from C_0 to that mean. Consider for example the categorization of stimulus b in the context of stimuli ranging from $b-2$ (blue) to $b+2$ (purple). In the context of $b-2$, the criterion will be drawn toward the blue end of the scale, so that b is more likely to be excluded (i.e., contrasted). However in the context of $b+2$ the criterion moves toward the purple end, and so b becomes more likely to be categorized as BLUE. This simple model generates the required contrast effect, and also accounts for the linear relation between normalized categorization probability and the position of the context stimulus. If one additionally assumes that the shift in criterion is subject to asymptote, then the model can not only account for the asymptotic effect in Experiment 2, but also the lack of different-range contrast effects in Experiment 3.

Conclusions

We have presented evidence for a dissimilarity-based contrast effect in simultaneous category judgments. Plausible extensions to traditional similarity-based decision models are unable to account for contrast effects in this task. However, the SD-GCM (Stewart & Brown, 2005) does provide a potential account of the data. The sign-and-magnitude version of the SD-GCM predicts increasing contrast with increasing inter-stimulus distance, and a negatively accelerated effect as distance becomes more extreme. These predictions were supported by our data. Depending on the precise values of the model parameters, the SD-GCM also predicts an assimilation effect when stimuli are very similar to each other.

Although we found little evidence of assimilation, we suggested that the observed monotonicity of responding would in itself introduce a short-range contrast effect that might counteract any assimilative tendency. Finally, we found that an additive model for combining target and context stimulus positions provided an excellent fit to our data in regression analyses. We outlined a criterion setting model in which the decision criterion is shifted in the direction of the presented stimuli. Further research will need to identify whether conditions can be found in which assimilation occurs as predicted by SD-GCM, and to use psychophysical methods to test models on extensive individual data sets, so that effects of averaging across individuals can be discounted.

References

Abramov, I. & Gordon, J. (1994). Color appearance: On seeing red—or yellow, or green, or blue. *Annual Review of Psychology, 45,* 451-485.

Ashby, F.G., & Gott, R.E. (1988). Decision rules in the perception and categorization of multi-dimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 33-53.

Berlin, B. & Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.

Bornstein, M. H. & Korda, N. O. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychological Research, 46,* 207-222.

DeCarlo, L. T. (2003). An application of a dynamic model of judgment to magnitude production. *Perception & Psychophysics, 65,* 152-162.

Estes, W. K. (1994). *Classification and Cognition.* New York: Oxford University Press.

Friedenberg, J., Kanievsky, S. & Kwasniak, M. (2002). Priming and perception of androgynous faces: What you see is not what you get. Poster presented at the 43rd Annual Meeting of the Psychonomic Society, Kansas City, MO.

Gardenfors, P. (2000) *Conceptual Spaces: The Geometry of Thought.* Cambridge Mass: MIT Press.

Geiselman, R. E., Haight, N. A. & Kimata, L. G. (1984). Context effects on the perceived physical attractiveness of faces. *Journal of Experimental Social Psychology, 20,* 409-424.

Helson, H. (1964). *Adaptation-Level Theory.* New York: Harper & Row.

Jameson, D. & Hurvich, L. M. (1959). Perceived color and its dependence on focal, surrounding, and preceding stimulus variables. *Journal of the Optical Society of*

America, 49, 890-898.

Jameson, D. & Hurvich, L. M. (1989). Essay concerning color constancy. *Annual Review of Psychology, 40,* 1-22.

Kay, P. & McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language, 54,* 610-646.

Lamberts, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General, 124,* 161-180.

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review, 107,* 227-260.

Laming, D. R. J. (1997). *The Measurement of Sensation.* London: Oxford University Press.

LeBoeuf, R. A. & Estes, Z. (2004). "Fortunately, I'm no Einstein:" Comparison relevance as a determinant of behavioral assimilation and contrast. *Social Cognition, 22,* 607-636.

MacLaury, R. E. (1991). Prototypes revisited. *Annual Review of Anthropology, 20,* 55-74.

Medin, D. L. & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207-238.

Murphy, G. L. (2002). *The Big Book of Concepts.* Cambridge: MIT Press.

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review, 110,* 472-489.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39-57.

Nosofsky, R. M. & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266-300.

