Categorical perception of facial expressions: Evidence for a “category adjustment” model

DEBI ROBERSON AND LJUBICA DAMJANOVIC
University of Essex, Colchester, England

AND

MICHAEL PILLING
Oxford Brookes University, Oxford, England

Four experiments probed the nature of categorical perception (CP) for facial expressions. A model based on naming alone failed to accurately predict performance on these tasks. The data are instead consistent with an extension of the *category adjustment model* (Huttenlocher et al., 2000), in which the generation of a verbal code (e.g., “happy”) activated knowledge of the expression category’s range and central tendency (prototype) in memory, which was retained as veridical perceptual memory faded. Further support for a memory bias toward the category center came from a consistently asymmetric pattern of within-category errors. Verbal interference in the retention interval selectively removed CP for facial expressions, under blocked, but not under randomized presentation conditions. However, verbal interference at encoding removed CP even under randomized conditions and these effects were shown to extend even to caricatured expressions, which lie outside the normal range of expression categories.

The extent to which the phenomenon of categorical perception represents underlying perceptual asymmetry, or merely the engagement of category labels has been the subject of much recent debate (Bornstein & Korda, 1984; Calder, Young, Perrett, Etcoff, & Rowland, 1996; Harnad, 1987; Munnich & Landau, 2003; Pilling, Wiggott, Özgen, & Davies, 2003; Roberson & Davidoff, 2000; Young et al., 1997). Categorical Perception (CP) is the name given to the increased sensitivity to a physical change found when that change crosses the boundary between two perceptual categories (Harnad, 1987). It has been demonstrated for simple perceptual continua such as phoneme blends (Pastore, 1987) or color categories (Roberson, Davies, & Davidoff, 2000; Pilling et al., 2003), and a number of recent studies have examined the phenomenon with regard to face processing. A steady change between two faces (differing in either identity or facial expression) can be created using morphing software (Beale & Keil, 1995; Calder et al., 1996). However, rather than being perceived as a monotonic linear progression, the continuum of identity or expression is perceived as an abrupt discontinuity at the boundary between two categories (e.g., happy–sad).

Demonstrations of CP for morphed facial expressions (Etcoff & Magee, 1992; Calder et al., 1996; Young et al., 1997) as well as for facial identity (Beale & Keil, 1995; Campanella, Hanoteau, Seron, Joassin, & Bruyer, 2003; Stevenage, 1998) have shown that it is relatively hard to

make discriminations near the center of a category and easy close to a boundary (Calder et al., 1996). Young et al. (1997) used an X–AB two-alternative forced-choice (2-AFC) memory paradigm in which a series of morphed facial expressions were created, with an equal amount of physical difference between each successive pair in the continuum (see Figure 1). Participants viewed one of the morphs as a target expression, followed by a test pair of images (target + distractor), differing only in morphed expression. Their task was to identify the position of the target face in the test pair by keypress. Participants showed superior identification of expressions from pairs that crossed a category boundary (e.g., happy–fearful) than from pairs where both members belonged to the same category (e.g., both happy). The finding is robust, but the locus of the effect remains unclear.

Some researchers have suggested that CP arises at an early perceptual level, or even that, for domains such as color, it might be innate and hardwired into the visual system (Bornstein, Kessen, & Weiskopf, 1976; Franklin & Davies, 2004). Young et al. (1997) argued for perceptually (rather than verbally) mediated CP for facial expressions, because labeling emotional facial expressions is actually more difficult at boundaries than at the center of categories, where a “prototypical” emotional face gives a more consistent and rapid response. Thus it seems unlikely that verbal labeling could underlie the superior discrimination of emotional expressions at category boundaries.

D. Roberson, robedd@essex.ac.uk

Figure 1. Morphed facial expressions 1–15 (with percentage naming agreement in brackets) moving in 20% steps from fear–happiness–anger–fear used in Experiments 1–3, taken from Young et al. (2002) *FEEST* test.

Others have proposed that CP is based, to a greater or lesser degree, on comparisons of verbal labels of the stimuli (Fujisaka & Kawashima, 1971; Kay & Kempton, 1984; Pilling et al., 2003; Roberson & Davidoff, 2000; Rosen & Howell, 1987). Roberson and Davidoff found that verbal, but not visual, interference in the interstimulus interval (ISI) between presentation of the target and test pair selectively removed CP for facial expressions, using an X–AB paradigm. They suggested that CP effects arose directly through verbal coding because the different labels assigned to stimuli that straddle category boundaries gave an additional tool for cross-category judgments that is not available for within-category judgments.

However, other evidence argues against a straightforward labeling account of CP. First, for complex stimuli, such as facial expressions, identification boundaries do not always align precisely with peaks in discrimination (Calder et al., 1996). Second, prelinguistic infants, at six months of age, appear to perceive happy expressions as

categorically distinct from other facial expressions (Kotsoni, de Haan, & Johnston, 2001). Roberson, Davidoff, and Braisby (1999) also found preserved CP for a morphed set of facial expressions in an amomic patient, who no longer had explicit access to emotion category labels. Moreover, for other sets of complex stimuli, differences in naming do not always lead to differences in categorization (Malt, Sloman, Gennari, Shi, & Wang, 1999). Furthermore, Goldstone, Lippa, and Shiffrin (2001) found evidence against labeling accounts of CP acquired through categorization training. Two faces in the same trained category were judged to be more similar to a third neutral stimulus after category training. Goldstone et al. (2001) argued that labeling could not account for this change because the neutral stimulus was never seen during training, and therefore could not have been labeled.

An alternative to a simple labeling explanation has been put forward by Huttenlocher, Hedges, and Vevea (2000), who suggested that naming a target at encoding activates a

representation of the (named) category prototype (defined as the central tendency of a known distributional range). As the veridical memory trace of the target fades over time, participants' estimations are biased by categorical information so that there is a consistent shift in recognition toward the category center. The greater the task difficulty (e.g., because of interference), the less exact the perceptual encoding of the target, and thus the stronger the influence of the activated prototypical representation. In this case CP would arise through labeling, but only indirectly, because for cross-category items the distractor would come from a different category to the activated prototype, whereas for within-category trials both target and distractor would come from the same category as the activated prototype. Huttenlocher et al. (2000) provided convincing evidence in support of the model for artificial, unidimensional stimuli, but it has not previously been tested with real-world stimuli.

Face processing is one domain in which it has been proposed that knowledge is stored in a framework around a prototypical representation, that is generated as the central tendency of all exemplars encountered (Valentine, 1991). Facial expressions may be stored in a similar "face-space" around prototypes. The prototypes might then be activated in response to a target stimulus and, when the target is a poor exemplar of the category (distant from the category center), the activated prototypical representation might influence memory for the target stimulus. In that case, where two test items came from the same category, but the distractor was closer to the category center than the target, subjects should often mistakenly select the more central distractor. They should seldom mistakenly choose a more distant distractor if the target was central. Activation of categorical knowledge should therefore result in an asymmetric pattern of errors for within-category targets. Further, if generation of a verbal label is necessary to activate the appropriate category representation, then verbal interference, by preventing label generation, should disrupt category activation, so that an asymmetric error pattern would not be observed under verbal interference conditions. Previous tests of CP for facial expressions have not considered the pattern of errors for within-category targets.

In the present set of experiments, we compared our results to a simple direct naming model but also sought evidence of the category adjustment model of Huttenlocher et al. (2000) by examining the pattern of within-category errors. If superior recognition of stimuli at category boundaries (the hallmark of CP) arises purely through the generation of verbal labels, then, without interference, a target item that is named differently from the distractor should always be successfully identified. When target and distractor share the same name, identification of the correct item should be at chance. This prediction would be modulated, however, by the degree of consistency with which labels are applied (Pilling et al., 2003). We therefore compare accuracy of identification to a model that takes into account the consistency with which individual targets are named.

