Emotion
2010, Vol. 10, No. 4, 544–554

1528-3542/10/$12.00 DOI: 10.1037/a0019017

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

Jennifer M. B. Fugate
Boston College

Harold Gouzoules
Emory University

Lisa Feldman Barrett
Boston College, and Massachusetts General Hospital, Charlestown, MA

Categorical perception (CP) occurs when 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 not facilitate CP. Only when perceivers first explicitly learned facial expression categories 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 facilitated by verbal labeling.

Keywords: facial expressions, emotion, categorical perception, verbal labeling, expertise

Supplemental materials: http://dx.doi.org/10.1037/a0019017.supp

Categorical perception (CP) occurs when stimuli that continuously vary are perceived as belonging to discrete categories. The two categories are marked by a sharp boundary (Harnad, 1987). As a result, perceivers are more accurate at discriminating between stimuli placed into different categories than between stimuli placed within the same category (i.e., a between-category advantage; Goldstone, 1994). 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 automatically "recognize" emotion (the *structural hypothesis*; Ekman, 1992; Izard, 1971; Tomkins, 1962). In this view, perceivers are merely translating the information about emotion that is carried in the facial movements of the sender. The idea that people easily and effortlessly perceive anger, sadness, or fear in another person's face is supported by studies which demonstrate CP for facial depictions of emotion (Calder, Young, Perrett, Etcoff, & Rowland, 1996; Etcoff & Magee, 1992; Young, Rowland, Calder, Etcoff, Seth, & Perrett, 1997). Alternatively, the categorization that occurs in emotion perception might arise from 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 judgment, 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; human facial movements are seen and interpreted as emotional 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 discrete emotion category judgments that people make about facial movements. 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 similar looking facial movements (Burrows, Waller, Parr, & Bonar, 2006; Parr, Waller, Vick, & Bard, 2007; Waller et al., 2006). In two studies, we tested the structural and conceptual hypotheses of CP. Specifically, we examined whether conceptual knowledge in the form of expertise

Jennifer M. B. Fugate, Department of Psychology, Boston College; Harold Gouzoules, Department of Psychology, Emory University; Lisa Feldman Barrett, Department of Psychology, Boston College, and Department of Psychiatry and the Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA.

Preparation of this article was supported by National Institutes of Health Director's Pioneer Award DP1OD003312 and Grant BCS 0721260 from the National Science Foundation to Lisa Feldman Barrett. We thank the Yerkes National Primate Center for original stimulus pictures. We thank Jack McDowell and Spencer Lynn for statistical advise and Colleen Madden and Katie Bessette for assistance with data collection and graphing, respectively. Portions of this data were presented at the meeting of the International Society for Research on Emotion in 2009.

Correspondence concerning this article should be addressed to Jennifer M. B. Fugate, Department of Psychology, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467. E-mail: fugatej@bc.edu

(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 use their expertise with faces to make effortlessly categorical judgments comes from experiments that manipulate identity. People easily show CP for familiar faces (e.g., Beale & Keil, 1995), but do not typically exhibit CP for unfamiliar faces. 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).[1]

One important facet of CP is that a perceiver does not lose the ability to discriminate among category exemplars. Rather, people are able to ignore differences across exemplars that are not psychologically meaningful to category membership. Said another way, CP results from cross-category expansion, not from within-category compression (Ozgen & Davies, 2002). This effect 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).[2]

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 (Barrett, 2006a, 2006b, 2009). In such cases, is repeated exposure (expertise) enough for a perceiver to understand which differences are psychologically meaningful from those differences that are not?

Verbal Labeling

A growing body of research suggests that categories, especially those with tremendous variability, are learned and anchored by a verbal label, such as a word (Booth & Waxman, 2003; Dewar & Xu, 2009; Fulkerson & Waxman, 2006; Waxman & Braun, 2005; Waxman & Markow, 1995). It is possible that applying the same word to physically different exemplars might be enough for a perceiver to arrive at discrete, psychologically meaningful categories. For example, Kikutani, Roberson, and Hanley (2008) presented perceivers with unfamiliar faces (e.g., two different identities), either by themselves or paired with a label, during a familiarization task completed prior to a CP experiment. Only when participants were exposed to the unfamiliar 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 facilitate CP has also been proposed in other domains, such as emotion and color (Pilling, Wiggett, Ozgen, & Davies, 2003; Roberson & Davidoff, 2000; Winawer, Witthoft, Frank, Wu, Wade, & Boroditsky, 2007). Consistent with this account, CP for human facial depictions of emotion and colors is eliminated when a secondary task is employed that disrupts verbal processing (Roberson & Davidoff, 2000).

Figure 1. Morphed stimuli. Reprinted with permission of Frans B. M. Waal.

The Present Experiments

In Experiment 1, we tested whether nonhuman primate "experts" or "novices" showed CP for chimpanzee facial expressions.[3] If the structural hypothesis is correct and the structural information in the face is sufficient for CP to occur, then both groups should show CP. If only experts, however, 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

[1] In some cases, people do show CP for unfamiliar faces without extensive training, however, the effect occurs when one category anchor is very distinctive (see Angeli, Davidoff, & Valentine, 2008), or when the task requires that category anchors be held in memory which can lead to a distortion of face space (see Angeli et al., 2008; McKone et al., 2001).

[2] Interestingly, Scott and Monesson (2009) showed that exposing infants to individually labeled faces prevented perceptual narrowing, whereas exposing infants to unlabeled or shared labeled faces did not prevent perceptual narrowing.

