



Published in final edited form as:

Emotion. 2010 August ; 10(4): 544–554. doi:10.1037/a0019017.

Reading Chimpanzee Faces: Evidence for the Role of Verbal Labels in Categorical Perception of Emotion

Jennifer M.B. Fugate^{1,*}, Harold Gouzoules², and Lisa Feldman Barrett^{1,3}

¹Department of Psychology, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467

²Department of Psychology, Emory University, 532 Kilgo Circle, Atlanta, GA, 30322

³Department of Psychiatry and the Martinos Center for Biomedical Imaging, Massachusetts General Hospital- East, Building 149 13th Street, Charlestown, MA 02129

Abstract

Categorical perception (CP) occurs when items in a series of continuously varying stimuli are perceived as belonging to discrete categories. Thereby, perceivers are more accurate at discriminating between stimuli of different categories than between stimuli within the same category (Harnad, 1987; Goldstone, 1994). The current experiments investigated whether the structural information in the face is sufficient for CP to occur. Alternatively, a perceiver's conceptual knowledge, by virtue of expertise or verbal labeling, might contribute. In two experiments, people who differed in their conceptual knowledge (in the form of expertise, Experiment 1, or verbal label learning, Experiment 2) categorized chimpanzee facial expressions. Expertise alone did result in enhanced CP. Only when perceivers were first trained to associate faces with a label were they more likely to show CP. Overall, the results suggest that the structural information in the face alone is often insufficient for CP; CP is enhanced by verbal labeling.

Keywords

facial expressions; emotion; categorical perception; verbal labeling; expertise

Introduction

Categorical perception (CP) occurs when objects that vary along a continuous dimension come to be experienced as similar enough that they are assigned to be members of the same category, as distinct from other images assigned to a different category. The two categories are marked by a sharp boundary (Harnad, 1987). As a result, perceivers are more accurate at discriminating between stimuli of different categories than between stimuli within the same category (i.e. a between-category advantage) (Goldstone, 1994). Humans perceive emotion in categorical terms (Calder, Young, Perrett, Etcoff, & Rowland, 1996; Etcoff & Magee, 1992; Young et al., 1997). We easily and effortlessly perceive anger, or sadness, or fear in another person's face. Many scientists believe that emotion perception is simple and undemanding because the facial muscle movements broadcast the internal state of the sender, thereby allowing the perceiver

* to whom correspondence should be addressed: fugatej@bc.edu, 617-552-1787.

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/pubs/journals/emo.

to automatically “recognize” emotion (the structural hypothesis) (Izard, 1971; Tomkins, 1962; Ekman, 1992). In this view, perceivers are merely translating the information about emotion that is carried in the facial movements of the sender. Alternatively, the categorization that occurs in emotion perception might arise from the conceptual knowledge that is evoked when the perceiver views the structural information in another’s face in context (the conceptual hypothesis) (Barrett, Lindquist, & Gendron, 2007). In this view, conceptual knowledge constrains the meaning of the structural information from the face, allowing a perceiver to arrive at a categorical distinction, even when the information in the face itself might not be sufficient for such distinctions. Conceptual knowledge can be evoked in many different ways, including via prolonged exposure to category members (i.e., expertise), verbal labeling, contextual priming, and so forth.

All adult humans (barring organic disturbances) have expertise with emotion perception to some degree because human facial movements are seen and interpreted as an emotional expressions on a regular basis. As a result, it is impossible to definitively determine whether the structural information in the face alone or automatically-activated conceptual knowledge is driving the category judgments that people make about emotion. Therefore, in the current studies, we had human perceivers view and make judgments about chimpanzee facial expressions. Chimpanzees and humans have nearly identical mimetic musculature and stimulation of some of these muscles in both species results in similarly-looking facial expressions (Burrows, Waller, Parr, & Bonar, 2006; Parr, Waller, Vick, & Bard, 2007; Waller, et al., 2006). In two studies, we tested the structural and conceptual hypotheses using these faces. Specifically, we examined whether conceptual knowledge in the form of expertise (Experiment 1) and verbal labeling (Experiment 2) plays a role in humans’ categorization of chimpanzee facial expressions.

Expertise

Indirect support for the idea that people are using their expertise with faces to make effortlessly categorical judgments comes from experiments that examine face identity. People easily show CP for familiar faces (e.g. Beale & Keil, 1995), but do not typically exhibit CP for unfamiliar faces of identity. Instead, CP becomes possible only when the unfamiliar faces (category anchors) are viewed during extended periods of training prior to the experiment, or when the individual stimuli are learned by repetition over the course of the experiment (McKone, Martini, & Nakayama, 2001; Stevenage, 1998; Viviani, Binda, & Borsato, 2007). When perceivers show CP for unfamiliar faces without extensive training, one category anchor is very distinctive or atypical when compared to the other (see Angeli, Davidoff, & Valentine, 2008), or category anchors are held in memory during the categorization judgments, which can lead to a distortion of the face space, making CP more likely (see Angeli, et al., 2008; McKone, et al., 2001).

CP requires ignoring differences across exemplars that are not meaningful to category membership. People can still discriminate individual exemplars from one another, but become more sensitive to differences that reside at the category boundary. Said another way, CP results from cross category expansion, not from within category compression (Ozgen & Davies, 2002). This is quite different from the “perceptual narrowing” that occurs across the course of development in which perceivers lose the ability to discriminate between category exemplars as they once did (e.g. the “other-race” or “other-species” effect (Kelly, Quinn, Slater, Lee, Ge, & Pascalis, 2007; Pascalis, de Haan, & Nelson, 2002).⁴ Expertise might help a perceiver

⁴Interestingly, Scott & Monesson (2009) showed that one way perceptual narrowing can be overcome is to exposure infants to individually labeled faces (compared to unlabeled faces or faces with the same label).

distinguish between those perceptual differences that are psychologically meaningful, and those that are not.

The problem of how people ignore variation across exemplars to allow CP is exacerbated when there is tremendous variability within a category, or the categories themselves are not grounded in perceptual regularities per se. This appears to be the case for emotion categories (cf. Barrett, 2006a, b, 2009). In such cases, what dictates the differences that are psychologically meaningful from those that are not? A growing body of research suggests a role for verbal labeling (Booth & Waxman 2003; Dewar & Xu, 2009; Fulkerson & Waxman, 2006; Waxman & Braun, 2005; Waxman & Markow, 1995).

Verbal Labeling

Applying the same word to physically different exemplars allows a perceiver to place them into the same category. Kitutani, Roberson, and Hanley (2008) presented unfamiliar faces (e.g. two different identities), either by themselves or paired with a label, during a familiarization task completed prior a CP experiment. Only when participants were exposed to the identities with a label did they subsequently show CP for the faces. Exposure to the identities by themselves was not sufficient to produce CP. The activation of verbal labels to produce CP has also been proposed in domains other than facial expression, such as color (Pilling, Wiggett, Ozgen, & Davies, 2003; Roberson & Davidoff, 2000; Winawer, Witthoft, Frank, Wu, Wade, Boroditsky, 2007)⁵. Consistent with this account, CP is eliminated when a secondary task is employed that disrupts verbal processing (Roberson & Davidoff, 2000). Whether or not language causes CP or simply supports it is still a matter of debate. Roberson, Damjanovic, and Pilling (2007) proposed a category “adjustment” model in which labels are less directly involved in CP, shaping already existing grouping.