Petzold, P. & Haubensak, G. (2001). Higher order sequential effects in psychophysical judgments. *Perception & Psychophysics, 63,* 969-978.

Petzold, P. & Haubensak, G. (2004). The influence of category membership of stimuli on sequential effects in magnitude judgment. *Perception & Psychophysics, 66,* 665-678.

Roberson, D., Davies, I. & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a Stone-Age culture. *Journal of Experimental Psychology: General, 129,* 369-398.

Rosch, E. (1975). The nature of mental codes for color categories. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 303-322.

Rosch Heider, E. (1972). Universals in color naming and memory. *Journal of Experimental Psychology, 93,* 10-20.

Stapel, D. A. & Winkielman, P. (1998). Assimilation and contrast as a function of context-target similarity, distinctness, and dimensional relevance. *Personality and Social Psychology Bulletin, 24,* 634-646.

Stewart, N. & Brown, G. D. A. (2004). Sequence Effects in the Categorization of Tones Varying in Frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 416-430.

Stewart, N. & Brown, G. D. A. (2005). Similarity and dissimilarity as evidence in perceptual categorization. *Journal of Mathematical Psychology,* in press.

Stewart, N., Brown, G. D. A. & Chater, N. (2002). Sequence effects in categorization of simple perceptual stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 3-11.

Treisman, M., & Williams, T.C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review, 91,* 68-111.

Wedell, D. H. (1995). Contrast effects in paired comparisons: Evidence for both stimulus-based and response-based processes. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1158-1173.

Wedell, D. H., Parducci, A. & Geiselman, R. E. (1987). A formal analysis of ratings of physical attractiveness: Successive contrast and simultaneous assimilation. *Journal of Experimental Social Psychology, 23*, 230-249.

Whittle, P. (2003). Contrast Colours. *In* R.Mausfeld & D.Heyer, (eds.) *Colour Perception, Mind and the Physical World.* Oxford: Oxford University Press (pp 115-138).

Author Notes

The first and second authors contributed equally to this research. Experiments 1 – 3 were presented at the 42nd Annual Meeting of the Psychonomic Society. We thank Greg Murphy, Steve Shevell, Neil Stewart, and anonymous reviewers for constructive comments on an earlier draft, and Vladimir Sloutsky for the stimuli of Experiment 5.

Footnotes

1. Notice that if the two hues were from opposite sides of the category, then contrast would be expected instead. For instance, suppose that S_1 is again midway between BLUE and PURPLE, but S_2 is on the boundary between BLUE and GREEN. S_1 would shift the BLUE category toward PURPLE, and S_2 would therefore become less similar to the BLUE category. The present experiments primarily concern the case in which S_1 and S_2 are from the same side of the category, hence our emphasis is on the prediction of assimilation.
2. Because the stimuli were quantified in terms of software-defined RGB values, rather than in absolute physical properties of the display, one cannot expect the stimuli to be exactly the same in any replication. However, this is not problematic for the present purposes. The critical points are that (1) the selection of hues was individually calibrated for each participant, and remained constant throughout a session, (2) the hues have the correct ordinal properties, and (3) inter-hue differences along the scale were approximately equal. The first point is a fact resulting from the methodology, whereas the latter two assumptions are supported by the behavioral data.
3. If the location of a participant's borderline within the range of nine hues did not allow three hues on either side, then that participant was assigned as her borderline the nearest possible hue that would allow three hues on both sides. For example, if one's borderline was hue #3, then only hues #1 and #2 were left on the blue side of the range. Because our experimental design required *three* hues on either side of the borderline, we assigned as that participant's borderline the nearest hue for which three others *were* available on either side. So in this example, hue #4 was designated the participant's borderline (rather than hue #3).
4. Although not specifically designed for this purpose, it is possible to test for a sequential

contrast effect within the calibration phase of Experiments 1a, 1b, and 2. However, due to the random presentation of hues in that phase, the data were sufficient to evaluate the sign-only model but not the sign-and-magnitude model (as in Friedenberg et al., 2002; Stewart et al., 2002). The data were collapsed across experiments and reverse scored in the case of Experiment 1b. Due to ceiling and floor effects only the five hues in the middle of the range could be analyzed. Furthermore, several participants had to be excluded due to insufficient data. Nevertheless, a significant contrast effect was observed, F(1,136) = 13.24, p = .001.