[3] Experts were defined as people who had worked with at least one nonhuman primate species consecutively for at least 12 months any time during the past 6 years and had some familiarity with the behaviors, including the facial expressions, of the species. Novices were defined as people who had never worked with any nonhuman primate species and had no formal training in nonhuman primate behaviors, including facial expressions.

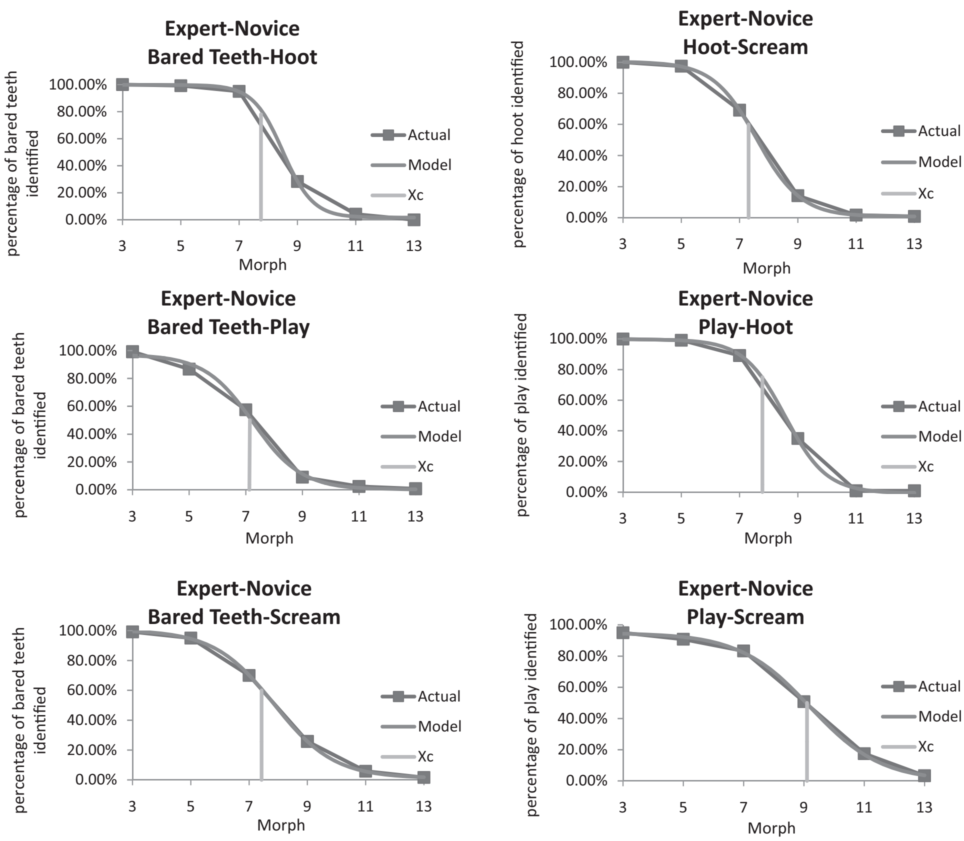

Figure 2. Identifications on each continuum for experts and novices combined.

("label" learners) or without a verbal label ("no-label" learners) showed CP. Once again, if the structural hypothesis is correct, then both groups of participants, regardless of training, should show CP. If only the label learners, however, show evidence of CP or CP increases significantly compared with the 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). Participants in each task viewed morphed faces created from pairs of photographs each depicting a different chimpanzee facial expression (bared teeth face, hoot face, scream face, and play face). Six morphs were created between each pair (X–Y) of facial expressions (Morph 1, 86%X–14%Y; Morph 2, 71%X–29%Y; Morph 3, 57%X–43%Y, Morph 4, 43%X–57%Y; Morph 5, 29%X–71%Y; Morph 6, 14%X–86%Y) (see Figure 1). In the *identification task*, participants saw each morph along with the two photographs from which the morph was created (e.g., category anchors) as comparison images. Participants judged whether the morph was more like one comparison image or the other. If participants are capable of showing CP, the number of identifications given to morphs along a X-Y continuum of facial expressions should shift abruptly, creating a *category boundary* (Harnad, 1987). In the AB-X *discrimination task,* participants first saw two morphed faces from the same continuum (Face A followed by Face B) that differed by one incremental step from each other. Next, either A or B was reshown (X), and participants judged whether this face was the same as face A or B.[4] If participants show CP, they should be more accurate in their judgments when A and B cross the category boundary compared with when A and B are from the same category, a *between-category advantage* (Goldstone, 1994).

Method

Participants

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

[4] There are several ways to assess a between-category advantage (e.g., AB-X, similarity, and better likeness; see McKone et al., 2001). We chose an AB-X task for several reasons: (a) It is widely used; (b) it did not require showing category anchors; (c) it did not require holding the images in memory for more than 2 s.