The Present Experiments

In Experiment 1, we tested whether non-human primate experts⁶ or novices⁷ showed CP for morphed (blended) chimpanzee faces. If the structural hypothesis is correct and the signal in the face is sufficient for CP to occur, then both groups should show CP. However, if only experts show CP or CP increases significantly compared with novices, then the results provide support for the conceptual hypothesis that expertise contributes to CP. In Experiment 2, we tested whether novices who first learned the chimpanzee facial expression categories either paired with a label (“label” learners) or without a verbal label (“no-label” learners) showed CP. If the structural hypothesis is correct, then both groups of participants, regardless of training, should show CP. However, if only the “label” learners show evidence of CP or CP increases significantly compared with “no-label” learners, then the results provide support for the conceptual hypothesis that verbal labeling contributes to CP.

CP was assessed by two widely used tasks (e.g. Calder, et al., 1996; Etcoff & Magee, 1992; Young et al., 1997). Each task involved participants viewing morphed chimpanzee facial expressions created from pairs of four well-documented expressions (bared teeth face, hoot face, scream face, and play face). Six morphs were created between each pair (X-Y) of facial expressions (morph 3, 86%X-14%Y; morph 5, 71%X-29%Y; morph 7, 57%X-43%Y, morph 9, 43%X-57%Y; morph 11, 29%X-71%Y; morph 13, 14%X-86%Y). In the *identification*

⁵We thank an anonymous reviewer for pointing out the complexities of verbal labeling on CP. Roberson, Davidoff, and Braisby (1999) showed that at least one patient with color anomia, who had difficulty with color grouping, showed CP for color when the procedure relied less on labeling and become more automatic, supporting the “category adjustment” model.

⁶People who had worked with at least one non-human primate species consecutively for at least 12 months any time during the past five years and have some familiarity with the behaviors, including the facial expressions, of the species.

⁷People who had never worked with any non-human primate species and had no formal training in non-human primate behaviors, including facial expressions.

task, participants saw each morph along with the two faces from which the morph was created (e.g. endpoints of the continuum) as comparison images. Participants judged whether the morph was more like one comparison image or the other. If participants are capable of showing CP, their identifications should shift abruptly at one point along the morphing continuum (Harnad, 1987). Such a discrete shift in identification is known as the categorical boundary. In the AB-X *discrimination task*, participants first saw two morphed faces (face A followed by face B) that differed by one incremental step from each other. Next, either A or B was re-shown (X) and participants judged whether this target face was the same as face A or B⁸. If participants show CP, they should be more accurate in discriminating between morphs that cross the categorical boundary compared with morphs that do not. Such an advantage is known as a *between category advantage* (Goldstone, 1994).

Methods

Participants

In Experiment 1, 15 “experts” (2 M, 13 F; average years expertise = 5.6 years, range 1 – 9 years) and 15 “novices” from Emory University volunteered to participate (4 M, 11F). Data from one participant were removed due to poor performance (less than 69%) on the control trials in one part of the experiment. In Experiment 2, 28 Boston College (8 M, 20 F) undergraduate students participated for research credit. Participants were randomly assigned to one of the two training groups. As per Experiment 1, data from four participants were removed due to poor performance.

Morph Construction

Morphs were created with commercial software (FantaMorph 3 Deluxe edition; Abrosoft, version 3.6.1 www.fantamorph.com) from pairs of four black and white pictures, each which depicted a different, prototypical chimpanzee facial expression (bared teeth face, hoot face, scream face, and play face). Each of the four endpoints was rated by chimpanzee experts as being the most prototypical of each expression category from a large database of photos maintained by the Yerkes National Primate Research Center. The result was six expression continua (bared teeth – hoot, bared teeth – play, bared teeth – scream, hoot- scream, play – hoot, play - scream). Six morphs were created for each of the six expression continua (see Figure 1).

In addition, a separate sample of undergraduates unfamiliar with chimpanzee expressions rated each pair of endpoints on their similarity to one another in order to assess whether any effects could be attributed to the distinctiveness of one endpoint over another (see Angeli, et al., 2008). The pairs that included the “hoot” expression (e.g. bared teeth-hoot, hoot-scream, play-hoot) were rated more different than those pairs which did not include the hoot expression (which did not differ from one another). Although participants’ ratings for the pairs which did not involve the hoot expression were judged as less similar, they were not judged as identical. In addition, participants in the actual experiments had no trouble discriminating between endpoints as each endpoint served in several control trials.

Procedure

Participants completed the discrimination task, followed by the identification task. In the ABX discrimination task, all three morphs were sequentially shown in the center of the computer

⁸There are several ways to assess the between category advantage (e.g. ABX, similarity, and better likeness) (see McKone, et al., 2008). We choose an ABX task for several reasons: 1.) it is most widely used; 2) it does not require showing the endpoint images (as do both the other tasks); 3.) it does not require holding the images in memory for more than one second (as does the better likeness task) which can cause distortions of the face space.

screen for 500 ms each, separated by 500 ms of blank screen. In response to the third image, participants pressed either the “1” key to select the first image, and the “2” key to select the second image. Participants discriminated all AB morph pairs eight times (where X was A on half the trials, and X was B on half). There was no time limit to indicate a response.

In the identification task, a morph was always presented on the top, center of the screen. Participants pressed the space bar to activate two comparison morphs that were displayed at the bottom of the screen and to the right and left of the initial morph. Participants pressed the “k” to indicate the comparison on the right or “d” to indicate the comparison on left. Participants identified all morphs four times. There was no time limit to indicate a response.

Experiment 2 Additional Training

In Experiment 2, participants were randomly assigned to one of two training groups. Both groups of participants were told that they would first undergo training in which their goal was to learn the different chimpanzee facial expressions presented. One group, “no-label” learners, was trained by viewing different faces belonging to the four different chimpanzee facial expressions categories (not the comparison images used in the identification task). The other group, “label” learners, saw the same faces presented with a category nonsense label. Nonsense labels were used because we could control the length of the words. In addition, real labels might have directed participants’ attention to certain facial features (e.g. bared teeth). Images were shown in the middle, center of the computer screen. Participants could view each image for as long as they wished; images were advanced by pressing the space bar. Four faces from each category were presented four times in random order.

Participants then completed an assessment to test whether they learned the facial expression categories. “Label” learners were shown a label in the center top of the screen, along with two of the previously-seen faces as comparisons presented on the bottom to the right and left of the label. Participants in the “no-label” group were shown a previously-seen face in place of the label. Therefore, participants in the “label” group were asked to match a category label to a face in the same category, whereas those in the “no-label” group were asked to match a face to another face from the same category. The “label” group’s assessment was purposefully constructed as such to emphasize learning the labels as category words. Participants pressed the “k” key to indicate the comparison on the right, and the “d” key for the one on the left. Participants were required to achieve at least 95.00% accuracy on the assessment before completing the two tasks. If participants did not pass on their first time, they repeated the viewing and assessment until they passed. Fifty three percentage of participants passed the assessment on the first time (max =4). The number of times required to pass the assessment did not vary between the two groups.

Results

Experiment 1

Identification Task—Repeated Measures ANOVAs were conducted on each expression continua separately using group (“expert” and “novice”) as a between subjects factor. Experts and novices showed no difference in the way they identified any morphs on any continuum. Therefore, we combined the number of identifications “experts” and “novices” gave to morphs along each continuum and recalculated the Repeated Measures ANOVAs. Not surprisingly, the effect was significant for each continuum, suggesting that as the proportion of an expression decreased, the number of identifications to that expression also decreased (bared teeth-hoot: $F(5,145) = 375.52$; bared teeth-play: $F(5,145) = 246.77$; bared teeth-scream: $F(5,145) = 195.034$, hoot-scream: $F(5,145) = 347.77$; play-hoot: $F(5,145) = 362.52$; play-scream: $F(5,145) = 140.28$, $p < 0.001$ in all cases) (see Figure 2).