5. Whereas the dissimilarity-based contrast effect was manifest as a decreasing slope (i.e. same hue range) in Figure 3, it was depicted as an increasing slope in Figure 2. This is because the target category was PURPLE in Experiment 3, but was BLUE in Experiment 2. So although the figures may appear to show opposite patterns, they actually show the same pattern with different target categories (cf. Figure 1).

6. Of the 23 participants, five exhibited category borderlines that did not allow two within-category context stimuli (i.e., their borderlines were *CD1* or *CD2*). Each of these participants was therefore assigned *CD3* as their borderline (cf. Experiments 1 – 3; see footnote 3). However, statistical results were unaffected by removing these five participants instead.

Table 1. Mean probabilities (and standard errors) of categorization of either hue as BLUE, Experiment 1a.

	Target						
	Blue						Purple
	b-3	*b-2*	*b-1*	*b*	*b+1*	*b+2*	*b+3*
Context							
Blue							
b-3	**.98 (.01)**						
b-2	.97 (.02)	**.94 (.02)**					
b-1	.99 (.01)	.88 (.06)	**.74 (.08)**				
b	.97 (.01)	.92 (.05)	.69 (.08)	**.31 (.09)**			
b+1	.99 (.01)	.91 (.05)	.73 (.09)	.28 (.09)	**.12 (.07)**		
b+2	.98 (.01)	.92 (.04)	.75 (.09)	.44 (.10)	.09 (.06)	**.06 (.06)**	
b+3	.96 (.01)	.93 (.03)	.80 (.08)	.54 (.10)	.13 (.07)	.07 (.05)	**.02 (.01)**
Purple							

Note. Hues ranged from clearly blue (i.e., hue *b-3*) through the borderline (i.e., hue *b*) to clearly purple (i.e., hue *b+3*). Target hue was defined as the bluer of the two hues. Figures in **bold** represent the baseline measure for that hue, where each is paired with itself.

Table 2. Mean probabilities (and standard errors) of categorization of both hues as PURPLE, Experiment 1b.

	Target						
	Blue						Purple
	b-3	*b-2*	*b-1*	*b*	*b+1*	*b+2*	*b+3*
Context							
Blue							
b-3	**.01 (.01)**						
b-2	.00 (.00)	**.00 (.00)**					
b-1	.01 (.01)	.06 (.04)	**.12 (.07)**				
b	.01 (.01)	.04 (.03)	.16 (.08)	**.45 (.10)**			
b+1	.01 (.01)	.02 (.02)	.13 (.06)	.51 (.08)	**.80 (.09)**		
b+2	.00 (.00)	.02 (.01)	.14 (.07)	.39 (.09)	.77 (.09)	**.89 (.07)**	
b+3	.00 (.00)	.01 (.01)	.12 (.07)	.30 (.11)	.52 (.11)	.74 (.09)	**.81 (.09)**
Purple							

Note. Hues ranged from clearly blue (i.e., hue *b-3*) through the borderline (i.e., hue *b*) to clearly purple (i.e., hue *b+3*). Target hue was defined as the less purple of the two. Figures in **bold** represent the baseline measure for that hue, where each is paired with itself.

Table 3. Mean probabilities (and standard errors) of categorization of the target hue as BLUE, Experiment 2.

	Target						
	Blue						Purple
	b-3	*b-2*	*b-1*	*b*	*b+1*	*b+2*	*b+3*
Context							
Blue							
b-3	**.98 (.02)**	.91 (.05)	.80 (.08)	.45 (.10)	.16 (.07)	.01 (.01)	.01 (.01)
b-2	.99 (.01)	**.95 (.03)**	.84 (.06)	.44 (.10)	.13 (.06)	.01 (.01)	.00 (.00)
b-1	.99 (.01)	.96 (.02)	**.80 (.08)**	.56 (.10)	.17 (.07)	.00 (.00)	.00 (.00)
b	.99 (.01)	.95 (.03)	.84 (.07)	**.60 (.11)**	.13 (.07)	.03 (.02)	.01 (.01)
b+1	.99 (.01)	.97 (.02)	.84 (.06)	.63 (.10)	**.20 (.07)**	.02 (.02)	.01 (.01)
b+2	.98 (.01)	.97 (.02)	.86 (.06)	.72 (.10)	.26 (.10)	**.02 (.02)**	.00 (.00)
b+3	.99 (.01)	.99 (.01)	.92 (.06)	.72 (.10)	.33 (.10)	.06 (.04)	**.00 (.00)**
Purple							