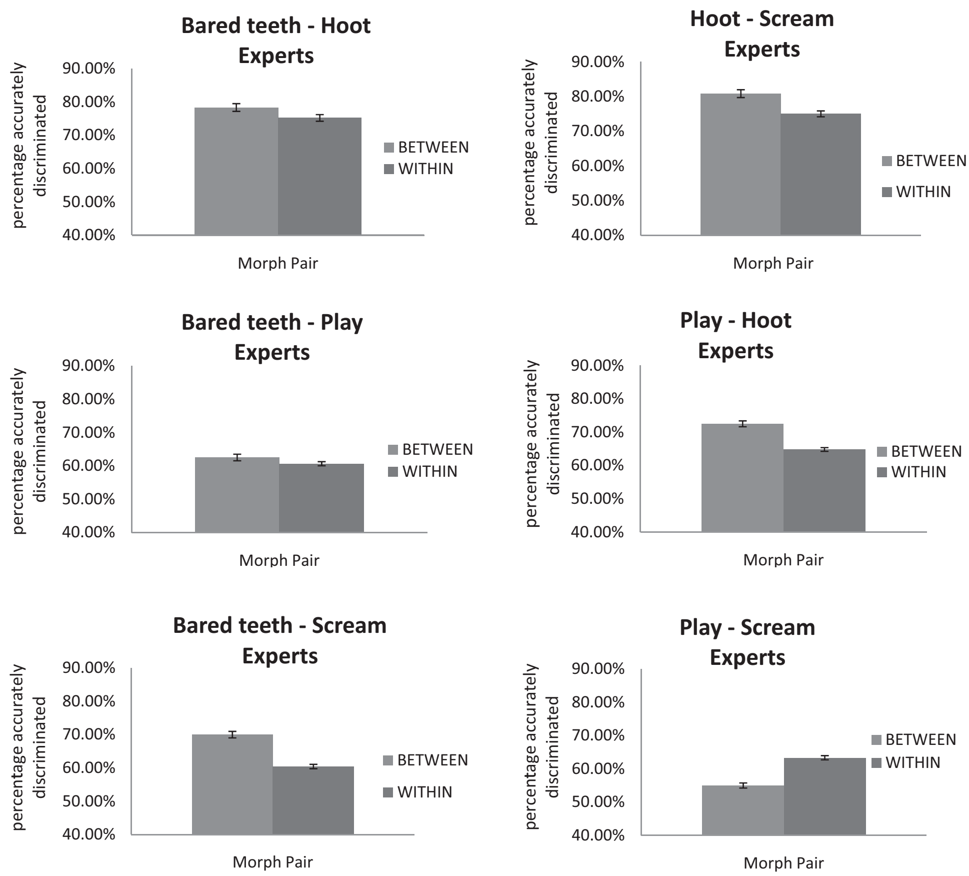

Figure 3. Accuracy on each continuum for experts.

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 photographs, each of which depicted a different, prototypical chimpanzee facial expression (bared teeth face, hoot face, scream face, and play face). Each of the four photographs was previously rated by three 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 (courtesy F. deWaal).

In addition, a separate sample of undergraduates unfamiliar with chimpanzee facial expressions rated each pair of photographs on their similarity to one another in order to assess whether any effects could be attributed to the distinctiveness of one photograph 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 less similar than pairs that did not include the hoot expression (in which the similarity ratings among these pairs did not differ). Although participants judged pairs that did not include the hoot expression as more similar, the two photographs were not judged as identical. In addition, participants in the actual experiments had no trouble discriminating between photographs as each served as a comparison image in the identification task.

Procedure

Participants completed the discrimination task, followed by the identification task. In the discrimination task, all three morphs were sequentially shown in the center of the computer 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 or the *2* key to select the second image. Participants discriminated all AB morph pairs 8 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 centered, at the top of the screen. Participants pressed the space bar to activate two comparison images displayed at the bottom of the screen and to the right and left of the initial morph. Participants pressed the *k* key to indicate the photograph on the right or the *d* key to indicate the photograph on left. Participants identified all morphs 4 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 prior to completing the discrimination and identification tasks. Both groups of participants were told that they would undergo training in which their goal was to learn the

Table 1
Pearson's Correlations Between Data From the Identification Task and Three Models

Morph	Experts/novices			Label/no-label learners		
	r^2	r^2	r^2	r^2	r^2	r^2
Bared teeth−hoot	.871	>.999	.959	.905	>.999	.939
Bared teeth−play	.913	.998	.903	.926	.997	.891
Bared teeth−scream	.939	>.999	.915	.900	>.999	.960
Hoot−scream	.904	>.999	.937	.914	.999	.870
Play−hoot	.894	>.999	.925	.895	>.999	.829
Play−scream	.931	>.999	.838	.943	.999	.773
Function	Linear	Logistic	Step	Linear	Logistic	Step

different categories of chimpanzee facial expressions. One group, no-label learners, was trained by viewing four exemplars from each of the categories (not the original photographs). The other group, label learners, saw the same exemplars as the no-label learners, but each face was presented with a category 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 in both groups could view each image for as long as they wished; images were advanced by pressing the space bar.

Participants then completed an assessment to test whether they learned the facial expression categories. Label learners were shown a label centered, at the top of the screen along with two of the previously seen faces as comparisons presented at the bottom of the screen to the right and left of the label. No-label learners 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 to emphasize learning the category labels. Participants in both groups pressed the *k* key to indicate the comparison on the right and the *d* key to indicate the one on the left. Participants were required to achieve at least 95% 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 percent of participants passed the assessment the first time (maximum = 4). The average number of times required to pass the assessment did not vary between the two groups.