A direct labeling account of CP not only predicts low within-category accuracy (because both target and distractor share the same name) but also that there should be no recognition advantage for either member of a within-category pair. In contrast a Category Adjustment model predicts an asymmetric pattern of within-category errors, biased toward the category center. We tested these predictions both with morphed expressions ranging between expressions (for which naming becomes more varied toward category boundaries, and with caricatured expressions, for which naming is invariant). We therefore examined the pattern of errors for within-category pairs of stimuli where one stimulus (e.g., 90% happy/10% fearful) was closer to the category center than the other (e.g., 70% happy/30% fearful). If labeling a target activates the central tendency of the category (Huttenlocher et al., 2000) and the memory representation is biased in the direction of the category center during the ISI, we would always expect more mistaken identifications of the stimulus closer to the category center than vice versa. If verbal interference prevents activation of categorical information, no such asymmetry of errors should occur in this condition.

Roberson and Davidoff (2000) proposed a direct labeling account because verbal, but not visual, interference in the ISI selectively removed CP. They suggested that participants always accessed labels for target expressions, even when no verbal response was required. Verbal interference in the ISI would not prevent verbal labels being generated (since the interference occurs after the target is encoded) but Pilling et al. (2003) suggested that, in blocked conditions, participants anticipate the interference and strategically choose to rely on either visual or verbal encoding. Under randomly interleaved interference conditions, such strategic switching should not be possible and discrimination of cross-category pairs should remain better than within-category pairs in all three conditions.

We therefore compared performance between blocked (Experiment 1) and randomized (Experiment 2) interference conditions, with secondary tasks that require the same modality of response and for which we could measure performance, to assess whether subjects strategically choose to rely on one or other code under blocked (but not random) interference conditions.

In Experiment 3, we investigated whether verbal interference at encoding, in the form of incongruent verbal labels can disrupt categorical perception, even under randomized conditions. If labeling facilitates CP indirectly, by activating a category prototype, then reading an incongruent emotional label should always impact CP negatively, because it would activate an inappropriate prototype.

Finally, in Experiment 4, we conducted a stronger comparison of a direct labeling account with a Category Adjustment model using caricatured expressions of emotion. Unlike for colors, it is possible to intensify facial expressions of emotion by computer enhancement (Calder et al., 2000; Calder, Young, Rowland, & Perrett, 1997). Thus it is possible to examine discrimination of facial expressions that lie outside any representational "face space" and thus

Figure 2. Format of 2-AFC trials with no interference, visual interference or verbal interference in Experiments 1, 2, and 4.

outside the range of previously encountered exemplars on which the central tendency of the category is formed. Since caricatures are consistently given the same label as the prototypical expressions, if CP were based on labeling alone, we would predict that caricatures and best examples should be equally well recognized and that no bias toward selecting a more central distractor would be observed when the target is a caricature that portrays a more intense emotion. An account based on activation of the prototype, would instead predict a consistent bias toward selecting a more central distractor, even when the target is a more intense caricature.

EXPERIMENT 1

Roberson and Davidoff (2000) reported that verbal, but not visual interference, selectively removed the discrimination advantage for stimuli that crossed a category boundary between facial expressions. Although both types of interference task were visually presented, in their experimental design the verbal interference condition required verbal output (either articulatory suppression or reading aloud a list of words) while the visual interference condition required only visual scanning. Overall, visual interference

affected accuracy less than verbal interference but, as there was no measure of accuracy on the two secondary tasks, greater difficulty of the verbal interference task cannot be ruled out. We first seek to replicate this result using the same facial expression stimuli, but interference conditions that are more closely balanced in task demand, because both verbal and visual tasks require a verbal output, for which performance can be systematically measured.¹

The experiment used a blocked X–AB design in which participants first viewed a target facial expression and were then required to select the target from two test images of the same face, differing only in expression. The target face was shown for 1 sec, followed by a 5-sec ISI before the appearance of the two test faces. The same stimulus trials were presented in three conditions: one with an unfilled ISI, one with a secondary verbal task in the ISI and one with a secondary visual task. Figure 2 shows the format of trials in each of the three conditions. Conditions were blocked and order of presentation of the three blocks (no interference, visual interference or verbal interference) was pseudorandomized so that each type of interference appeared equally often in each position. There was a short break between each block of trials.

Table 1
Composition of Within- and Cross-Category Pairs of Faces
Used in Experiments 1–3

Within-Category Pairs	
90% fear/10% happy (1) + 30% angry/70% fear (14)	
70% fear/30% happy (2) + 10% angry/90% fear (15)	
30% fear/70% happy (4) + 90% happy/10% angry (6)	
10% fear/90% happy (5) + 30% angry/70% happy (7)	
30% happy/70% angry (9) + 90% angry/10% fear (11)	
10% happy/90% angry (10) + 70% angry/30% fear (12)	
Cross-Category Pairs	
70% fear/30% happy (2) + 30% fear/70% happy (4)	
70% happy/30% angry (7) + 30% happy/70% angry (9)	
70% angry/30% fear (12) + 30% angry/70% fear (14)	

Note—Numbers in parentheses correspond to the numbering of the faces in Figure 1.

Method

Participants. Eighteen undergraduate students (7 males, 11 females, mean age = 20.7 years) from the University of Essex subject pool participated in the experiment for course credits. All had normal or corrected-to-normal vision.

Stimuli and Apparatus. All stimuli were displayed on a Sony Trinitron SVGA monitor controlled by a 100-MHz 80486 IBM PC compatible computer. The experiments were run from specially created programs, written in Microsoft Visual BASIC. The facial expressions used were identical to those used in Roberson and Davidoff (2000) Experiment 5 and were taken from the Facial Expressions of Emotion: Stimuli and Tests (FEEST—Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002). Black and white photographs of a Caucasian male face, showing typical expressions of happiness, fear or anger, were morphed in equal steps in a continuum that crossed the boundaries between three facial expressions: fearful–happy; happy–angry; angry–fearful (see Figure 1).

Expression naming data for each of the 15 morphed faces were taken from Calder et al. (1996) (for 24 participants). Additional expression naming data were collected for each face from 24 undergraduate students at the University of Essex (9 males, 15 females). Participants were required to decide, for each face, what emotional expression it conveyed. The percentage name agreement obtained correlated highly with that reported by Calder et al. (1996) ($r = .92$). Based on the combined expression naming data we were able to calculate the probability, for each stimulus, that successful identification could be achieved based on naming alone.

Pairs of faces, two morphed steps apart, were constructed so that the two members of the pair contained either predominantly the same category of facial expression (e.g., 90% happy/10% angry + 70% happy/30% angry) or predominantly different categories (e.g., 60% happy/40% angry + 40% happy/60% angry). Table 1 shows the composition of all pairs of test faces. There were 8 repetitions of each pair in each condition giving 72 trials per condition and 216 trials in all.

Stimuli for the secondary interference task were taken from Pilling et al. (2003). For the verbal-interference task, 72 word pairs (half rhyming, half not rhyming) were displayed sequentially on the computer monitor in Arial font on a neutral background. For the visual interference task, 72 shape pairs were used. All shapes were two-dimensional outline drawings of closed abstract shapes in white lines on a neutral gray background. The shapes in each pair were always physically different from one another. For half the shape pairs the first shape could fit into the second shape, while for the other half it could not.