We followed up the main effect with planned, multiple paired t-tests (corrected for multiple comparisons) to determine which pair(s) of adjacent morphs was/were significantly different. A discrete categorical boundary would be reflected in a step-like shape, with only one significant difference between pairs of adjacent morphs; all morphs up to that certain point would be identified similarly to one endpoint, all morphs exceeding that point would be identified as similar to the other endpoint (but dissimilar from the first). A weaker shift in perception would be reflected in a sigmoidal shape, characterized by multiple significant differences among several adjacent pairs⁹. Our data conformed to the latter function, with multiple significant differences between pairs of adjacent morphs (see SI1).

Because we did not find that one pair of adjacent morphs which differed in their identification (a discrete boundary) for any of the morphed expression continua, we estimated the category boundary using a proxy. We identified the two adjacent morphs that were judged as maximally different (even when the differences between other adjacent morph pairs were also statistically significant) (see † in SI1). We used these two morphs to test whether experts and/or novices were more accurate at discriminating this pair (between category pair) compared to morph pairs which did not cross the boundary (within category pairs). For example, participants identified morphs 7 and 9 to be maximally different along the bared teeth - hoot continuum, and so we examined in the discrimination task whether experts and/or novices were significantly more accurate in discriminating the between category morph pair created from morphs 7 and 9 compared to the average of the within category morph pairs created from morphs of the same physical distance but that did not span the boundary (e.g. morphs 3 and 5, morphs 5 and 7, morphs 9 and 11, and morphs 11 and 13).

As a second way to more precisely determine the category boundary, we fitted three known functions through the data to best predict the underlying distribution (e.g. a step function, a logistic function, and a linear function). From the line of best fit, we calculated the center point, X_c , which would be the category boundary (McKone, et al., 2001). A step function should be the best fit if the data show a discrete categorical boundary, whereas a logistic function should be best fit if the data show a weaker shift in perception. Lastly, a linear function should best fit the data if the morphs are identified in the way in which they were created.

The results showed that every continuum was best fitted with a logistic function (see Table 1), and the pattern of residuals produced a random pattern, indicating that the sigmoid shape was a suitable description of the data. The category boundaries (X_c) for each continuum can be found on Figure 2. For all but one continuum (play-scream), X_c matched the boundary we found using the largest t-statistic as a proxy. In this case, X_c fell nearly on one of the morphs, so we used the largest t-statistic to indicate which two morphs spanned the boundary.

Overall, the results of the identification task showed that both experts and novices were fairly good (although not perfect) at sorting morphs into two classes when asked to do so. In addition, experts and novices did not differ on the way they sorted any of the morphs.

Discrimination Task—Data from the discrimination task are necessary to determine whether the between category advantage has occurred. To assess whether there was a between category affect, we used one sample t-tests in which the mean accuracy of the between category pairs was compared to the average accuracy from the within category pairs for each continuum for the experts and novices separately. Experts showed a between category advantage on three of the continua (bared teeth-scream: $t(59) = -4.229$, $p < 0.001$; hoot-scream: $t(59) = -2.512$, p

⁹Although debate surrounds the ideal expected shape (see Harnard, 1987 for discussion), researchers typically look for either a step or logistic function to predict more of the variance than a linear function (which would occur if the perception matched the manner in which stimuli were created).

= 0.008; play-hoot: $t(59) = -3.651$, $p < 0.001$). The other three continua (bared teeth-hoot: $t(59) = -1.308$, $p = 0.098$; bared teeth-play: $t(59) = -0.745$, $p = 0.230$; play-scream: $t(59) = 3.566$ ¹⁰, $p < 0.001$) did not show the between category advantage (see Figure 3). Novices, on the other hand, showed a between-category advantage on two of the continua (bared teeth – play: $t(59) = -4.575$, $p < 0.001$; play - hoot: $t(59) = -6.177$, $p < 0.001$). The other four continua did not show a between category advantage (bared teeth-hoot: $t(59) = 2.712$, $p = 0.005$ ¹⁰); bared teeth-scream: $t(59) = -1.187$, $p = 0.120$; hoot-scream: $t(59) = 1.788$, $p = 0.40$; play-scream: $t(59) = 0.487$, $p = 0.314$) (see Figure 4).

The fact that both experts and novices were able to discriminate many morph pairs above chance (not just the predicted pairs) (see SI2) suggested that even novices could detect structural differences in the morphs along a continuum, but neither they nor the experts knew which discriminations were psychologically meaningful (i.e. mapping to category membership)¹¹.

Overall Advantage: We also combined all six expression continua to test whether experts and/or novices showed an *overall* between category advantage. We combined the average accuracies from the within category pairs on each continua in a paired t-test with the accuracy for the between category pair on each continua. When we did this, neither experts nor novices were more accurate at discriminating the between category morph pairs compared to the within morph pairs: experts: $t(5) = -1.2373$, $p = 0.130$; novices: $t(5) = -0.767$, $p = 0.478$. In addition, we thought it possible that perhaps experts or novices might show evidence for the between category effect overall if we compared the accuracy for the between category pairs from each continua with only the within category pair at the continua ends (where the morphs in each pair were identified most similarly). When we did this, experts were still not more accurate at discriminating the between category morph pairs compared to within category pairs created from either end of the continuum ($t(89) = -1.653$, $p = 0.102$; $t(89) = -1.033$, $p = 0.305$). Using the same comparisons, novices were also not more accurate at discriminating the between category morph pairs compared to within category morphs at the continua ends ($t(89) = -0.763$, $p = 0.448$; $t(89) = -0.100$, $p = 0.920$). These results suggest that at a global level, neither experts nor novices showed CP.

Summary—To summarize findings across both tasks, experts and novices did not differ in the way they identified the morphs, such that both were able to identify structural changes at multiple points along each continuum. Both novices and experts, while being able to detect differences among most of the morphs, failed to understand which changes mapped on to category membership. When we took into account the largest of these identification changes (which coincided with Xc from the sigmoid function), experts and novices showed the between category advantage for three and two continua, respectively. Experts and novices agreed in their CP for one of these continua (play – hoot). When we combined all six continua, neither experts and novices showed an *overall* between category advantage, suggesting that at a global level neither group showed CP. Our results suggest that the structural information in the face was not sufficient for the categorical perception of chimpanzee facial expressions in the majority of cases. Conceptual knowledge, in the form of expertise, did not markedly improve CP however.

¹⁰The test is significant but in the direction opposite to predictions, such that the within are more accurate than the between pairs.

¹¹Above chance discriminations, even among within category pairs, is not unusual in CP experiments. Such performance means that participants are still able to discriminate among individual exemplars. Unique to the between category advantage, however, would be whether participants are able to generalize across exemplars to form meaningful categories.

Experiment 2

Identification Task—Similar to the results observed for experts and novices in Experiment 1, “label” and “no-label” learners did not differ in the number of identifications given to morphs along any continuum. Therefore, we combined the two groups and recalculated the Repeated Measures ANOVAs for each expression continuum separately. Again, not surprisingly, the over effect was highly significant for each continuum, suggesting that as the proportion of an expression decreased, the number of identifications to that expression also decreased (bared teeth-hoot: $F(5,135) = 224.177$; bared teeth-play: $F(5,135) = 91.675$; bared teeth-scream: $F(5,135) = 242.323$; hoot-scream: $(5,135) = 176.616$; play-hoot, $F(5,135) = 279.402$; play-scream, $(5,135) = 100.743$, $p < 0.001$ in all cases) (see Figure 5).

As Experiment 1, participants identified several adjacent morphs along every continuum as statistically different from one another (see SI1). Following the logic laid out in Experiment 1, we identified the two adjacent morphs that were judged as maximally different (when the differences between other adjacent morph pairs were also statistically significant) (see † in SI1), and used this difference to test whether “label” learners and/or “no-label” learners showed the between-category advantage in the discrimination task. We also calculated the amount of variance each of a step, linear, and logistic function predicted. As In Experiment 1, a logistic function best fit the data for all continua (see Table 1). In all but one case (hoot-scream), X_c fell between the two morphs that were judged as maximally different (the largest t statistic between controlled comparisons between adjacent morphs). In this case, we used the largest t -statistic to indicate which two morphs span the boundary.