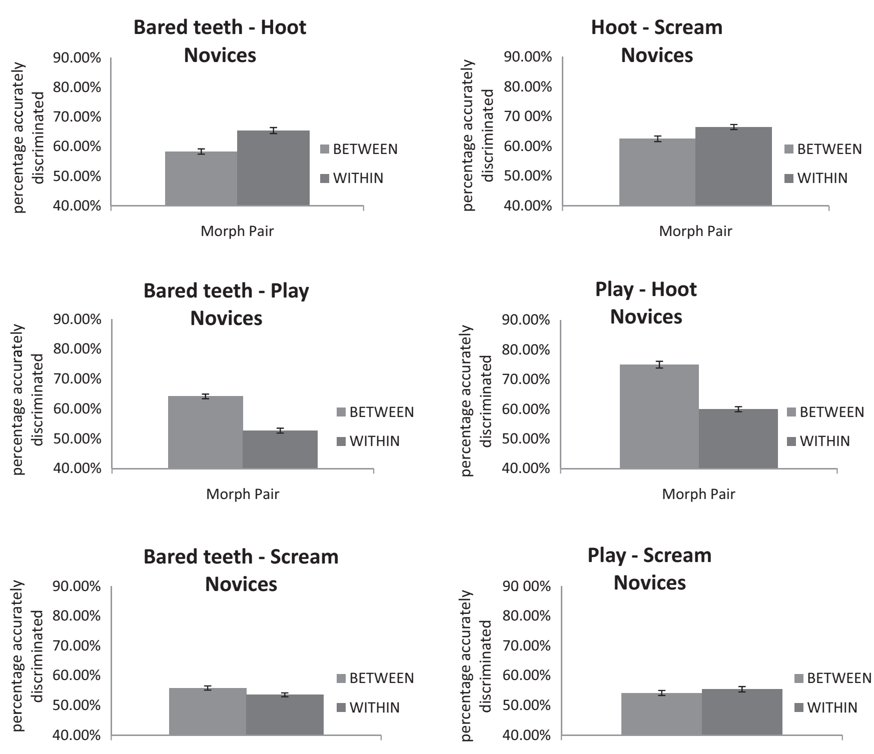

Figure 4. Accuracy on each continuum for novices.

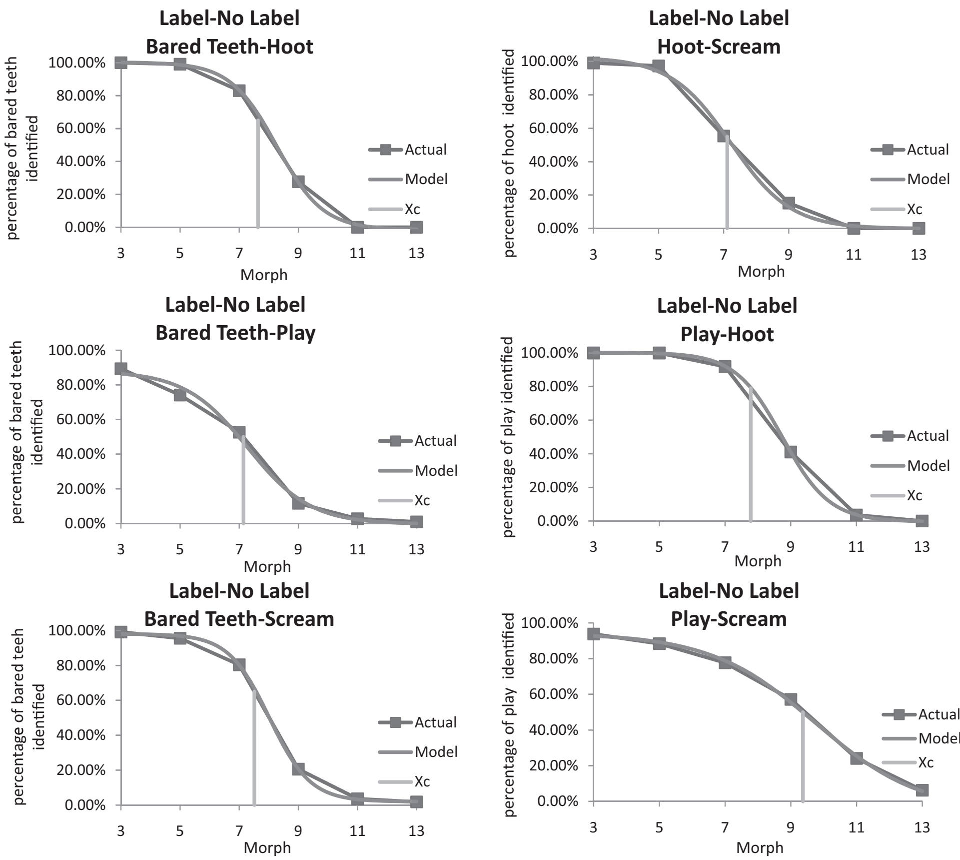

Figure 5. Identifications on each continuum for label and no-label learners combined.

Results

Experiment 1

Identification task. A repeated measures analyses of variance (ANOVA) was conducted on the number of times each morph was identified as one of the comparison images in an expression continuum. Participant group (novice or expert) was entered as a between-subjects factor. Experts and novices showed no difference in the way they identified any morphs on any continuum. Therefore, we recalculated the repeated measures ANOVAs without a between-subjects factor. Not surprisingly, the effect was significant for each continuum, suggesting that participants identified morphs as less like a comparison image as the amount of the same expression decreased (see Figure 2): bared teeth–hoot, $F(5, 145) = 375.524$; bared teeth–play, $F(5, 145) = 246.769$; bared teeth–scream, $F(5, 145) = 195.034$; hoot–scream, $F(5, 145) = 347.768$; play–hoot, $F(5, 145) = 362.522$; play–scream, $F(5, 145) = 140.280$; $ps < .001$.