Procedure. Participants carried out practice trials for each type of interference, before the start of the experimental trials, using stimuli that did not appear in subsequent experimental trials. Participants received instructions for the secondary task (shape fit judgments, rhyme judgments or no task) before the beginning of each block of trials. All trials commenced with a fixation cross, presented in

the center of the screen for 300 msec followed by a target stimulus, presented in the center of the screen for 1,000 msec. In the no interference condition, participants saw a blank neutral gray screen for the duration of the ISI (5,000 msec). In the verbal-interference condition, 300 msec after the offset of the target, participants saw one word (for 1,000 msec), in the center of the screen, followed (after 300 msec), by a second word (for 3,400 msec or until a response was made) and were required to respond, by keypress, whether the second word did, or did not, rhyme with the first. After a response was made the screen went blank until the end of the 5,000-msec ISI. Word pairs were randomly selected, and each pair appeared only once in each condition. In the visual-interference condition 300 msec after the offset of the target, participants saw a shape (for 1,000 msec), followed by a second shape (for 3,400 msec or until a response was made) and were required to respond, by keypress, whether the first shape would, or would not fit into the second. After a response was made the screen went blank until the end of the 5,000-msec ISI. Shape pairs appeared in the center of the screen. Shape pairs were randomly selected and each pair appeared only once in each condition. No feedback for performance on the secondary task was given. At the end of the ISI, the test pair of faces (target + distractor) appeared on the screen and participants were required to select, by keypress, whether the (identical) target face appeared on the left or the right of the pair. Test pairs of faces differed only in expression.

The position of the target face within the test pair (left or right) was pseudorandomized so that it appeared in each position on half the trials. The inter trial interval was 600 msec. Instructions emphasized accuracy for the primary task and both speed and accuracy for the secondary task.

Results and Discussion

Accuracy on the verbal and visual interference tasks was 86% and 82%, respectively, and there was no significant difference in accuracy on the two tasks by repeated measures t test [$t(17) = 1.49, p > .1$]. The two interference tasks thus appeared to be of equivalent difficulty.

For the primary 2-AFC task, with well-equated types of interference, in blocked conditions, visual interference significantly impaired performance for both cross- and within-category pairs of stimuli, but verbal interference had a greater impact on cross-category pairs and reduced discrimination for cross-category stimuli to the same level as within-category pairs. Figure 3 illustrates these results.

The proportion of mean correct identifications on the primary task was above chance in all conditions [all one-tailed $ts(17) > 2.5$, all $ps < .05$]. The data were analyzed in a 3 (interference type: none vs. visual vs. verbal) $\times$ 2 (pair type: cross- vs. within-category) within-subjects ANOVA. There was a significant effect of interference [$F(2,34) = 22.67, MS_e = .01, p < .001$], a significant effect of pair type [$F(1,17) = 31.78, MS_e = .01, p < .001$] as well as a significant interaction [$F(2,34) = 4.02, MS_e = .01, p < .05$]. Cross-category pairs were better recognized than within-category pairs overall, and both types of interference caused a significant decrement in performance, compared to an unfilled ISI. Examination of the simple main effects revealed that cross-category pairs were discriminated better than within-category pairs without interference [$F(1,17) = 29.17, p < .01$] and also with visual interference [$F(1,17) = 8.06, p < .05$], but not with verbal interference [$F(1,17) = 2.33, p > .1$].

Asymmetry of within-category errors. To explore the possibility that accessing the name of a stimulus acti-

Figure 3. Mean proportion correct identifications (with standard error bars) for cross- and within-category pairs of facial expressions in Experiment 1.

vated a representation of the category prototype, we examined the data for within-category judgments, considering whether the distractor stimulus was closer to the category center than the target (e.g., distractor = 90% fear, target = 70% fear), or closer to the category periphery than the target (e.g., distractor = 70% fear, target = 90% fear). Activation of a category prototype during the ISI was predicted to result in an asymmetric pattern of errors, so that more errors would be made to more central distractors. The proportion of total within-category errors made to distractors that were closer to the category center than the target was .62 without interference, .59 with visual interference and .55 with verbal interference. All of these were significantly above chance, by one-tailed *t* test [$t(17) = 5.61, p < .01$; $t(17) = 5.97, p < .01$; $t(17) = 2.55, p < .05$, respectively].

The proportion of errors to central distractors was compared by a one-factor, within-subjects ANOVA (proportion of central distractors: no interference vs. visual interference vs. verbal interference). There was a significant effect of condition [$F(2,34) = 4.75, p < .05$]. A Newman-Keuls comparison of the main effect revealed that a significantly greater proportion of mistaken identifications of central distractors were made with no interference than with verbal interference ($p < .05$). There were no other significant differences. Verbal, but not visual interference, significantly reduced the apparent shift in memory representation toward the category prototype over the ISI.

Test of direct naming. Following the formula used by Pilling et al. (2003), the probability of a correct identification from judgments based on naming alone was calculated for each stimulus and compared to the observed accuracy. The model makes the assumption that nothing except the target's verbal label (e.g., "happy") is remembered across the ISI, and that labels for the two test stimuli are subsequently compared to the target label. When such a comparison yields an unambiguous choice (because target and distractor have different labels) then the target is chosen. However, when the same label applies to both test stimuli, discrimination should

be at chance. The probability of a correct discrimination on an unambiguous trial is calculated as

$$[p(T - N1) \times p(T - N1) \times p(D - N2)] \\ + [p(T - N2) \times p(T - N2) \times p(D - N1)],$$

where N1 and N2 are the two possible names, T is the target, D is the distractor, and $p(T - N1)$ is the probability that the target is named N1 etc.). In this case, a correct discrimination is made either if all the stimuli are named correctly (first half of the equation) or if all the stimuli are named incorrectly (second half of equation). The probability of correctly identifying a target on an ambiguous trial is calculated as

$$\{[p(T - N1) \times p(T - N1) \times p(D - N1)] \\ + [p(T - N2) \times p(T - N2) \times p(D - N2)]\} / 2.$$

The overall probability of a correct discrimination is the sum of the results of the two equations.

Predicted scores for every stimulus combination were calculated. The predicted and observed accuracy data from the no-interference condition were as follows: Cross-category pairs – predicted = .803, observed = .812. Within-category pairs – predicted = .447, observed = .647. The difference between predicted and observed values did not differ significantly from 0 for cross-category stimuli [$t(5) = 1.38, p > .05$] so the model predicted accuracy for cross-category stimuli well, but the model significantly underpredicted accuracy for within-category stimuli [$t(11) = 6.41, p < .01$]. Thus participants must have used more than a verbal code alone to remember targets across the ISI.

In this experiment, performance on the two secondary tasks did not differ significantly, but slightly better performance was achieved on the verbal interference task. The findings of Roberson and Davidoff's (2000) Experiment 3 were replicated, even with secondary tasks equated for difficulty, so the previous findings were not an artifact of greater difficulty of the verbal interference task. When

conditions were blocked, verbal interference selectively removed the advantage for cross-category pairs of stimuli in recognition memory, suggesting that this advantage arose as a function of verbal coding of the stimuli. Without interference, within-category discrimination was significantly above chance [$t(17) = 7.30, p < .05$] and visual interference caused a significant decrement in performance, suggesting that visual coding supports the level of discrimination achieved for within-category pairs. Further evidence for the use of both a visual and verbal code comes from the failure of a model based on labeling alone to accurately predict performance for both types of pair. The model predicted cross-category naming well, but underpredicted the accuracy of within-category judgments. If targets were coded both visually and verbally, then verbal coding might account for the additional discriminability of cross-category pairs of stimuli, over and above within-category pairs.