Note. Hues ranged from clearly blue (i.e., hue *b-3*) through the borderline (i.e., hue *b*) to clearly purple (i.e., hue *b+3*). Figures in **bold** represent the baseline measure for that hue, where each is paired with itself.

Table 4. Mean probabilities (and standard errors) of categorization of the target hue as PURPLE, Experiment 3.

	Blue-purple target					Purple-red target				
	Blue				Purple	Purple				Red
	b-2	*b-1*	*b*	*b+1*	*b+2*	*b+2*	*b+1*	*b*	*b-1*	*b-2*
Blue-purple context										
Blue										
b-2	**.01 (.01)**	.09 (.02)	.56 (.08)	.92 (.04)	.96 (.04)	.94 (.05)	.71 (.07)	.36 (.09)	.06 (.04)	.02 (.02)
b-1	.01 (.01)	**.07 (.03)**	.41 (.07)	.84 (.06)	.96 (.04)	.93 (.05)	.74 (.08)	.35 (.08)	.06 (.04)	.01 (.01)
b	.03 (.02)	.13 (.03)	**.38 (.07)**	.71 (.07)	.93 (.05)	.92 (.05)	.74 (.06)	.33 (.08)	.08 (.04)	.00 (.00)
b+1	.03 (.01)	.05 (.02)	.41 (.07)	**.81 (.06)**	.93 (.05)	.89 (.05)	.64 (.08)	.31 (.07)	.08 (.04)	.01 (.01)
b+2	.02 (.01)	.03 (.01)	.24 (.07)	.71 (.07)	**.93 (.03)**	.81 (.07)	.64 (.07)	.31 (.08)	.07 (.05)	.01 (.01)
Purple										
Purple-red context										
Purple										
b+2	.03 (.02)	.08 (.03)	.21 (.05)	.52 (.07)	.83 (.05)	**.90 (.05)**	.71 (.07)	.26 (.07)	.04 (.02)	.00 (.00)
b+1	.01 (.01)	.09 (.03)	.34 (.07)	.69 (.08)	.86 (.05)	.89 (.05)	**.74 (.07)**	.35 (.07)	.06 (.03)	.01 (.01)

b	.01 (.01)	.07 (.02)	.27 (.06)	.67 (.07)	.88 (.06)	.92 (.05)	.73 (.08)	**.35 (.07)**	.10 (.04)	.00 (.00)
b-1	.02 (.01)	.05 (.02)	.22 (.05)	.68 (.07)	.88 (.05)	.93 (.05)	.81 (.06)	.37 (.08)	**.07 (.05)**	.00 (.00)
b-2	.01 (.01)	.01 (.01)	.22 (.04)	.63 (.06)	.91 (.05)	.93 (.05)	.79 (.07)	.53 (.09)	.13 (.06)	**.00 (.00)**
Red										

Note. Hues were defined within their respective hue ranges, such that hue *b+2* from the blue-purple range, for instance, was *not* identical to hue *b+2* from the purple-red range. Figures in **bold** represent the baseline measure for that hue, where each is paired with itself.

Table 5. Mean probabilities (and standard errors) of categorization of the target image as a CAT, Experiment 5.