To determine whether the shift in identification among morphs in a continuum was linear or more categorical, we determined whether a step, logistic, or linear function best fit the data. A *strong* categorical shift in perception would be best fit by a step-like function, in which all morphs up to a certain point would be identified as similar to one comparison image, and all morphs exceeding that point would be identified as similar to the other comparison image. A *weak* categorical shift in perception would be best fit by a logistic function, in which the transition from one category to the other might be less defined.[5]

We found that every continuum was best fit with a logistic function (see Table 1), and the pattern of residuals produced a random pattern. From the function, we calculated the center point, *Xc*, which would be the category boundary (McKone et al., 2001) (see Figure 2). Logistic fits were consistent with our finding that there were multiple pairs of adjacent morphs that were identified as significantly different from one another on every continuum (see SI1). Using *Xc* from the model or the largest difference between identifications given to adjacent morphs produced the same category boundary in all but one case (play-scream). In this case, *Xc* was nearly on one of the morphs, so we used the largest difference in identification as the boundary.

Discrimination task. To assess whether experts and/or novices showed a between-category advantage, we performed one-sample *t* tests comparing the mean accuracies of the within-category pairs with the mean accuracy of the between-category pair. A separate *t* test was performed for each continuum. Experts showed a between-category

[5] Although debate surrounds the ideal expected shape (see Harnad, 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).

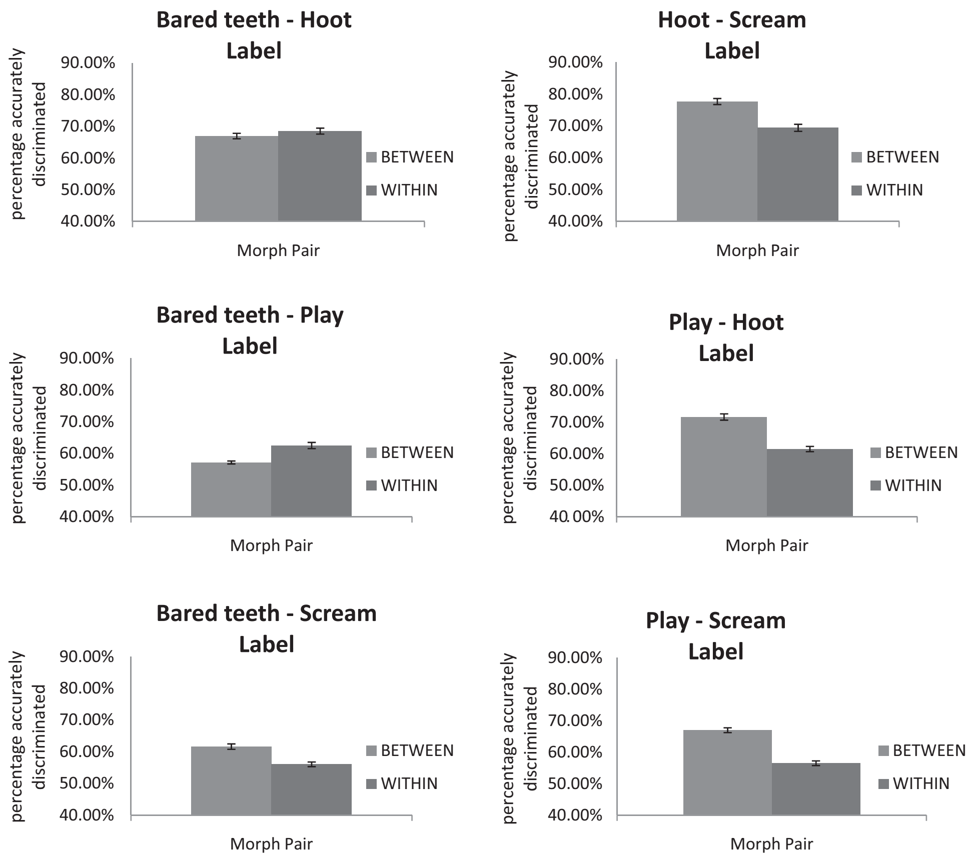

Figure 6. Accuracy on each continuum for label learners.

advantage on three of the continua (see Figure 3): bared teeth–scream, $t(59) = -4.229$, $p < .001$; hoot–scream, $t(59) = -2.512$, $p = .008$; play–hoot, $t(59) = -3.651$, $p < .001$. Experts did not show a between-category advantage on the other three continua: bared teeth–hoot, $t(59) = -1.308$, $p = .098$; bared teeth–play, $t(59) = -0.745$, $p = .230$; play–scream, $t(59) = 3.566$, $p < .001$.[6] Novices, on the other hand, showed a between-category advantage on two of the continua (see Figure 4): bared teeth—play, $t(59) = -4.575$, $p < .001$; play–hoot, $t(59) = -6.177$, $p < .001$. Novices did not show a between-category advantage on the other four continua, or the effect was in the reverse direction: bared teeth–hoot, $t(59) = 2.712$, $p = .005$; bared teeth–scream, $t(59) = -1.187$, $p = .120$; hoot–scream, $t(59) = 1.788$, $p = .40$; play–scream, $t(59) = 0.487$, $p = .314$.

Both experts and novices were able to discriminate many morph pairs above chance (not just the between-category pairs; see SI2). This finding is not uncommon given that within-category discrimination is typically not compromised in CP. Rather, a between-category advantage arises from an expansion of differences that distinguish the categories. Thus, a failure to show a between-category advantage in the presence of above chance discrimination suggests that participants did not know which differences mapped to category membership.