However, there was also evidence that the memory representation shifted toward the category center during the ISI in all but the verbal interference condition, because distractors closer to the category center were misrecognized more often than distractors closer to the category boundary. This suggests that verbal coding of targets may have activated the relevant category prototype and thus indirectly contributed to CP, because both the label and the activated categorical representation supported identification of the target in cross-category pairs. For within-category pairs fewer errors would be expected to central than to peripheral targets. This possibility is further addressed in Experiment 3.

EXPERIMENT 2

Experiment 1 replicated the findings of Roberson and Davidoff's (2000) Experiment 3, with blocked conditions of visual and verbal interference tasks that both required a verbal output. However, under blocked verbal interference, participants may have strategically avoided generating verbal labels and tried to rely solely on visual coding.

With randomized interference conditions, participants are unaware, at encoding, what type of interference will occur and therefore cannot adopt an efficient coding strategy. If participants continue to code targets both visually and verbally, then CP should survive verbal interference. We therefore wished to inquire whether CP for facial expressions would survive verbal interference under randomized interference conditions,² using the same types of interference.

Method

Participants. Eighteen undergraduate students (3 males, 15 females, mean age = 18.8 years) from the University of Essex participated for course credits. All had normal or corrected-to-normal vision. None had participated in any previous experiment.

Stimuli and Apparatus. The stimuli and apparatus for both the primary and secondary tasks were identical to those used in Experiment 1.

Procedure. Stimulus pairs were randomly presented within conditions, but all pairs appeared with equal frequency in all conditions. Each stimulus was shown as a target and distractor with equal frequency. On each trial, following the 300-msec fixation cross, the target was shown for 1,000 msec, followed by a randomly selected filled or unfilled 5,000-msec ISI and then two test stimuli (target plus distractor) for 1,000 msec, or until a response was made. Target position within the test pair (left or right) was pseudorandomized so that it appeared in each position on half the trials. The intertrial interval was 600 msec. Instructions emphasized accuracy for the primary task and both speed and accuracy for the secondary task. There were two blocks of 108 trials, with a short break between blocks. Interference type was randomized within each block.

Participants carried out practice trials for each type of interference, before the start of the experimental trials, using stimuli that did not appear in subsequent experimental trials and received instructions for the secondary task (shape fit judgments, rhyme judgments or no task) before the beginning of each block of trials.

Results and Discussion

Accuracy on the verbal and visual interference tasks was 81% and 73%, respectively and there was no significant difference in accuracy on the two tasks [$t(17) = 1.59, p > .1$]. As in Experiment 1, there was no apparent difference in the difficulty of the two secondary tasks.

Figure 4. Mean proportion correct identifications (with standard error bars) for cross- and within-category pairs of faces in Experiment 2.

For the primary 2-AFC task, cross-category pairs of stimuli were better discriminated than within-category pairs in all three conditions. Figure 4 illustrates these results.

The proportion of mean correct identifications on the primary task was above chance in all cross-category conditions, and in the within-category condition without interference [all one-tailed $t(17) > 2.5$, all $ps < .05$]. With both visual and verbal interference, within-category identification was at chance. Randomly switching interference types should have increased the task demands in Experiment 2, compared to Experiment 1, causing less exact perceptual encoding of targets and greater reliance on verbal labels. This impacts particularly on within-category pairs, where both target and distractor have the same verbal label, so that only perceptual information can differentiate target and distractor.

The data were analyzed in a 3 (interference type: none vs. visual vs. verbal) $\times$ 2 (pair type: cross- vs. within-category) within-subjects ANOVA. There was a significant effect of interference [$F(2,34) = 38.50$, $MS_e = .01$, $p < .001$] and a significant effect of pair type [$F(1,17) = 40.08$, $MS_e = .01$, $p < .001$] but no significant interaction [$F(2,34) = 1.32$, $MS_e = .01$, $p > .1$]. A Newman-Keuls post hoc analysis of the main effect of interference type revealed that significantly more targets were discriminated in the No Interference condition than in either the visual or verbal interference conditions (both $ps < .01$) and significantly more targets were discriminated in the visual than in the verbal interference condition ($p < .01$). Both verbal and visual interference impaired performance, although verbal interference produced the larger reduction in performance overall.

Asymmetry of within-category errors. Asymmetry of within-category errors was again examined by considering whether the distractor was closer to the category center than the target or closer to the category periphery. Significantly more errors were made to distractors closer to the category center without interference (.55) [$t(17) = 2.6$, $p < .05$] and with visual interference (.53) [$t(17) = 3.03$, $p < .05$], but not with verbal interference (.52) [$t(17) = .452$, $p > .05$]. However, a one-factor (interference type: none vs. visual vs. verbal) within-subjects ANOVA found no significant main effect of interference [$F(2,34) < 1$]. The proportion of mistaken identifications of central distractors did not differ significantly across conditions.

Test of direct naming. The predicted and observed accuracy data from the no-interference condition were as follows: Cross-category pairs – predicted = .684, observed = .615. Within-category pairs – predicted = .561, observed = .646. The difference between predicted and observed values did not differ significantly from 0 for cross-category stimuli [$t(5) = 1.97$, $p > .05$] so the model predicted accuracy for cross-category stimuli well, but the model significantly underpredicted accuracy for within-category stimuli [$t(11) = 4.32$, $p < .01$].

In Experiment 2, the effects of CP survived verbal interference under random presentation conditions, although the advantage for cross- over within-category items was

reduced under verbal interference. Both types of interference also impaired accuracy for within-category discrimination, suggesting that both visual and verbal codes are used when interference is unpredictable. An asymmetric pattern of errors was observed both without interference and under visual interference, suggesting that labeling activated the category prototype. This tendency was reduced under verbal interference, so the secondary verbal task may have suppressed the activation of the prototype during the ISI. If the verbal label were recalled, but verbal interference blocked activation of the prototype representation in memory, then CP should still be found, but to a lesser extent than in the blocked no-interference condition of Experiment 1 (where both verbal label retention and prototype activation would facilitate CP). Without interference in Experiment 1 the mean difference in accuracy between cross- and within-category pairs was .153; under verbal interference in this experiment it was .125. Although the difference between these two means was not significant [$t(17) = 1.027$, $p > .05$] the trend was in the predicted direction. Experiment 3 directly examines the possibility that presenting an inappropriate label alongside the target might activate a different category prototype.

EXPERIMENT 3

If verbal labeling of targets facilitates CP, both directly, through name retention, and indirectly, because it activates a category prototype; then reading an incongruent label should impact heavily on CP in this task, because it could activate an inappropriate prototype. However, if all stimuli are routinely labeled anyway, then reading a congruent label would not be expected to further increase CP. Experiment 3 therefore investigated the same primary task as Experiments 1 and 2, but using a paradigm where congruent or incongruent emotional labels appeared at the same time as the target stimulus. Participants were instructed to ignore the words and attend only to the target face.

There were two reasons to expect a stronger effect from the to-be-ignored words in this paradigm than from the interference in Experiments 1 and 2. First, the incongruent word was the label of a different emotion from that displayed by the target (whereas the verbal interference in Experiments 1 and 2 contained no emotional content) and second, in this experiment, the words were presented at the same time as the target, whereas in Experiments 1 and 2 the interference occurred in the ISI, after the target stimulus had been encoded. Greater disruption of CP was therefore predicted for this paradigm.

Method

Participants. Twenty-four volunteers (5 males and 19 females) mean age = 18.4 years) from the University of Essex subject pool participated in the experiment. All had normal or corrected-to-normal vision. None had participated in any previous experiment.

Stimuli and Apparatus. Apparatus and stimulus pairs for the primary task were identical to those used in Experiments 1 and 2. In addition, at the time the target stimulus was presented, a word (HAPPY, ANGRY, OF AFRAID in experimental conditions; XXXXX in con-

Figure 5. Proportion of correct identifications (with standard error bars) in Experiment 3 with neutral, congruent, or incongruent verbal interference at target presentation.

trol condition) was simultaneously presented immediately below the target stimulus, printed in 16-point bold Times font.