Overall, “label” and “no-label” learners did not differ on the way they identified any of the morphs on any of the continua. In addition, as with the experts and novices, both trained groups here were fairly good at sorting morphs reliability across trials and similarity with one another.

Discrimination Task—To assess whether “label” learners and/or “no-label” learners showed a between category advantage, we used the same analyses described in Experiment 1. “Label” learners showed a between category advantage for four of the six morphed continua (bared teeth-scream: $t(55) = -2.385$, $p = 0.011$; hoot-scream: $t(55) = -2.957$, $p = 0.003$; play-hoot: $t(55) = -2.908$, $p = 0.003$; play-scream: $t(55) = -4.704$, $p < 0.001$) (see Figure 6). For the other two continua, “label” learners showed no between category advantage (bared teeth-hoot: $t(55) = 0.614$, $p = 0.271$; bared teeth-play: $t(55) = 2.011$, $p = 0.025^{10}$). “No-label” learners, on the other hand, only showed a between category advantage on two continua (bared teeth-scream: $t(55) = -1.834$, $p = .036$; play-scream: $t(55) = -2.247$, $p = .015$). For the other four continua, “no-label” learners did not show a between category advantage (bared teeth-hoot: $t(55) = 2.155$, $p = 0.018^{10}$; bared teeth-play: $t(55) = 1.374$, $p = 0.09$; hoot-scream: $t(55) = -1.060$, $p = 0.147$; play-hoot: $t(55) = -0.359$, $p = 0.361$) (see Figure 7).

As in Experiment 1, both “no-label” and “label” learners discriminated many morph pairs above chance levels. The discrimination for between and within category pairs can be seen in SI2.

Overall advantage: As in Experiment 1, we also combined all six continua to test whether “label” and “no-label” learners showed an *overall* between category advantage. Using the same comparisons as in Experiment 1, the “no-label” learners were not more accurate at discriminating the between category morph pairs compared to within morph pairs ($t(5) = -0.574$, $p = 0.300$). However, “label” learners were marginally more accurate at discriminating the between category morph pairs compared to morph pairs ($t(5) = -1.696$, $p = 0.075$). Moreover, when we compared the between pairs with only the within category pairs at the tail ends of the continua (where morphs in the pair was identified most similarly), they was discriminated significantly more accurately ($t(83) = -2.700$, $p = 0.008$; $t(83) = -2.416$, $p =$

0.018). Importantly, this was not the case when we performed the same type of analysis for the “no-label” group ($t(83) = -0.463, p = 0.645, t(83) = 0.284, p = 0.777$). These results suggest that at a global level, only “label” learners showed CP.

Summary—To summarize both tasks in Experiment 2, “no-label” and “label” learners did not differ in the way they identified morphs, and identified significant structural changes at multiple points along each continuum. When we took into account largest change in identification, (coinciding with the Xc form the sigmoid function), only “label” learners showed a between category advantage on the majority (four of six) of continua. “No-label” learners showed the effect on two continua. This effect cannot be explained by the fact that labels enhanced learning, as the number of training attempts to reach above 95% accuracy did not differ between the label and no-label groups. In addition, the “label” learners, compared to the “no-label” learners, discriminated only a few more morph pairs above chance (by shear count) (see also SI2). When we combined all the continua, “no-label” learners did not an *overall* effect. “Label” learners showed a marginal *overall* effect; moreover, this overall effect was highly significant effect when we limited the comparison to only the within category pairs at the two ends of the continua. Our results suggest that, in the majority of the cases, conceptual knowledge in the form of verbal labeling supported CP.

Discussion

For many years, scientists have debated how a human perceiver sees emotion in a face. In this paper, we used a novel technique for evaluating the structural and conceptual hypotheses by having perceivers, who differ in their conceptual knowledge (in the form of expertise in Experiment 1 or verbal labeling learning in Experiment 2), categorize chimpanzee facial expressions. Chimpanzee expressions are morphologically similar to human expressions, but provide the advantage that human perceivers do not routinely have experience with these expressions. Thus, they allow a more controlled and precise test of the structural hypothesis. Across two experiments, we found little evidence to support the structural hypothesis, and more evidence to support the conceptual hypothesis. Specifically, participants who learned category members with a verbal label showed CP on four of the six expression continua. In comparison, participants who learned only category members without a label showed CP on two of the continua. In fact, when we combined all continua to assess an *overall* between category advantage, only “label” learners showed the between category advantage. Thus, at the most general level, only the “label” learners showed CP. Expertise (in the form of prolonged experience working with non-human primates), on the other hand, did not show much change in CP. Experts showed CP on three of the six continua, whereas novices showed the effect on two continua. Neither group showed an *overall* between category advantage.

It should not be surprising that words support categorical perception. Words have a powerful effect on a person’s ability to group together objects during the learning of a new category, even when the objects do not share perceptual features (Booth & Waxman 2003; Dewar & Xu, 2009; Fulkerson & Waxman, 2006; Waxman & Markow, 1995). Words direct an infant’s ability to categorize animals and objects by acting as “essence placeholders”, such that a word allows an infant to make inferences about a new object on the basis of prior experience with objects of the same kind (Xu, Cote, & Baker, 2005). Most recently, Plunkett, Hu, and Cohen (2008) showed that labels override perceptual categories, and even play a causal role in category membership in preverbal infants. In a recent review, Barrett, et al. (2007) summarized a number of different lines of evidence to support the idea that language is a key component of the conceptual knowledge hypothesis. Simply stated, accessible language provides an “internal context” that shapes emotion perception.

Why might it be that when novices (in Experiment 1) and “no-label” learners (in Experiment 2) showed a CP effect it was on different continua? One reason might be sampling differences such that a certain number of continua by chance alone would show significant effects. The reason might also be related to the training the “no-label” learners underwent prior to the CP tasks. Recall that the novices did not receive any training, but the “no-label” learners studied and learned unlabeled exemplars from each category. We did not attempt to combine samples from the two experiments as a result of these differences. Whatever the reasons, it is more important to note that the differences did not arise from one endpoint being rated as more distinctive than another. Recall that only pairs of endpoints containing the “hoot” expression were rated as less similar than those pairs which did not contain the “hoot” expression; however, the effects we found were not limited to (nor were they always present) for the hoot continua.

Why did the experts not show an overall effect of CP in Experiment 1? Given the findings of Experiment 2 (along with other supporting evidence, e.g., Barrett, et al., 2007; Russell, 1994), one possibility is that experts might not have been cued to bring to bear their considerable knowledge without the use of words in the identification task. This is in line with the conclusions other researchers have drawn on the role of expertise in CP (Goldstone, 1994; Levin & Beale, 2000). Thus, it appears that CP occurs when there is some explicit reference to language or context. Such reasoning is consistent with findings from Roberson, et al. (1999) who found that a patient with color anomia had difficulty sorting colors into groups but still showed a between category advantage once the process became automated (and therefore less dependent or independent on labeling). Their findings supported a category-adjustment model for the role of labeling on CP, in which labels are indirectly involved in CP (see Roberson, Damjanovic, & Pilling, 2007 for further details). Specifically, such an “adjustment” model suggests that within category pairs near the boundary should be discriminated less accurately than those within category pairs at the ends. In fact, our overall analyses (in which we found a significant between category advantage when we compared the between category pairs with only the within pairs created from the morphs near the continua ends, but only a marginal effect when we compared the between category pair to the average of all the within category trials) is in line with this argument.