Overall advantage. We also combined all six expression continua to test whether experts and/or novices showed an *overall* between-category advantage. We performed a paired t test in which we compared the averaged mean accuracy of the within-category pairs with the mean accuracy of the between-category pair on each continuum in the same analysis. When we did this, neither experts nor novices were more accurate at discriminating the between-category pairs compared with the within–category pairs overall: experts, $t(5) = -1.2373$, $p = .130$; novices, $t(5) = -0.767$, $p = .478$. It might be possible, however, that experts and/or novices would show an overall between-category advantage if we limited the within-category pairs to only those at the continua ends (in which the morphs in the pair were identified most similarly). Even with this more liberal analysis, experts were still no more accurate at discriminating the between-category pairs compared with within-category pairs from either of the two ends across continua, $t(89) = -1.653$, $p = .102$; $t(89) = -1.033$, $p = .305$. Using the same comparisons, novices were also no more accurate at discriminating the between-category pairs compared with within-category pairs from either of the two ends across continua, $t(89) = -0.763$, $p = .448$; $t(89) = -0.100$, $p = .920$. 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, with both groups identifying structural changes at multiple points along each continuum. Novices and experts,

[6] The test is significant but in the direction opposite to predictions, such that participants were more accurate on the within-category pairs than the between-category pairs.

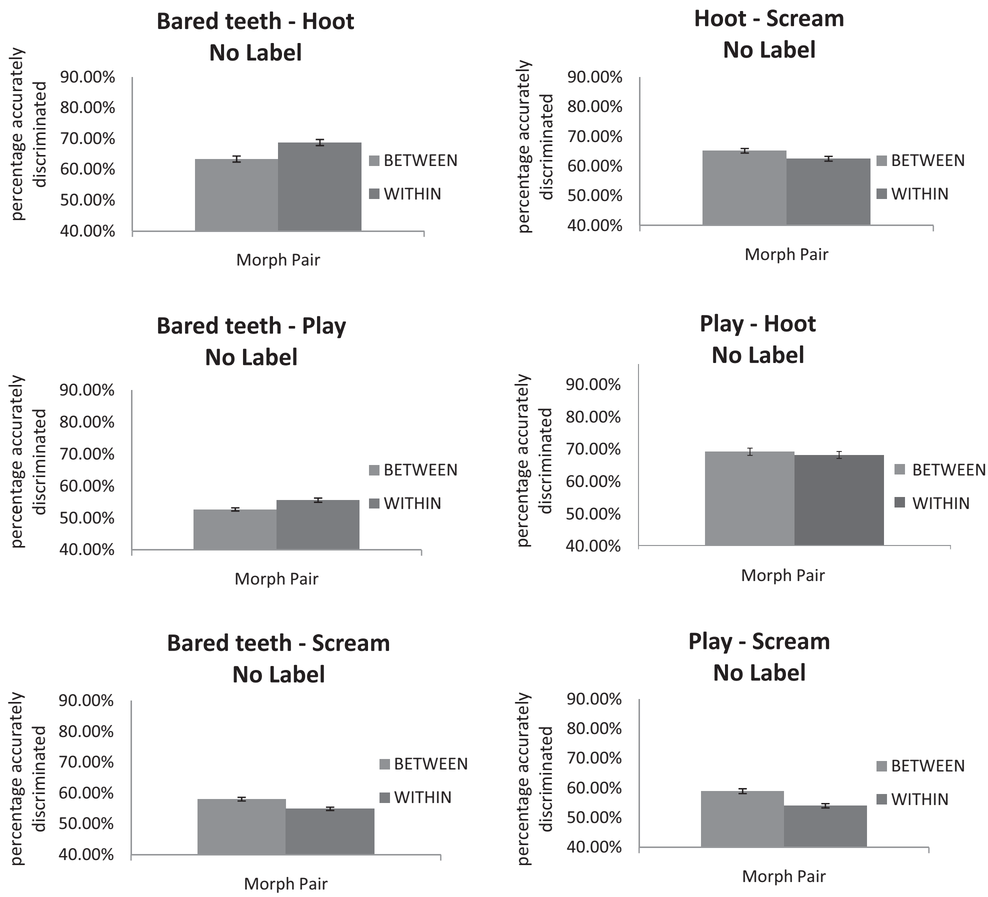

Figure 7. Accuracy on each continuum for no-label learners.

although being able to detect differences among most of the morphs, failed to understand which changes mapped on to category membership for most of the continua. When we took into account the largest of these identification changes (which coincided with *Xc*), experts and novices showed a between-category advantage for three and two continua, respectively (both groups showed CP on the play–hoot continuum). When we combined all six continua, neither experts nor 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 in the majority of cases for CP. Additionally, conceptual knowledge in the form of expertise did not markedly improve CP.

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 way they identified any morphs on any of the continua. Therefore, we combined the two groups and recalculated the repeated measures ANOVA for each expression continuum separately. Again, not surprisingly, the overall effect was highly significant for each continuum, suggesting that participants identified morphs as less like a comparison image as the amount of that same expression decreased (see Figure 5): 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, $(5, 135) = 100.743$; $ps < .001$.

As in Experiment 1, we found that every continuum was best fit with a logistic function (see Table 1), and the pattern of residuals produced a random pattern. A logistic shape was consistent with the finding that for every continua, there were multiple pairs of adjacent morphs that were identified as significantly different from one another (see SI1).

Next, we calculated X*c* for each continua (see Figure 5). In all but one case (hoot–scream), *Xc* fell between the two morphs that were judged as maximally different (see SI1). In this case, as in Experiment 1, we used the largest difference in identification between two adjacent morphs as the boundary.

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 (see Figure 6): bared teeth–scream, $t(55) = -2.385$, $p = .011$; hoot–scream, $t(55) = -2.957$, $p = .003$; play—hoot, $t(55) = -2.908$, $p = .003$; play—scream, $t(55) = -4.704$, $p < .001$. For the other two continua, label learners showed no between-category advantage or showed a significant effect in the reverse direction: bared teeth–hoot, $t(55) = 0.614$, $p = .271$; bared teeth–play, $t(55) = 2.011$, $p = .025$. No-label learners, on the other hand, only showed a between-category advantage on two continua (see Figure 7): 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 or showed a significant effect in the reverse direction: bared teeth–hoot, $t(55) = 2.155$, $p = .018$; bared teeth–play, $t(55) = 1.374$, $p = .09$; hoot–scream, $t(55) = -1.060$, $p = .147$; play–hoot, $t(55) = -0.359$, $p = .361$.