	Target						
	Cat						Dog
	CD1	*CD2*	*CD3*	*CD4*	*CD5*	*CD6*	*CD7*
Context							
Cat							
CD1	**.82 (.07)**	.65 (.07)	.33 (.07)	.17 (.07)	.04 (.03)	.02 (.02)	.02 (.02)
CD2	.78 (.08)	**.70 (.08)**	.50 (.08)	.15 (.06)	.11 (.05)	.04 (.03)	.02 (.02)
CD3	.91 (.04)	.70 (.08)	**.51 (.10)**	.35 (.07)	.13 (.05)	.02 (.02)	.04 (.03)
CD4	.83 (.07)	.74 (.08)	.57 (.08)	**.38 (.10)**	.13 (.06)	.13 (.06)	.04 (.03)
CD5	.83 (.07)	.78 (.06)	.61 (.09)	.35 (.09)	**.14 (.05)**	.13 (.06)	.07 (.04)
CD6	.96 (.03)	.80 (.06)	.67 (.08)	.37 (.08)	.22 (.06)	**.00 (.00)**	.04 (.04)
CD7	.96 (.03)	.91 (.05)	.74 (.07)	.43 (.08)	.26 (.08)	.02 (.02)	**.05 (.03)**
Dog							

Note. Stimuli ranged from a photo of a cat (i.e., *CD1*) through intermediate morphed images (i.e., *CD2 – CD6*) to a photo of a dog (i.e., *CD7*). Figures in **bold** represent the baseline measure for that stimulus, where each is paired with itself.

Table 6. Results of regression analyses applied to standardized normal transformations of the choice proportion data from Experiments 2, 3 and 5.

Experiment	Regression coefficients			Target range used		Adjusted
- condition	Target	Context	Intercept	from	to	R square
Experiment 2	0.98	-0.11	0.00	*b-2*	*b+2*	.97
Experiment 3 - blue	1.18	-0.16	-0.29	*b-1*	*b+1*	.97
Experiment 3 - red	0.95	-0.11	-0.40	*b-1*	*b+2*	.98
Experiment 5	0.58	-0.14	0.49	*CD1*	*CD5*	.96

Notes: Stimuli were coded so that the center of the scale *b* was zero, and steps towards the target category were positive.

Figure Captions

Figure 1. Probability of positive categorization of the borderline hue as a function of context hue, Experiments 1a (BLUE) and 1b (PURPLE).

Figure 2. Probability of positive categorization (BLUE) of the borderline hue as a function of context hue, Experiment 2.

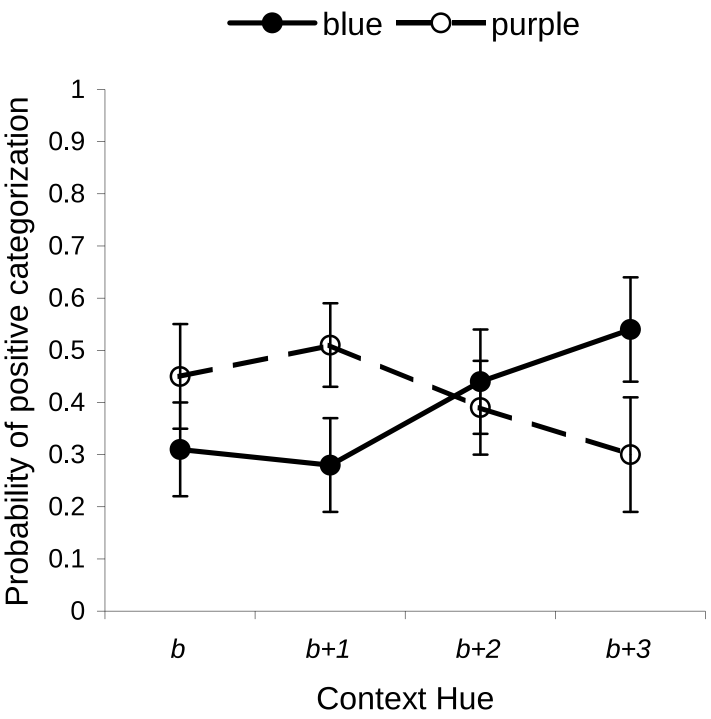
Figure 3. Probability of positive categorization (PURPLE) of the borderline hue as a function of context hue and hue range (i.e., “same” or “different”), Experiment 3.

Figure 4. Percent correct detection of a target hue change with separated and embedded presentation, as a function of hue change direction, Experiment 4.

Figure 5. Stimuli, Experiment 5.

Figure 6. Probability of positive categorization (CAT) of each target image as a function of context image, Experiment 5.

Figure 1.



Note. “blue” = probability that either the context or the borderline hue was judged blue; “purple” = probability that both the context and the borderline hues were judged purple. Error bars indicate +/- one standard error of the mean (in all figures).