Procedure. Participants were instructed that their task was to select, by keypress, whether the target face appeared on the left or the right of the test pair. In addition, they were instructed to ignore the words appearing on the screen and attend only to the faces. Participants carried out practice trials, before the start of the experimental trials, using stimuli that did not appear in subsequent experimental trials. On each trial, following the fixation cross, the target face and the word stimulus were shown together for 1,000 msec, followed by a 5,000-msec blank ISI and then two test stimuli (target plus distractor) for 1,000 msec, or until a response was made. The word appeared only during presentation of the initial target, not during the ISI or during presentation of the test pair. Target position within the test pair (left or right) was pseudorandomized so that it appeared in each position on half the trials. The intertrial interval was 600 msec. Instructions emphasized accuracy. Each target face appeared with the neutral letter string (XXXXX), the congruent expression word (e.g., HAPPY) and both incongruent words (e.g., ANGRY, AFRAID). There were two blocks of 72 trials, with a short break between blocks. Interference type was randomized across blocks. Participants carried out practice trials for each type of interference, before the start of the experimental trials, using stimuli that did not appear in subsequent experimental trials.

Results and Discussion

Figure 5 shows mean correct identifications for each of the three to-be-ignored word conditions. Mean correct identifications were significantly above chance in all conditions [all one-tailed $t(17) > 3$, all $ps < .05$] and were analyzed in a 3 (word type: neutral vs. congruent vs. incongruent) $\times$ 2 (pair type: cross- vs. within-category) within-subjects ANOVA. There was a significant effect of word type [$F(2,46) = 6.52$, $MS_e = .006$, $p < .01$], a significant effect of pair type [$F(1,23) = 114.19$, $MS_e = .06$, $p < .001$] and a significant interaction [$F(2,46) = 10.97$, $p < .001$]. Inspection of the simple main effects revealed that cross-category pairs were discriminated significantly better than within-category pairs with congruent words [$F(1,16) = 10.621$, $p < .01$] and the effect approached significance with neutral letter-strings [$F(1,16) = 3.36$, $p < .07$]; however, with incongruent words there was no

difference in the discrimination rates for cross- versus within-category pairs [$F(1,16) < 1$]. Newman–Keuls post hoc analysis of the main effect of word type revealed that a significantly greater proportion of stimuli were correctly identified in both congruent and neutral word conditions than in the incongruent condition (both $ps < .01$).

Consistent with the hypothesis that the superior discrimination of cross-category pairs arises because the verbal label activates a prototypical representation of the category, incongruent labels significantly reduced the discrimination of cross-category pairs, but had no effect on the discrimination of within-category pairs, presumably because, if the wrong prototype were activated, both members of the within-category pair would fail to match it to the same extent. Congruent labels did not give any extra facilitation of discrimination over neutral rows of XXXXX, presumably because, if participants were already routinely generating a verbal label for all stimuli, the appropriate prototype would be activated and provision of the written label added nothing more.

Asymmetry of within-category errors. Asymmetry of within-category errors was again examined by considering whether the distractor was closer to the category center than the target or closer to the category periphery. More errors were made to central distractors with neutral letter strings (.57) [$t(23) = 2.51$, $p < .05$] and with congruent words (.57) [$t(23) = 2.83$, $p < .05$], but not with incongruent words (.50) [$t(23) = .66$, $p > .1$]. A one-factor (word type: neutral vs. congruent vs. incongruent) within-subjects ANOVA found a significant main effect of word type [$F(2,46) = 2.54$, $p < .05$]. A Newman–Keuls comparison of the main effect of word type revealed that, while there was no significant difference between congruent and neutral word conditions, significantly more central errors were made with both congruent and neutral words than with incongruent words (both $ps < .05$).

As in Experiments 1 and 2, generation of a stimulus label appeared to bias memory toward the category center. When an incongruent word was presented, this bias was removed, as it was with blocked verbal interference in

Experiment 1. Failure to generate the appropriate category label, either strategically or because an inappropriate label is presented both appeared to block the activation of category relevant information.

Test of direct naming. The predicted and observed accuracy data from the neutral condition were as follows: Cross-category pairs – predicted = .684, observed = .667. Within-category pairs – predicted = .561, observed = .642. The difference between predicted and observed values did not differ significantly from 0 for cross-category stimuli [$t(5) = 1.74, p > .05$] so the model predicted accuracy for cross-category stimuli well, but the model significantly underpredicted accuracy for within-category stimuli [$t(11) = 2.51, p < .05$].

Again, although labeling alone predicted cross-category identification well in the neutral letter string condition, it underpredicted within-category identification, further supporting the findings of previous experiments.

If labeling a stimulus activates a representation of the central tendency of the category and thus biases estimates toward this (Huttenlocher et al., 2000), then such bias should be observed regardless of the stimulus set used. If, however, the bias arises simply because the stimulus is labeled (e.g., “happy”) and stimuli close to the category center are labeled more consistently than those at the category periphery, then any stimuli that are labeled equally consistently, should be equally well recognized. In Experiment 4, we tested categorical perception of facial expressions with a set of stimuli that included caricatures of the three expressions used in previous experiments (happiness, anger, fear). Since these stimuli are labeled just as consistently as the central examples used in Experiments 1–3, an account based on labeling would not predict more within-category errors from caricatured targets to distractors closer to the category center than vice versa. If labeling activates a category prototype, however, then one would expect a bias in estimation to more central distractors for both peripheral and caricatured targets.

EXPERIMENT 4

Whilst the asymmetry of within-category errors observed in Experiments 1–3 are consistent with an extended category adjustment model, in which generation of a verbal label is required to activate categorical information, it could also be consistent with an account based on labeling alone, even though a simple labeling model consistently underpredicts recognition accuracy for within-category pairs of expressions. It is possible to provide another test of this hypothesis with facial expression stimuli since, unlike for colors, it is possible to create caricatures of facial stimuli that exaggerate the expression of emotion portrayed.

Calder et al. (2000) created a set of caricatures of facial expressions of emotion, and showed that intensity ratings for the caricatures increased monotonically with the degree of exaggeration (+15%, +30%, +50%). In a pilot study, 24 students (6 males, 18 females, aged between 19 and 23 years) from the University of Essex named the emotions displayed in a set of morphed faces containing

100%, 110%, 120%, 130%, and 150% of happiness, fear, and anger, respectively (taken from the FEEST test; Young et al., 2002). These faces are illustrated in Figure 6. All five “happy” expressions received 100% “happy” labels; the five “fearful” expressions received 100% “fearful” labels, and the five “angry” expressions received 100% “angry” labels. Thus, under a labeling account, no asymmetry of within-category errors would be predicted for these stimuli. Under a prototype-activation account, asymmetric errors would still be predicted where the target is a caricature and the distractor lies toward the category center. Half the participants in Experiment 4 carried out the experiment with blocked interference conditions and half with randomly interleaved interference conditions.

Method

Participants. Thirty-four undergraduate students (13 males, 21 females, mean age = 20.7 years) from the University of Essex subject pool participated in the experiment for course credits. All had normal or corrected-to-normal vision. Seventeen participants carried out the experiment with blocked conditions and 17 with randomly presented conditions.