In future studies we hope to address the scope that words have on CP more directly. Specifically, we are interested in assessing whether we can enhance CP when the learned labels for category anchors are present in the identification task. It is possible, that in the current study, “label” learners did not explicitly recall the learned labels since they were not used in the task. We believe that requiring participants to use verbal labels in the identification task would activate their conceptual knowledge. As previously mentioned, most studies showing CP use the category labels instead of the endpoint images in the identification task, which would force participants to explicitly access their conceptual knowledge (Calder, et al., 1996; Etcoff & Magee, 1992; Young et al., 1997).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Preparation of this manuscript was supported by the National Institutes of Health Director’s Pioneer Award (DP1OD003312), a grant from the National Science Foundation (BCS 0721260), and a contract with the Army Research Institute (W91WAW) to Lisa Feldman Barrett. We thank the Yerkes National Primate Center for original stimulus pictures. The authors would like to thank Jack McDowell and Spencer Lynn for statistical advice, and Colleen Madden and Katie Bessette for assistance with data collection and graphing, respectively. Portions of this data were presented at the biannual meeting of the International Society for Research on Emotion in 2009. The views, opinions, and/or findings contained in this article are solely those of the authors and should not be construed as an official Department of the Army or DOD position, policy, or decision.

References

- Angeli A, Davidoff J, Valentine T. Face familiarity, distinctiveness, and categorical perception. *The Quarterly Journal of Experimental Psychology* 2008;61(5):690–707.10.1080/17470210701399305 [PubMed: 17853232]
- Barrett LF. Emotions as natural kinds? *Perspectives on Psychological Science* 2006a;1:28–58.10.1111/j.1745-6916.2006.00003.x
- Barrett LF. Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review* 2006b;10:20–46.10.1207/s15327957pspr1001_2 [PubMed: 16430327]
- Barrett LF. The future of psychology: Connecting mind to brain. *Perspectives in Psychological Science* 2009;4:326–339.10.1111/j.1745-6924.2009.01134.x
- Barrett LF, Lindquist K, Gendron M. Language as context for the perception of emotion. *Trends in Cognitive Science* 2007;11:327–332.10.1016/j.tics.2007.06.003
- Beale JM, Keil FC. Categorical effects in the perception of faces. *Cognition* 1995;57(3):217–239.10.1016/0010-0277(95)00669-X [PubMed: 8556842]
- Booth AE, Waxman S. Object names and object functions serve as cues to categories in infants. *Developmental Psychology* 2003;38(6):948–957.10.1037//0012-1649.38.6.948 [PubMed: 12428706]
- Burrows AM, Waller BM, Parr LA, Bonar CJ. Muscles of facial expression in the chimpanzee: Descriptive, ecological, and phylogenetic contexts. *Journal of Anatomy* 2006;208(2):153–167.10.1111/j.1469-7580.2006.00523.x [PubMed: 16441560]
- Calder AJ, Young AW, Perrett DI, Etcoff NL, Rowland D. Categorical perception of morphed facial expressions. *Visual Cognition* 1996;3:81–117.10.1080/713756735
- Dewar K, Xu F. Do early nouns refer to kinds of distinct shapes? *Psychological Science* 2009;20(2):252–257.10.1111/j.1467-9280.2009.02278.x [PubMed: 19175526]
- Ekman P. Are there basic emotions? *Psychological Review* 1992;99(3):550–553.10.1037/0033-295X.99.3.550 [PubMed: 1344638]
- Etcoff NL, Magee JJ. Categorical perception of facial expressions. *Cognition* 1992;44:227–240.10.1016/0010-0277(92)90002-Y [PubMed: 1424493]
- Fulkerson AL, Haaf RA. The influence of labels, non-labeling sounds, and source of auditory input on 9- and 15-month olds' object categorization. *Infancy* 2003;4(3):349–369.10.1207/S15327078IN0403_03
- Fulkerson AL, Waxman SR. Words (but not tones) facilitate object categorization: Evidence from 6- and 12- month-olds. *Cognition* 2006;105:218–228.10.1016/j.cognition.2006.09.005 [PubMed: 17064677]
- Goldstone R. Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General* 1994;123:178–200.10.1234/12345678 [PubMed: 8014612]
- Harnad, S. *Categorical perception*. Cambridge: Cambridge University Press; 1987.
- Izard, CE. *The face of emotion*. New York: Appleton-Century-Crofts; 1971.
- Kelly DJ, Quinn PC, Slater AM, Lee K, Ge L, Pascalis O. The other-race effect develops during infancy. *Psychological Science* 2007;18:1084–1089.10.1111/j.1467-9280.2007.02029.x [PubMed: 18031416]
- Kitutani M, Roberson D, Hanley JR. What's in a name? Categorical perception for unfamiliar faces can occur through labeling. *Psychonomic Bulletin & Review* 2008;15(4):787–794.10.3758/PBR.15.4.787 [PubMed: 18792505]
- Levin DT, Beale JM. Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Perception & Psychophysics* 2000;62(2):386–401.10.3758/PBR.15.4.787 [PubMed: 10723217]
- McKone E, Martini P, Nakayama K. Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology: Human Perception and Performance* 2001;27(3):573–599. 10.1037//0096-1523.27.3.573. [PubMed: 11424647]
- Ozgen E, Davies IRL. Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General* 2002;131(4):477–493.10.1037/0096-3445.131.4.477 [PubMed: 12500859]

- Parr LA, Waller BM, Vick SJ, Bard K. Classifying chimpanzee facial expressions using muscle action. *Emotion* 2007;7(1):172–181.10.1037/1528-3542.7.1.172 [PubMed: 17352572]
- Pascalis O, de Hann M, Nelson CA. Is face processing species-specific during the first year of life? *Science* 2002;296:1321–1323.10.1126/science.1070223 [PubMed: 12016317]
- Pilling M, Wiggett A, Ozgen E, Davies IRL. Is color “categorical perception” really perceptual? *Memory & Cognition* 2003;3:538–551.
- Plunkett K, Hu J-F, Cohen LB. Labels can override perceptual categories in early infancy. *Cognition* 2008;106:665–681.10.1016/j.cognition.2007.04.003 [PubMed: 17512515]
- Preuschoft S, van Hooff JARAM. Homologizing primate facial displays: A critical review of the methods. *Folia Primatologica* 1995;65:121–137.10.1159/000156878
- Roberson D, Davidoff J. The categorical perception of color and facial expressions: The effect of verbal interference. *Memory and Cognition* 2000;28:977–986. apa.org.
- Roberson D, Davidoff J, Braisby N. Similarity and categorization: neuropsychological evidence for a dissociation in explicit categorization tasks. *Cognition* 1999;71:1–42.10.1016/S0010-0277(99)00013-X [PubMed: 10394708]
- Roberson, D.; Damjanovic, L.; Pilling, M. Categorical perception of facial expressions: Evidence for a “category adjustment” model; *Memory & Cognition*. 2007. p. 1814-1829.mc.psychonomic-journals.org
- Russell J. Is there universal recognition of emotion from facial expression - A review of the cross-cultural studies. *Psychological Bulletin* 1994;115(1):102–141.10.1037/0033-2909.115.1.102 [PubMed: 8202574]
- Scott L, Monesson A. The origin of biases in face perception. *Psychological Science* 2009;20(6):676–680. [PubMed: 19422630]
- Stevenage SV. Which twin are you? A demonstration of induced categorical perception of identical twin faces. *British Journal of Psychology* 1998;89:39–57. apa.org.
- Tomkins, SS. *Affect, imagery, consciousness. Vol.1: The positive affects*. New York: Springer; 1962.
- Viviani P, Binda P, Borsato T. Categorical perception of newly learned faces. *Visual Cognition* 2007;15:420–467.10.1080/13506280600761134
- Waller BM, et al. Intramuscular electrical stimulation of facial muscles in humans and chimpanzees: Duchene revisited and extended. *Emotion* 2006;6(3):367–382.10.1037/1528-3542.6.3.367 [PubMed: 16938079]
- Waxman SR, Braun I. Consistent (but not variable) names as invitations to form object categories: new evidence from 12-month-old infants. *Cognition* 2005;95(3):B59–B68.10.1016/j.cognition.2004.09.003 [PubMed: 15788158]
- Waxman SR, Markow DB. Words as invitations to form categories: Evidence from 12 to 13 month old infants. *Cognitive Psychology* 1995;29:257–302.10.1006/cogp.1995.1016 [PubMed: 8556847]
- Winawer J, Witthoft N, Frank MC, Wu L, Wade AR, Boroditsky L. Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Science* 2007;104:7780–7785.10.1073/pnas.0701644104
- Xu F, Cote A, Baker A. Labeling guides object individuation in 12-month old infants. *Psychological Science* 2005;16:372–377. [PubMed: 15869696]
- Young AW, et al. Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition* 1997;63:271–313.10.1016/S0010-0277(97)00003-6 [PubMed: 9265872]