Figure 2.

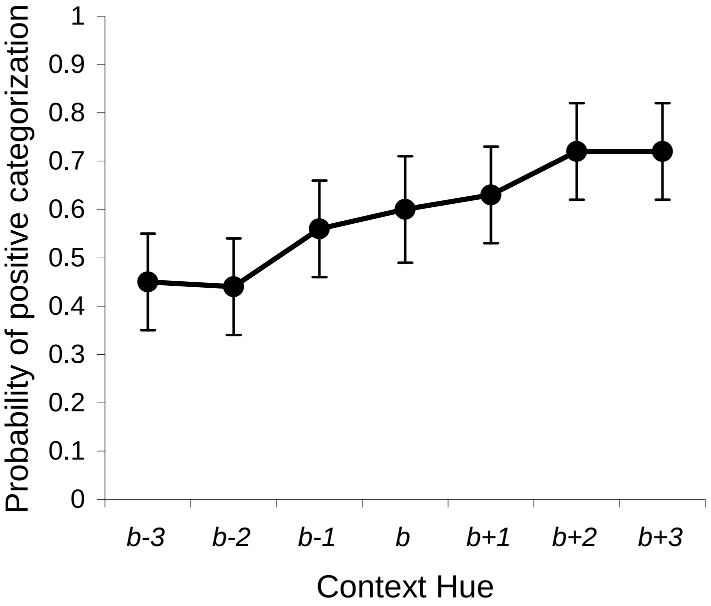
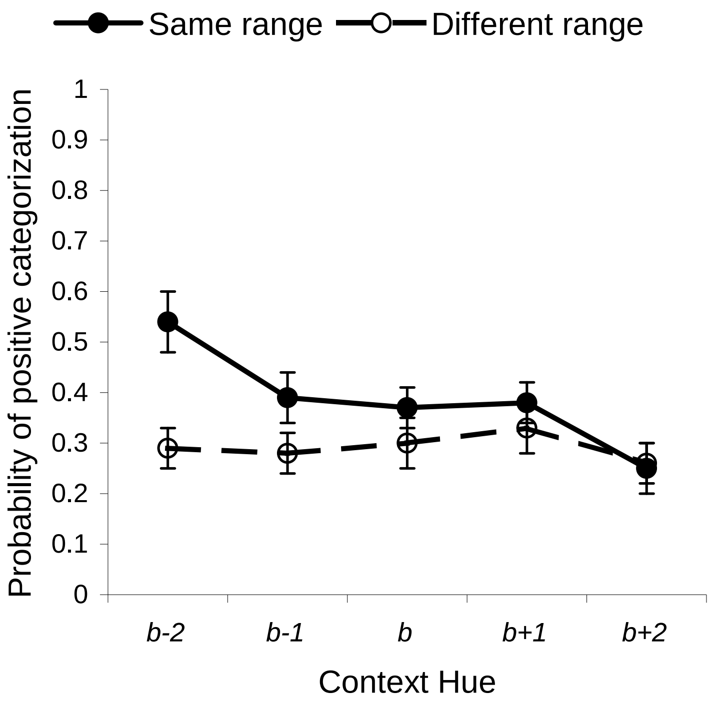
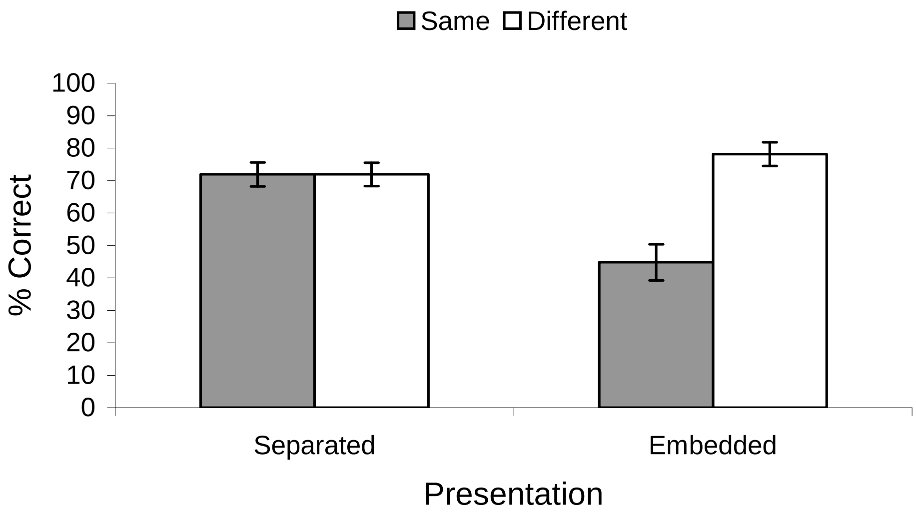