Stimuli and Apparatus. All stimuli were displayed on a Sony Trinitron SVGA monitor controlled by a 100MHz 80486 IBM PC compatible computer. The experiments were run from specially created programs, written in Microsoft Visual BASIC. Nine pairs of faces, two morphed steps apart, were used: three within- and three cross-category pairs of faces from Experiments 1, 2, and 3; together with 3 pairs of within-category faces that included a caricature (90% + 110%; 100% + 120%; 130% + 150%). These faces are illustrated in figure 6. Each pair of faces was repeated 8 times (4 with each face as target) and each target appeared equally often on the left and right sides of test pairs, giving 72 trials/condition and 216 trials in all.

The visual and verbal interference tasks were identical to those used in Experiments 1 and 2.

Procedure. The instructions and procedure for the blocked conditions were identical to that in Experiment 1. Those for randomly ordered conditions were identical to that in Experiment 2. All participants carried out practice trials for each type of interference, before the start of the experimental trials, using stimuli that did not appear in subsequent experimental trials, including three pairs of caricatured expressions (1 happy, 1 fearful, 1 angry).

Results and Discussion

Blocked interference conditions. Data from 1 participant, whose performance was at chance in 4/6 conditions was discarded. For the remaining 16 subjects, accuracy on the verbal and visual interference tasks was 78.4% and 74.2%, respectively and there was no significant difference in accuracy on the two tasks [$t(16) = 1.71, p > .1$].

Figure 7A illustrates the results of the primary, 2-AFC task for the blocked conditions. Figure 7B illustrates the results of the primary 2-AFC task for randomized conditions.

The proportion of mean correct identifications on the primary task was significantly above chance for all cross- and within-category pairs and for caricatured pairs without interference and with verbal interference [all one-tailed $t(16) > 2.01$, all p s $< .05$] but did not differ from chance for caricatured pairs with visual interference. Overall lower performance on caricatured stimuli may reflect increased task difficulty, since these stimuli lie outside the previously encountered distribution of category exemplars. This may

Figure 6. Morphed facial expression caricatures (100%, 110%, 120%, 130%, and 150%) of fear–happiness–anger used in Experiment 4.

have lead to poorer perceptual encoding of targets, making them more susceptible to visual interference.

The data were analyzed in a 3 (interference type: none vs. visual vs. verbal) $\times$ 3 (pair type: cross- vs. within-category vs. caricature) within-subjects ANOVA. There was no significant effect of interference [$F(2,30) = 2.01$, $MS_e = .01$, $p > .05$] but a significant effect of pair type [$F(2,30) = 14.02$, $MS_e = .01$, $p < .001$] and a significant interaction [$F(4,60) = 3.08$, $MS_e = .01$, $p < .05$]. A Newman–Keuls post hoc analysis of the interaction revealed that, without interference, significantly fewer caricatured targets were discriminated than either within-category targets ($p < .05$) or cross-category targets ($p < .01$) and significantly fewer within-category than cross-category targets were discriminated ($p < .05$). In the visual interference condition, significantly more cross-category targets were recognized than within-category targets or caricatures ($p < .05$ and $p < .01$, respectively). There were no significant differences between the three types of pairs in the verbal interference condition. For cross-category pairs of facial expressions, significantly more targets were discriminated without interference or with visual interference than with verbal interference (both $ps < .05$) but there were no sig-

nificant differences between interference conditions for within-category pairs or caricatured pairs (all $ps > .05$). Without interference, cross-category pairs were better recognized than any within-category pairs (caricatured or not), replicating and extending the result of Experiment 1. Verbal, but not visual interference reduced performance on cross-category pairs to a similar level to the within-category pairs, again replicating and extending the result of Experiment 1. For pairs of faces including caricatures, performance was particularly poor under visual interference. Possible reasons for this are addressed in the general discussion.

Asymmetry of within-category errors. The proportion of errors made to distractors closer to the category center was .54 without interference (.54 for caricature pairs alone), .57 with visual interference (.57 for caricature pairs alone) and .49 with verbal interference (.48 for caricature pairs alone). Without interference or with visual interference it was significantly above chance, by one-tailed t test [$t(15) = 1.85$, $p < .05$; $t(15) = 2.54$, $p < .05$] but with verbal interference it was not [$t(15) = 1.04$, $p > .05$].

However, a one-factor (interference type: none vs. visual vs. verbal) within-subjects ANOVA found no significant

Figure 7. (A) Proportion of correct identifications (with standard error bars), using blocked conditions, for cross-category, within-category, and caricatured pairs of faces in Experiment 4. (B) Proportion of correct identifications (with standard error bars), using randomly interleaved conditions, for cross-category, within-category, and caricatured pairs of faces in Experiment 4.

main effect of interference [$F(2,30) = 1.98$, $MS_e = .01$, $p > .05$]. Under blocked conditions there was a tendency to select central distractors in all conditions and this effect was equally strong for pairs including a caricatured target as it was for pairs including a peripheral target. Verbal interference only slightly reduced the apparent shift in memory representation toward the category prototype over the ISI.

Test of direct naming. The predicted and observed accuracy data from the no-interference condition were as follows: Cross-category pairs – predicted = .808, observed = .773. Within-category pairs – predicted = .435, observed = .663. Caricatured pairs – predicted = .467, observed = .515. The difference between predicted and observed values did not differ significantly from 0 for cross-category stimuli [$t(11) = .651$, $p > .05$] or for caricatured pairs [$t(11) = .256$, $p > .05$] so the model predicted accuracy for both cross-category and caricatured pairs well, but significantly underpredicted accuracy for within-category stimuli [$t(11) = 8.63$, $p < .01$].

Discrimination of targets from caricatured pairs of faces is well predicted by naming alone, which would predict chance performance when both stimuli are always given the same name. However, discrimination of within-category pairs was significantly above chance, and thus better than predicted, dropping to chance only for the visual interference condition, suggesting that participants may rely more heavily on visual coding when discriminating caricatures.

The error bias toward central distractors, found in Experiments 1 and 2 was also found with caricatured pairs of test stimuli, where one of the test pair lies “beyond” the center of the established category distribution. This pattern would not be predicted by a simple naming model but was still observed, even with a relatively high proportion of errors for caricatured pairs of stimuli. However, the reduction in such bias under verbal interference was less than that observed in Experiment 1.

Randomized interference conditions. Accuracy on the verbal and visual interference tasks was 84% and 81%,

respectively, and there was no significant difference in accuracy on the two tasks [$t(17) = 1.47, p > .1$].

The proportion of mean correct identifications on the primary task was significantly above chance for all pair types [all one-tailed $t(16) > 2.5$, all $ps < .05$]. The proportion of mean correct identifications on the primary task was analyzed in a 3 (interference type: none vs. visual vs. verbal) $\times$ 3 (pair type: cross- vs. within-category vs. caricature) within-subjects ANOVA. There was a significant effect of interference [$F(2,32) = 5.10, MS_e = .01, p > .05$] and a significant effect of pair type [$F(2,32) = 37.36, MS_e = .02, p < .001$] as well as a significant interaction [$F(4,64) = 2.54, MS_e = .02, p < .05$]. A Newman-Keuls post hoc analysis of the interaction revealed that, without interference and with verbal interference, a significantly greater proportion of cross-category targets were discriminated than either within-category or caricatured targets (all $ps < .01$). With visual interference, significantly more cross-category targets were recognized than within-category targets or caricatures (both $ps < .01$), and more within-category targets were discriminated than caricatures ($p < .01$). For cross-category pairs of facial expressions, significantly more targets were discriminated without interference than with verbal interference ($p < .05$), but there were no other significant differences. There were no significant differences between interference types for within-category pairs of stimuli, but for caricatured pairs, discrimination in the visual interference condition was significantly worse than with either verbal or without interference (both $ps < .01$). Cross-category pairs were always better recognized than any within-category pairs (caricatured or not) regardless of interference type. This pattern replicates and extends that found in Experiment 2. Verbal interference significantly reduced discrimination of cross-category pairs, again replicating and extending the findings of Experiment 2, but for caricatured faces, visual interference significantly reduced accuracy. Consistent with the findings of the blocked condition, this suggests that participants rely more on perceptual coding of the caricatured stimuli. This finding is further addressed in the general discussion.