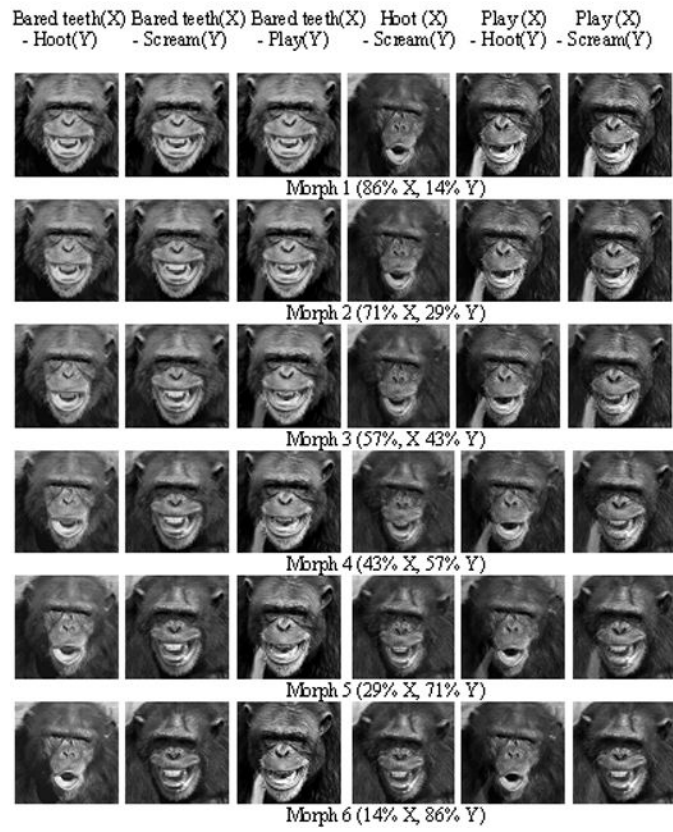


Figure 1.
Morphed Stimuli.

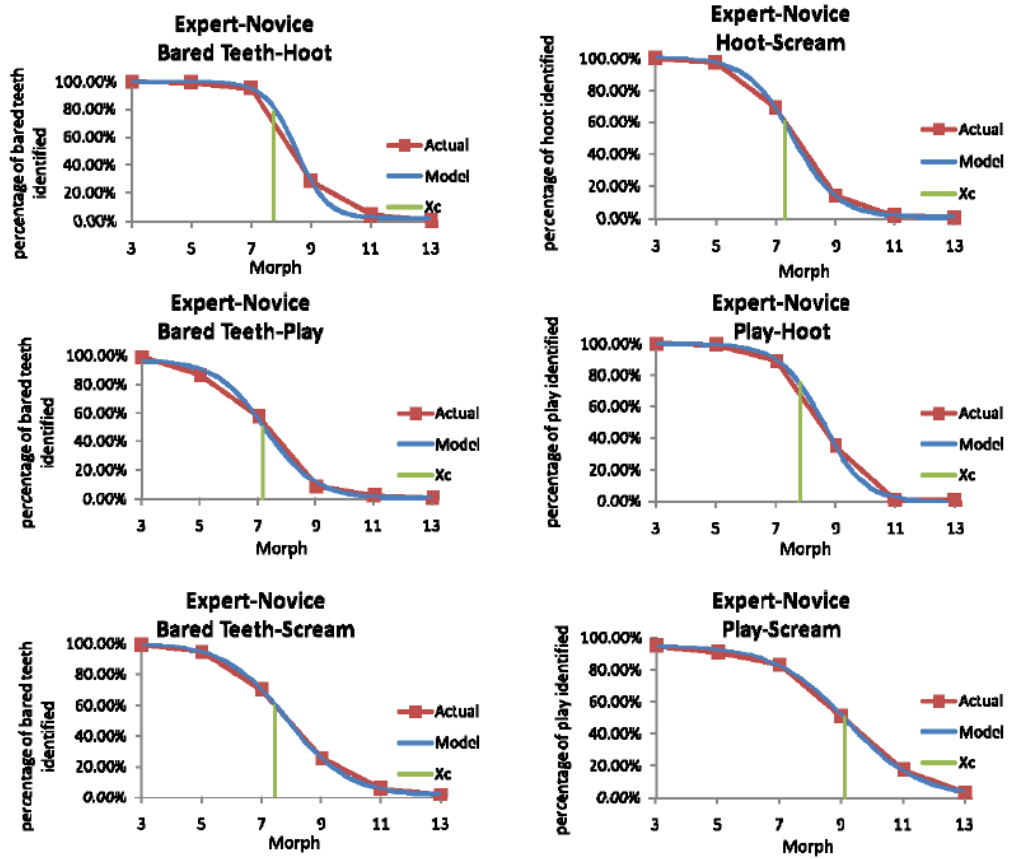


Figure 2. Identifications on Each Continuum for Experts/Novices Combined.

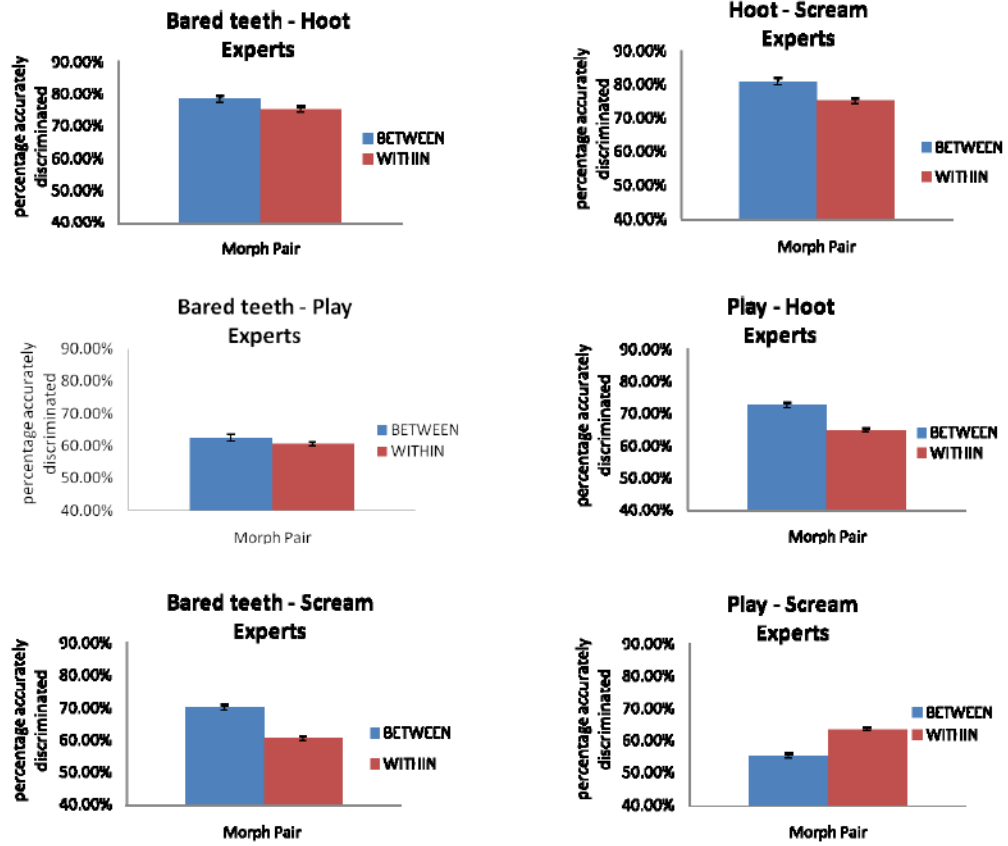


Figure 3.
Accuracy on Each Continuum for Experts.