As in Experiment 1, both no-label and label learners discriminated many morph pairs above chance levels. The discrimination for between-category pairs 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/or 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 pairs compared with within-category pairs, $t(5) = -0.574$, $p = .300$. Label learners, however, were marginally more accurate at discriminating the between-category pairs compared with within-category pairs, $t(5) = -1.696$, $p = .075$. Moreover, when we compared the between-category pairs with only the within-category pairs from either of the two ends across continua, label learners showed the between-category advantage, $t(83) = -2.700$, $p = .008$; $t(83) = -2.416$, $p = .018$. Importantly, this was not the case when we performed the same analysis for the no-label group, $t(83) = -0.463$, $p = .645$; $t(83) = 0.284$, $p = .777$. These results suggest that at a global level only label learners showed CP.

Summary. To summarize findings across both tasks, 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 the largest change in identification (coinciding with the *Xc*), 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 at least 95% accuracy did not differ between the label and no-label groups. In addition, the label learners, compared with the no-label learners, discriminated only a few more morph pairs above chance (by shear count; see SI2). When we combined all the continua, no-label learners did not show an overall effect. Label learners showed a marginal overall effect; moreover, this effect was highly significant when we limited the comparison to only the within-category pairs from both ends. Our results suggest that, in the majority of the cases, conceptual knowledge in the form of verbal labeling facilitated CP.

Discussion

For many years, scientists have debated how a human perceiver sees emotion in a face. Many studies have shown CP for human facial depictions of emotion (Calder et al., 1996; Etcoff & Magee, 1992; Young et al., 1997). The typical conclusion from these studies is that the structural information in the face is sufficient for CP. In the current study, we used a novel technique for evaluating the structural and conceptual hypotheses of emotion perception by having perceivers, who differed in their conceptual knowledge (in the form of expertise in Experiment 1 or verbal labeling learning in Experiment 2), categorize chimpanzee facial expressions. Chimpanzee facial expressions are morphologically similar to human facial depictions of emotion, but provide the advantage that human perceivers do not routinely have experience with these faces. 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 one variant of the conceptual hypothesis. Specifically, participants who learned category members with a verbal label showed a between-category advantage on four of the six expression continua. In comparison, participants who learned only category members without a label showed a between-category advantage on two of the continua. Moreover, when we combined all continua to assess an overall between-category advantage, only label learners showed the effect. Expertise (in the form of prolonged experience working with nonhuman primates), on the other hand, did not enhance CP. Experts showed a between-category advantage on three of the six continua, whereas novices showed the effect on two continua. Neither the experts or the novices showed an overall between-category advantage.

Rather than suggesting that the structural information in the face is sufficient for CP, the results of the current studies suggest that CP will occur even when the structural information is on its own insufficient but a label has been learned and applied. Since all adult humans (barring organic disturbances) have separate words for discrete emotion categories (e.g. *anger, sad, fear, etc.*) it is not surprising that many previous studies using human facial depictions of emotion have found CP, especially when the emotion words are used in the task (as is almost always the case). The results of our current study suggest, however, that even if emotion words did not appear in the task, but as long as their presence was brought to bear, they would facilitate CP. This suggestion is supported by a series of experiments that show words also affect perceptual emotion judgments in which words themselves are not required to solve the task (Lindquist et al., 2006; Fugate & Barrett, in preparation).

It should not be surprising that words support categorical perception. Words have a powerful effect on a child's ability to group 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). More 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 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 CP 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 prior to the CP tasks, 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 important to note that the differences did not arise from one photograph being rated as more distinctive than another. Recall that pairs of photographs containing the hoot expression were rated as less similar than those pairs that did not contain the hoot expression; however, the effects we found were not limited to (nor where 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 conceptual knowledge.

Whether or not language causes CP or simply supports it is still a matter of debate. For example, a patient with color anomia who had difficulty sorting colors into groups still showed a between-category advantage once the process became automated (and therefore less dependent on or independent of labeling) (Roberson, Davidoff, & Braisby, 1999). Roberson, Damjanovic, and Pilling (2007) proposed a category "adjustment" model in which labels shape already existing groupings. 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 of the continuum. 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-category pairs at the continua ends, but only a marginal effect when we compared the between-category pairs with the average of all the within-category pairs) is in line with this argument.

In future studies we hope to address the effect that words have on CP more directly. Specifically, we are interested in assessing whether we can further enhance CP when the learned labels appear in the task.