Asymmetry of within-category errors. The proportion of errors made to distractors closer to the category center was .59 without interference (.58 for caricature pairs alone), .56 with visual interference (.58 for caricature pairs alone), and .52 with verbal interference (.56 for caricature pairs alone). Without interference and with visual interference this proportion was significantly above chance [$t(17) = 3.70, p < .05$; $t(17) = 2.69, p < .05$, respectively] but with verbal interference it was not [$t(17) = .84, p > .05$].

A one-factor (interference type: none vs. visual vs. verbal) within-subjects ANOVA found no significant main effect of interference [$F(2,32) = 1.60, MS_e = .01, p > .05$]. As was the case with blocked conditions, under randomly interleaved conditions verbal interference did not significantly reduce the apparent shift in memory representation toward the category prototype over the ISI. However, without interference, or with visual interference the bias to select central distractors was equally strong

for pairs including a caricatured target as it was for pairs including a peripheral target.

Test of direct naming. The predicted and observed accuracy from the no-interference condition were as follows: Cross-category pairs – predicted = .808, observed = .789. Within-category pairs – predicted = .435, observed = .618. Caricatured pairs – predicted = .467, observed = .568. The difference between predicted and observed values did not differ significantly from 0 for cross-category stimuli [$t(11) = .386, p > .05$], so the model predicted accuracy for cross-category pairs well, but the model significantly underpredicted accuracy for within-category stimuli [$t(11) = 4.18, p < .01$] and caricatured pairs [$t(11) = 3.84, p < .01$].

When interference conditions were randomly interleaved, without interference, discrimination of caricatured pairs was above chance, which would not be predicted by a simple naming model. Both this comparison and the analysis of error bias above suggest that memory for targets is biased toward the central tendency of the established category.

The category adjustment model suggests that the degree of shift toward the category center in memory representation is a function of the quality of perceptual encoding of the target stimulus. To maximize accuracy “people use prior information in estimating stimulus values that are represented inexactly” (Huttenlocher et al., 2000, p. 220). The model predicts that with higher task demands, a less exact representation of the target stimulus is maintained, and thus the influence of categorical knowledge on memory is stronger. In the case of Experiment 4, both for blocked and randomized interference conditions, the addition of pairs of test stimuli that included a caricatured expression should have increased task demands, since these stimuli lie outside the previously encountered distribution of exemplars in an established categorical “face space.” Less exact perceptual encoding in this more difficult task might lead to a more persistent shift toward the established category center over the ISI. The categorical adjustment model fits the data from both blocked and randomized presentation conditions, for all types of stimuli, better than a model based on naming alone.

GENERAL DISCUSSION

The results of all four experiments support an extended category adjustment model, in which CP for facial expressions arises because individuals label expressions as “happy” or “sad” and, in so doing, access their knowledge of the known distribution and central tendency of categories of facial expressions. Over the course of retention, memory for targets is systematically biased toward the category center and this tendency increases when task conditions reduce the exactitude of the perceptual encoding of stimuli. Experiment 1 replicated the findings of Roberson and Davidoff (2000), using blocked conditions, but novel interference tasks, and ruled out the possibility that the selective effects of verbal interference on CP were due to differential task demands of the interference conditions. There was no significant difference in performance

levels of these new secondary tasks, indeed performance on the verbal (rhyme judgment) task was slightly more accurate than on the visual (shape fitting) task, so the selective effect of verbal interference was not due to the greater difficulty of this task.

When conditions were randomized, in Experiment 2, CP survived the effect of verbal interference. This suggests that, under blocked conditions, participants strategically avoided naming targets, as suggested by Pilling et al. (2003) and tried instead to rely wholly on perceptual encoding of the stimuli. When conditions were randomized, participants continued to use both verbal and visual codes, and thus cross-category pairs of stimuli were always discriminated better than within-category pairs. Moreover, superior discrimination of cross-category pairs cannot be due to purely image-based perceptual factors (such as the amount of perceptual distance between members of a pair) since verbal interference (under blocked conditions) reduces discrimination of cross-category pairs of stimuli to the same level as within-category pairs. Purely image-based perceptual processing of the stimuli, without accessing labels for the emotions expressed, would not account for the selective effect of verbal interference.

Roberson and Davidoff (2000) and Pilling et al. (2003) concluded that CP arose directly from retention of a verbal code for the target. However, the results of the present study are not consistent with an explanation based solely on labeling, which would predict that within-category discrimination should be at chance. Instead the data are consistent with an extended category adjustment model in which verbal coding of targets activates knowledge of the range and central tendency of category exemplars. A bias toward the central tendency and away from the veridical perceptual representation of the target during retention accounts for the asymmetric pattern of within-category errors found in all experiments, where central targets were correctly identified more often than distractors, but targets that were further from category centers (either toward the boundary or toward an extreme caricature) were less often correctly recognized. The results of all four experiments suggest that naming a target facial expression activated information about the central tendency of the category, which was then retained over the ISI.

In Experiment 1 and Experiment 4A, where participants may have strategically avoided naming targets under verbal interference, CP disappeared under verbal interference, and the asymmetric pattern of within-category errors observed in the other conditions was also reduced. In the other two conditions, where participants are presumed to have named targets, there were more mistaken identifications of central than of peripheral distractors. Similar effects were found in the randomly interleaved interference trials in Experiment 2 and Experiment 4B, although CP survived verbal interference. This suggests that categorical information was activated specifically when a target's label was accessed and that verbal interference, even after encoding, may still have disrupted such activation to some extent.

The asymmetry of errors in Experiments 1 and 2 might be explained by verbal coding alone, because cen-

tral distractors would also be labeled more consistently than peripheral targets. However, in Experiment 4, which used caricatured expressions, the same asymmetry was observed for pairs where the (caricature) distractor was labeled with the same consistency as the target. A straightforward labeling account cannot explain this pattern, but a model based on category adjustment would predict this pattern of performance, because the caricatures lie outside the normal range of exemplars in the category and therefore a memory bias toward the central tendency of the category would still be expected. The results of Experiment 4 not only replicated the main findings of Experiments 1 and 2, but also supported an extension of the category adjustment model (Huttenlocher et al., 2000). The model proposes that people use stored category information to adjust inexact representations of stimuli, to minimize errors of judgment. The model predicts that increased task demands should result in less exact encoding of targets, and therefore greater memory bias toward the central (prototypical) representation.

In Experiment 3, task demands were increased in two ways. First, the verbal interference was presented at the same time as the target whereas, in Experiments 1 and 2, interference was presented after the target had disappeared, when encoding should have already occurred. Second, the incongruent verbal label referred to another emotion (which should activate inappropriate categorical information). Under these conditions CP was eliminated, even in randomly interleaved interference trials. The activation of a conflicting category by incongruous emotion labels in Experiment 3 reduced discrimination of cross-category pairs to the same level as within-category discrimination, even in randomly interleaved conditions.

In contrast, under blocked verbal interference conditions (in Experiments 1 and 4A) when participants may have avoided labeling and strategically relied solely on visual encoding of the stimulus, the perceptual representation should have been much more exact, so less categorical bias would be predicted.