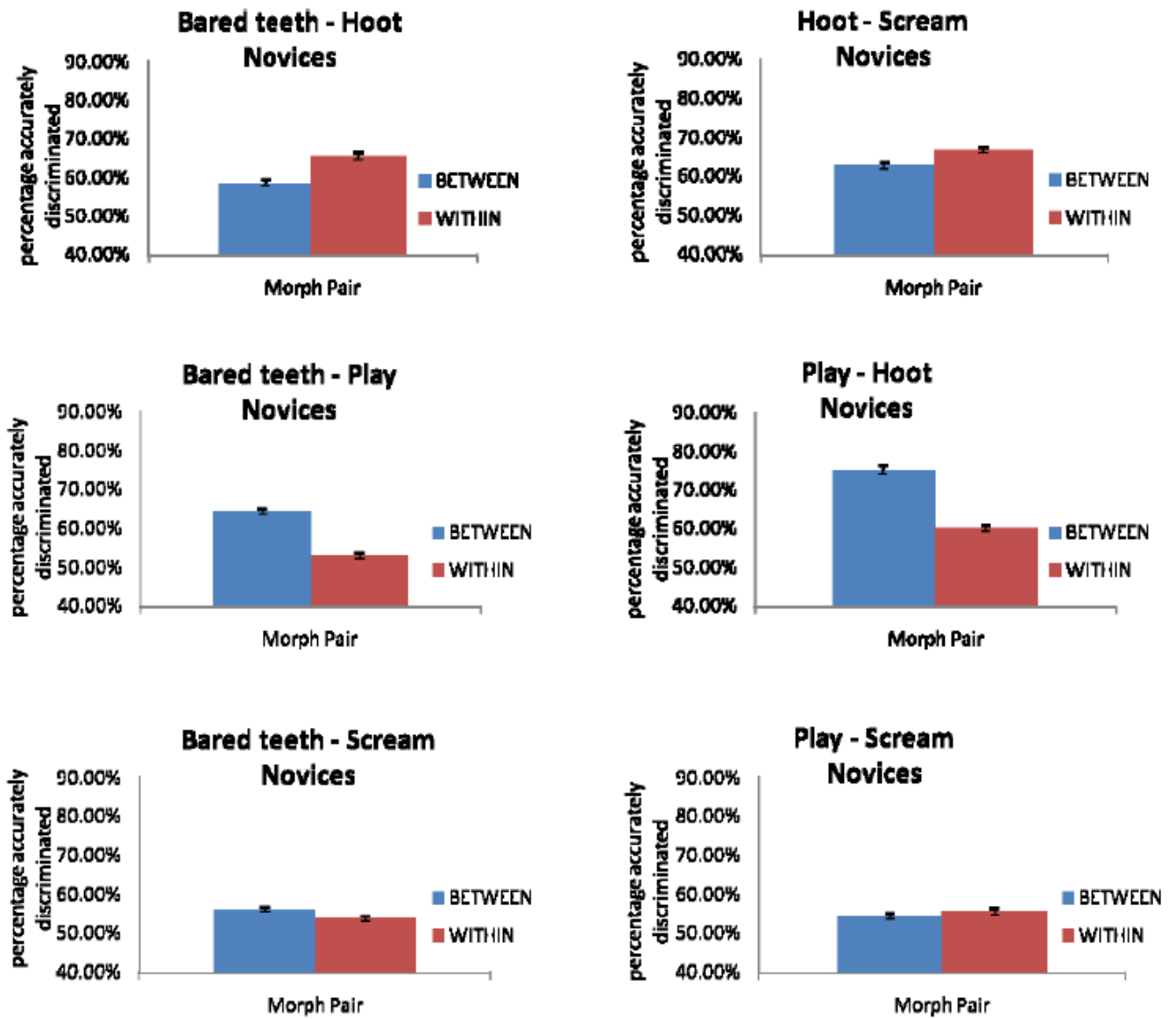


Figure 4.
Accuracy on Each Continuum for Novices.

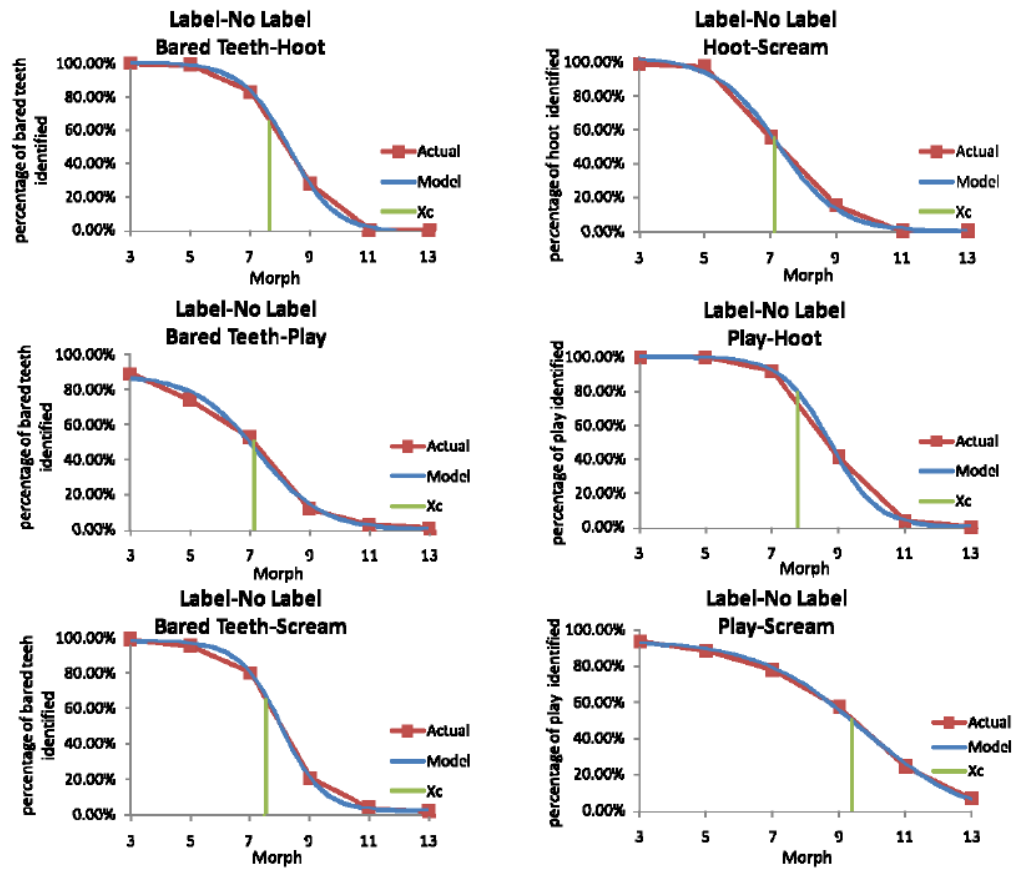


Figure 5. Identifications on Each Continuum for Label/No-Label Learners Combined.