References

Angeli, A., Davidoff, J., & Valentine, T. (2008). Face familiarity, distinctiveness, and categorical perception. *Quarterly Journal of Experimental Psychology, 61,* 690–707. doi:10.1080/17470210701399305

Barrett, L. F. (2006a). Emotions as natural kinds? *Perspectives on Psychological Science, 1,* 28–58. doi:10.1111/j.1745–6916.2006.00003.x

Barrett, L. F. (2006b). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10,* 20–46. doi:10.1207/s15327957pspr1001_2

Barrett, L. F. (2009). The future of psychology: Connecting mind to brain. *Perspectives in Psychological Science, 4,* 326–339. doi:10.1111/j.1745–6924.2009.01134.x

Barrett, L. F., Lindquist, K., & Gendron, M. (2007). Language as context for the perception of emotion. *Trends in Cognitive Science, 11,* 327–332. doi:10.1016/j.tics.2007.06.003

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition, 57,* 217–239. doi:10.1016/0010–0277(95)00669-X

Booth, A. E., & Waxman, S. (2003). Object names and object functions serve as cues to categories in infants. *Developmental Psychology, 38,* 948–957. doi:10.1037//0012–1649.38.6.948

Burrows, A. M., Waller, B. M., Parr, L. A., & Bonar, C. J. (2006). Muscles of facial expression in the chimpanzee: Descriptive, ecological, and phylogenic contexts. *Journal of Anatomy, 208,* 153–167. doi:10.1111/j.1469–7580.2006.00523.x

Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition, 3,* 81–117. doi:10.1080/713756735

Dewar, K., & Xu, F. (2009). Do early nouns refer to kinds of distinct shapes? *Psychological Science, 20,* 252–257. doi:10.1111/j.1467–9280.2009.02278.x

Ekman, P. (1992). Are there basic emotions? *Psychological Review, 99,* 550–553. doi:10.1037/0033-295X.99.3.550

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44,* 227–240. doi:10.1016/0010–0277(92)90002-Y

Fugate, J. M. B., & Barrett, L. F. (in preparation). Seen and unseen words serve as a context for perceptual judgments of emotion.

Fulkerson, A. L., & Waxman, S. R. (2006). Words (but not tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition, 105,* 218–228. doi:10.1016/j.cognition.2006.09.005

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123,* 178–200. doi:10.1234/12345678

Harnad, S. (1987). *Categorical perception.* Cambridge, England: Cambridge University Press.

Izard, C. E. (1971). *The face of emotion.* New York: Appleton-Century-Crofts.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy. *Psychological Science, 18,* 1084–1089. doi:10.1111/j.1467–9280.2007.02029.x

Kikutani, M., Roberson, D., & Hanley, J. R. (2008). What's in a name? Categorical perception for unfamiliar faces can occur through labeling. *Psychonomic Bulletin & Review, 15,* 787–794. doi:10.3758/PBR.15.4.787

Levin, D. T., & Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Perception & Psychophysics, 62,* 386–401.

McKone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 573–599. doi:10.1037//0096-1523.27.3.573

Ozgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General, 131,* 477–493. doi:10.1037/0096–3445.131.4.477

Parr, L. A., Waller, B. M., Vick, S. J., & Bard, K. (2007). Classifying chimpanzee facial expressions using muscle action. *Emotion, 7,* 172–181. doi:10.1037/1528–3542.7.1.172

Pascalis, O., de Hann, M., & Nelson, C. A. (2002, May 17). Is face processing species-specific during the first year of life? *Science, 296,* 1321–1323. doi:10.1126/science.1070223

Pilling, M., Wiggett, A., Ozgen, E., & Davies, I. R. L. (2003). Is color "categorical perception" really perceptual? *Memory & Cognition, 3,* 538–551.

Plunkett, K., Hu, J.-F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition, 106,* 665–681. doi:10.1016/j.cognition.2007.04.003

Roberson, D., Damjanovic, L., & Pilling, M. (2007). Categorical perception of facial expressions: Evidence for a "category adjustment" model. *Memory & Cognition, 35,* 1814–1829.

Roberson, D., & Davidoff, J. (2000). The categorical perception of color and facial expressions: The effect of verbal interference. *Memory & Cognition, 28,* 977–986.

Roberson, D., Davidoff, J., & Braisby, N. (1999). Similarity and categorization: Neuropsychological evidence for a dissociation in explicit categorization tasks. *Cognition, 71,* 1–42. doi:10.1016/S0010-0277(99)00013-X

Russell, J. (1994). Is there universal recognition of emotion from facial expression—A review of the cross-cultural studies. *Psychological Bulletin, 115,* 102–141. doi:10.1037/0033–2909.115.1.102

Scott, L., & Monesson, A. (2009). The origin of biases in face perception. *Psychological Science, 20,* 676–680.

Stevenage, S. V. (1998). Which twin are you? A demonstration of induced categorical perception of identical twin faces. *British Journal of Psychology, 89,* 39–57.

Tomkins, S. S. (1962). *Affect, imagery, consciousness: Vol. 1. The positive affects.* New York: Springer.

Viviani, P., Binda, P., & Borsato, T. (2007). Categorical perception of newly learned faces. *Visual Cognition, 15,* 420–467. doi:10.1080/13506280600761134

Waller, B. M., Vick, S. J., Parr, L. A., Bard, K. A., Pasqualini, M. C., Gothard, K. M., & Fuglevand, A. J. (2006). Intramuscular electrical stimulation of facial muscles in humans and chimpanzees: Duchene revisited and extended. *Emotion, 6,* 367–382. doi:10.1037/1528–3542.6.3.367

Waxman, S. R., & Braun, I. (2005). Consistent (but not variable) names as invitations to form object categories: New evidence from 12-month-old infants. *Cognition, 95,* B59–B68. doi:10.1016/j.cognition.2004.09.003

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29,* 257–302. doi:10.1006/cogp.1995.1016

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences, USA, 104,* 7780–7785. doi:10.1073/pnas.0701644104

Xu, F., Cote, A., & Baker, A. (2005). Labeling guides object individuation in 12-month-old infants. *Psychological Science, 16,* 372–377.

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., & Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition, 63,* 271–313. doi:10.1016/S0010-0277(97)00003–6

Received April 13, 2009
Revision received October 16, 2009
Accepted December 15, 2009 ■