Probability of positive categorization
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
b-3
b-2
b-1
b
b+1
b+2
b+3
Context Hue

Figure 3.



Note. “Same range” indicates that target and context hues were from the same hue range (i.e., blue-purple or purple-red); “Different range” indicates that target and context hues were from different hue ranges. Data are collapsed across the blue-purple and purple-red borderline hues.

Figure 4.



Note. “Same” indicates that context and target hues changed in the same direction (i.e., both blue to purple, or both purple to blue); “Different” indicates that context and target hues changed in different directions (e.g., one from blue to purple, and one from purple to blue).

Figure 5.

CD1

CD2

CD3

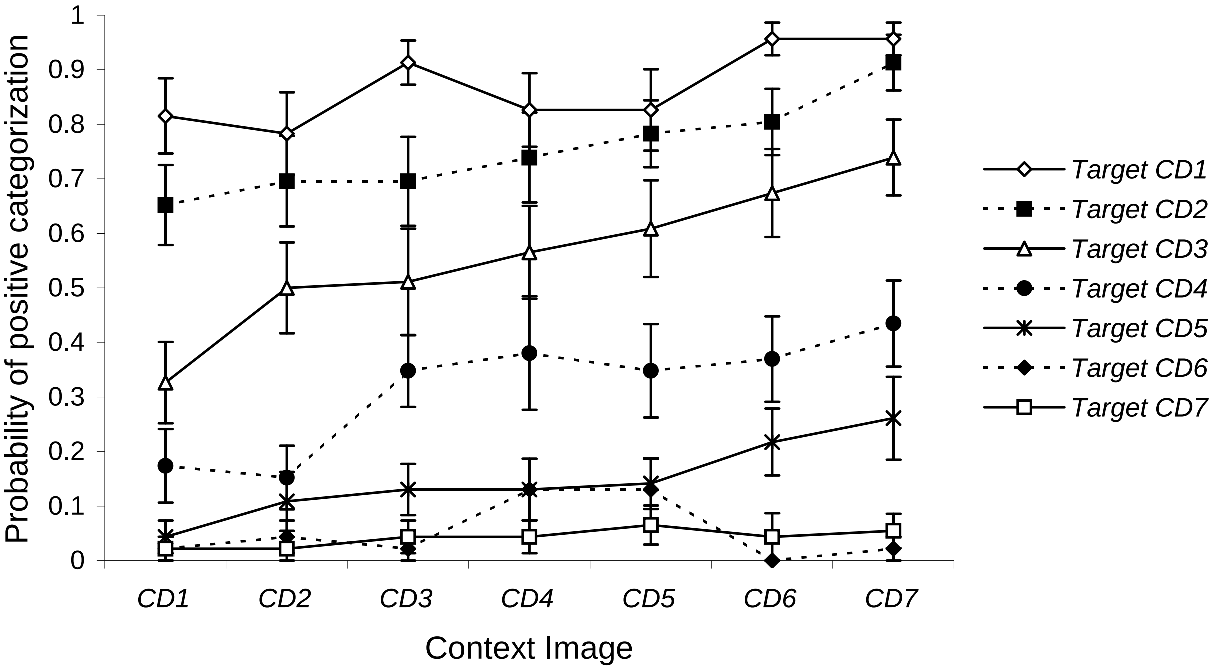
CD4

CD5

CD6

CD7

Figure 6.


Probability of positive categorization
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
CD1
CD2
CD3
CD4
CD5
CD6
CD7
Context Image
Target CD1
Target CD2
Target CD3
Target CD4
Target CD5
Target CD6
Target CD7