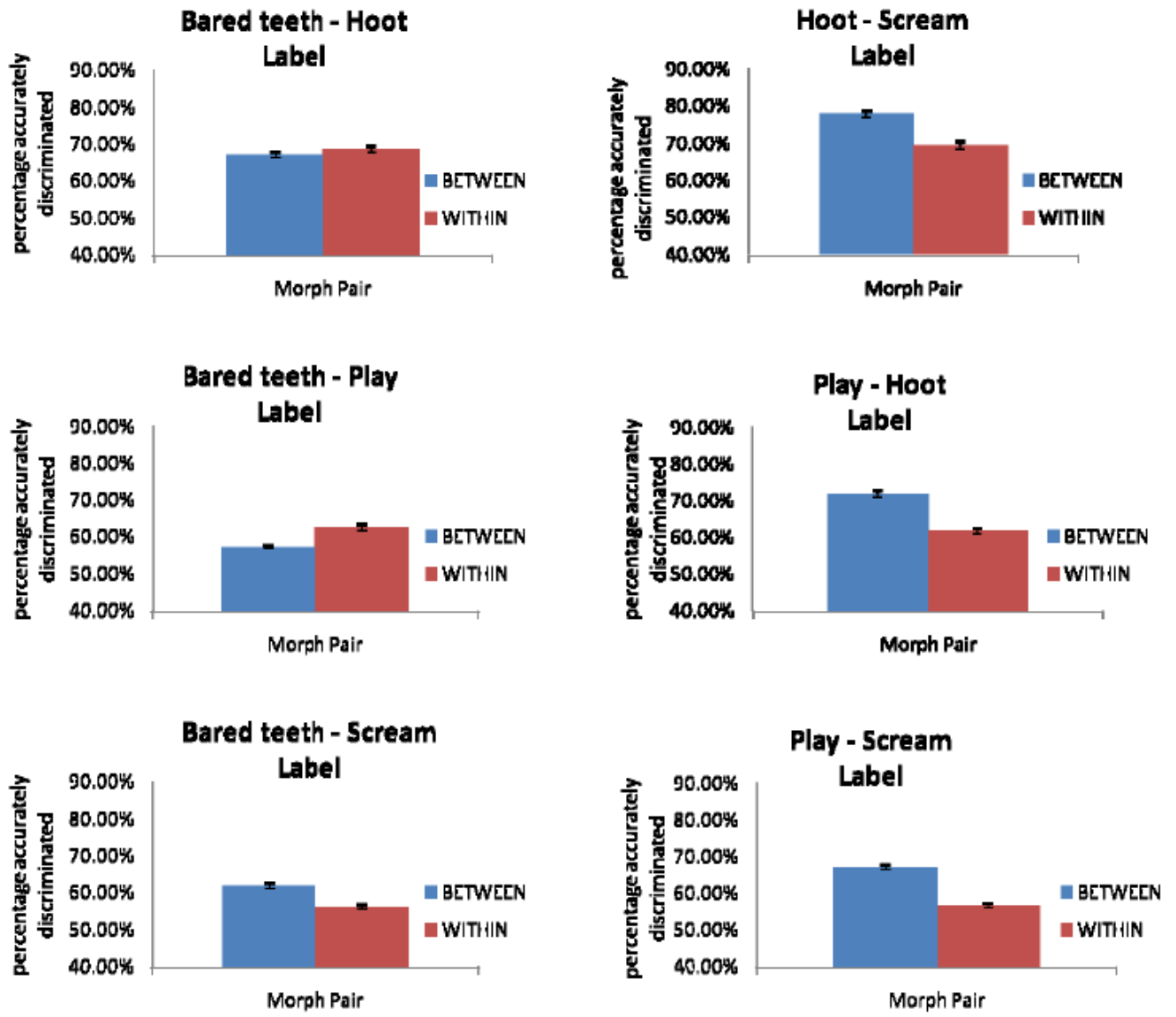


Figure 6. Accuracy on Each Continuum for “Label” Learners.

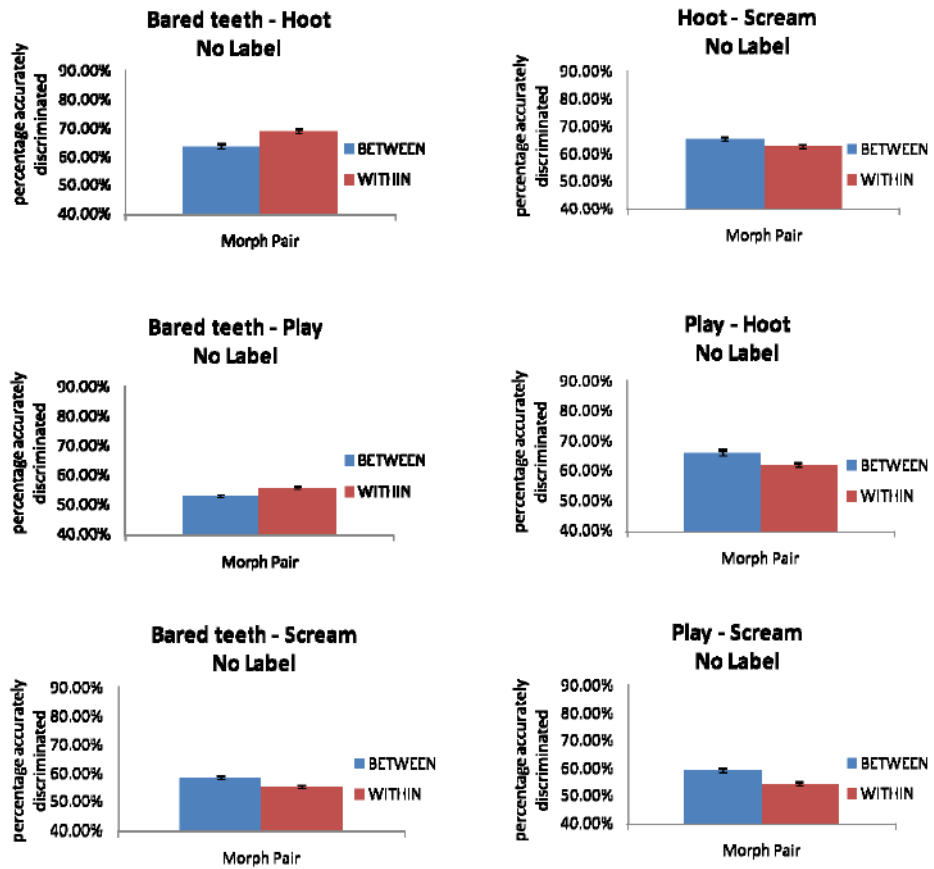


Figure 7. Accuracy on Each Continuum for “No-Label” Learners.

Table 1

Pearson's Correlations Between Identification Task and Three Equation Models

Experts/Novices	r ² linear	r ² logistic	r ² step	"Label"/"No-label" Learners	r ² linear	r ² logistic	r ² step
Bared teeth-hoot	.871	> .999	.959	Bared teeth-hoot	.905	> .999	.939
Bared teeth-play	.913	.998	.903	Bared teeth-play	.926	.997	.891
Bared teeth-scream	.939	> .999	.915	Bared teeth-scream	.900	> .999	.960
Hoot-scream	.904	> .999	.937	Hoot-scream	.914	.999	.870
Play-hoot	.894	> .999	.925	Play-hoot	.895	> .999	.829
Play-scream	.931	> .999	.838	Play-scream	.943	.999	.773