Another reason why our results are consistent with an extended category adjustment model, is that a model based on naming alone failed to predict discrimination, particularly for within-category pairs of facial expressions, for which naming alone consistently underpredicted discrimination. For the caricatured pairs of expressions used in Experiment 4 the model also underpredicted discrimination in randomized, but not blocked interference conditions. For these pairs of stimuli, a simple naming account would predict chance discrimination, but a category activation model would predict more accurate discrimination of targets central to the previously experienced range of exemplars, rather than of caricatured targets. This effect was found, with both blocked and randomized conditions, both without any interference and with visual interference, but not with verbal interference. If category-specific information is activated by labeling, then disruption of labeling by verbal interference would reduce the bias toward the category center.

Any prototypical representation must be general enough to capture what is common to the expression across differ-

ent individuals, of different gender, age and race and it has been suggested that caricatures, which contain exaggerated categorical information might be the most "prototypical" expressions of all (Calder et al., 2000), and thus that any categorical memory bias would favor these stimuli. Our data rather suggest that participants form categories on the basis of their experienced range of encountered exemplars, and the prototypical representations of facial expressions lie at the center of the normally experienced range, since errors from caricatured pairs are also biased toward the category center.

The present findings contrast with those of Pilling et al. (2003) for colors, where no evidence of a categorical bias in memory was found, and they challenge the proposal that CP is directly due to retention of a verbal label. One reason why the results of the present experiments may differ from those of Pilling et al. for colors is that the range of face stimuli used in these experiments includes examples that are very close to the category center (90% of the prototypical expression), whereas those used for color CP were restricted to an area around the category boundary, far from the best examples of the category. Such a narrow range may have caused temporary prototypical representations to be created based on the average of the set actually seen (Spencer & Hund, 2002). This may be more likely to occur with a narrow range of unidimensional stimuli, such as colors varying only in hue (but of constant brightness and saturation), than with a large range of multidimensional stimuli such as faces, for which it is more difficult to establish temporary representations (Campanella et al., 2003). Other authors have also argued that, at least for complex stimuli, CP does not arise directly from verbal labeling (Calder et al., 1996; Goldstone et al., 2001; Malt et al., 1999).

Our data suggest that the effect of labeling on memory is to activate categorical information that is retained alongside the visual representation of the target and systematically biases perceptual representations over the retention interval. This leads both to better discrimination of targets from cross-category pairs (because what is retained is effectively further from the distractor than the original target) and to asymmetric errors for within-category targets, such that more typical faces are more often selected in error than vice versa. Evidence for such category adjustment has previously been demonstrated for simple patterns (Huttenlocher et al., 2000) but the current data represent the first demonstration of the phenomenon with facial expressions. The present experiments extend the category adjustment model to define the role of category labels in activating information about the range and central tendency of categories and biasing memory for facial expressions in the direction of the category center.

AUTHOR NOTE

This research was partly supported by a University of Essex RPF grant to the first author. We are grateful to Komal Agarwal and Kodjo Fossu-Mensah for assistance with data collection and to three anonymous reviewers for their comments on an earlier draft of the manuscript. Correspondence concerning this article should be addressed to D. Roberson, Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, England (e-mail: robedd@essex.ac.uk).

REFERENCES

- BEALE, J. M., & KEIL, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*, 217-239.
- BORNSTEIN, M. H., KESSEN, W., & WEISKOPF, S. (1976). Color vision and hue categorization in young human infants. *Journal of Experimental Psychology: Human Perception & Performance*, *2*, 115-129.
- BORNSTEIN, M. H., & KORDA, N. O. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychological Research*, *46*, 207-222.
- CALDER, A. J., ROWLAND, D., YOUNG, A. W., NIMMO-SMITH, I., KEANE, J., & PERRETT, D. I. (2000). Caricaturing facial expressions. *Cognition*, *76*, 105-146.
- 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.
- CALDER, A. J., YOUNG, A. W., ROWLAND, D., & PERRETT, D. I. (1997). Computer-enhanced emotion in facial expressions. *Proceedings of the Royal Society B: Biological Sciences*, *264*, 919-925.
- CAMPANELLA, S., HANOTEAU, C., SERON, X., JOASSIN, F., & BRUYER, R. (2003). Categorical perception of unfamiliar facial identities, the face-space metaphor, and the morphing technique. *Visual Cognition*, *10*, 129-156.
- ETCOFF, N. L., & MAGEE, J. J. (1992). Categorical perception of facial expressions. *Cognition*, *44*, 227-240.
- FOX, E., & DAMJANOVIC, L. (2006). The eyes are sufficient to produce a threat superiority effect. *Emotion*, *6*, 534-539.
- FRANKLIN, A., & DAVIES, I. R. L. (2004). New evidence for infant color categories. *British Journal of Developmental Psychology*, *22*, 349-377.
- FUJISAKA, H., & KAWASHIMA, T. (1971). On the modes and hearing mechanisms of speech perception. In *Annual Report of the Engineering Research Institute* (p. 28). University of Tokyo, Faculty of Engineering.
- GOLDSTONE, R. L., LIPPA, Y., & SHIFFRIN, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*, 27-43.
- HARNAD, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 1-28). Cambridge: Cambridge University Press.
- HUTTENLOCHER, J., HEDGES, L. V., & VEVEA, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*, 220-241.
- KAY, P., & KEMPTON, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, *86*, 65-78.
- KOTSONI, E., DE HAAN, M., & JOHNSON, M. H. (2001). Categorical perception of facial expressions by 7-month-old infants. *Perception*, *30*, 1115-1125.
- MALT, B. C., SLOMAN, S. A., GENNARI, S., SHI, M., & WANG, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artefacts. *Journal of Memory & Language*, *40*, 230-262.
- MUNNICH, E., & LANDAU, B. (2003). The effects of spatial language on spatial representation: Setting some boundaries. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 113-156). Cambridge, MA: MIT Press.
- PASTORE, R. E. (1987). Categorical perception: Some psychophysical models. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 29-52). Cambridge: Cambridge University Press.
- PILLING, M., WIGGETT, A., ÖZGEN, E., & DAVIES, I. R. L. (2003). Is color categorical perception really perceptual? *Memory & Cognition*, *31*, 538-551.
- ROBERSON, D., & DAVIDOFF, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition*, *28*, 977-986.
- ROBERSON, D., DAVIDOFF, J., & BRAISBY, N. (1999). Similarity and categorisation: Neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition*, *71*, 1-42.
- ROBERSON, D., DAVIES, I. R. L., & DAVIDOFF, J. (2000). Colour categories are not universal: Replications and new evidence from a Stone-age culture. *Journal of Experimental Psychology: General*, *129*, 369-398.
- ROSEN, S., & HOWELL, P. (1987). Auditory, articulatory, and learning explanations of categorical perception in speech. In S. Harnad (Ed.),

- Categorical perception: The groundwork of cognition* (pp. 113-160). Cambridge: Cambridge University Press.
- SPENCER, J. P., & HUND, A. M. (2002). Prototypes and particulars: Geometric and experience-dependent spatial categories. *Journal of Experimental Psychology: General*, **131**, 16-37.
- STEVENAGE, S. (1998). Which twin are you? A demonstration of induced categorical perception of identical twin faces. *British Journal of Psychology*, **89**, 39-57.
- VALENTINE, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology*, **43A**, 161-204.
- YOUNG, A. W., PERRETT, D. I., CALDER, A. J., SPRENGELMEYER, R., & ETCOFF, N. L. (2002). *Facial Expressions of Emotion: Stimuli and Tests (FEEST)*. London: Harcourt Assessment.
- 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.

NOTES

1. The interference tasks used here were designed by Pilling et al. (2003) for an investigation of the categorical perception of colors.
2. As had been found for color (Pilling et al., 2003).

(Manuscript submitted April 4, 2006;
revision accepted for publication January 10, 2007.)