

Learning & Behavior
2006, 34 (3), 229-240

Determinants of range effects in face recognition

ERIC L. VERBEEK and MARCIA L. SPETCH
University of Alberta, Edmonton, Alberta, Canada

KEN CHENG
Macquarie University, North Ryde, New South Wales, Australia

and

COLIN W. G. CLIFFORD
University of Sydney, Sydney, New South Wales, Australia

The effects of test stimulus range on generalization gradients in humans were assessed for discriminations between faces that varied in brightness, faces that varied in orientation in the picture plane, and morphed faces. In Experiment 1, significant range effects, predicted by adaptation level theory, occurred when faces varied along the brightness or orientation dimension, but not for morphed faces. Changing the difficulty of discrimination of both training and test stimuli for Experiment 2 produced range effects in morphed faces. Experiment 3 explored training and testing stimulus factors as determinants of range effects in morphed faces. The results suggest that sufficiently biased testing ranges create shifts in response distributions (generalization gradients), and this may be amplified by using relatively difficult discriminations between training stimuli.

Responding to a stimulus as though it were another previously experienced stimulus, referred to as *generalization*, is a major feature of behavior. The likelihood of making a false positive identification or a generalized response to a stimulus is a function of its similarity to the target stimulus, and this function has been theoretically described as either a Gaussian or an exponential distribution (Ghirlanda & Enquist, 2003; Shepard, 1987). This distribution is normally centered on the target stimulus, but particular manipulations may reliably shift the distribution. For example, a participant may be trained to discriminate between the target stimulus (S+) and a foil (S-) that differs along a particular stimulus dimension (brightness, vertical orientation, volume). The participant is reinforced for responding to the S+, and not reinforced for responding to the S-. When good discrimination develops, a larger range of stimuli along the manipulated dimension is presented. A common finding is that the distribution of responding is systematically shifted away from the S+. For example, many studies have found that the peak of responding is to a stimulus beyond the S+, in the opposite direction of the S-; this is referred to as the *peak shift* effect (Hanson, 1959; Spence, 1937). This effect is interesting, because the maximum of responding occurs

to a stimulus that is different from the S+. The closer together the S+ and S- are along the stimulus dimension, the greater the peak shift tends to be during testing (Thomas, 1962). Sometimes, the peak of responding does not shift away from the S+, but participants nevertheless respond more to stimuli on the S+ side of the distribution than to stimuli an equivalent distance away from the S+ on the S- side. This effect is referred to as *area shift* (Nallan, McCoy, Pace, & Welch, 1979). Peak or area shifts occur with intradimensional testing wherein test stimuli vary along the dimension in which the discrimination training stimuli differed; these effects do not typically occur with interdimensional testing, which varies the training stimuli in one dimension and the test stimuli in another (Ghirlanda & Enquist, 2003).

Two theories have been proposed that can account for peak shift: Thomas's (1993) adaptation level hypothesis, which is often used to explain peak shift in humans, and Spence's (1937) conditioning account, which is often used to explain the effect in nonhuman animals. Spence's account claims that an excitatory gradient forms around the S+ and an inhibitory gradient forms around the S-. His theory (which was proposed before peak shift had been demonstrated) predicted that if the S+ and S- were sufficiently close, the combination of the two gradients would result in a net excitatory gradient that is shifted away from the S+ in the direction opposite to the S- (Spence, 1937). Thomas (1993) offered a very different account of peak shift, based on Helson's (1947) notion of adaptation level. Thomas's account holds that responding is based on the relationship between the positive stimulus and an adaptation level that develops during training (Thomas, 1993). This adaptation level is thought to reflect the subjective

This research was supported by an NSERC research grant to M.L.S. and a Queen Elizabeth II Fellowship from the Australian Research Council to C.W.G.C. We are grateful to Joel Pearson for help in generating the face stimuli, and to Alisha Brown, Scott McAnsh, and Carlo Dimailig for their help in data collection and experiment preparation. Correspondence concerning this article should be addressed to E. L. Verbeek, Department of Psychology, University of Alberta, Edmonton, AB, T6G 2E9 Canada (e-mail: everbeek@ualberta.ca).

average of the stimulus values experienced during training. When testing proceeds, the adaptation level changes due to exposure to a new set of stimuli, but the learned relation between the adaptation level and the positive stimulus is still used to determine responses. If the new adaptation level differs from that in training, the peak of responding will shift in the same direction, to about the same degree that the adaptation level itself has changed. This hypothesis thus predicts that an asymmetrical change in test range relative to training range will be a critical determinant of the resulting response gradient. Several experiments with humans, using stimulus dimensions such as weight, line orientation, brightness, and color, have confirmed that manipulations of test stimulus range can shift the peak of responding in a manner consistent with the adaptation level hypothesis (MacKinnon, 1972; Thomas, Mood, Morrison, & Wiertelak, 1991; Thomas & Jones, 1962). Particularly powerful evidence for the adaptation level hypothesis comes from demonstrations that, with the appropriate test range manipulation, the response distribution can shift toward the S- side of the distribution, a result that is inconsistent with Spence's theory.

Lewis and Johnston (1999) demonstrated a peak shift effect in humans' face recognition following training in which human participants discriminated between two morphed faces. The S+ (target) and S- (nontarget) faces were selected from a series of blended faces. Blended faces vary in numerous ways, including changes in local features and global facial configuration, and hence morphed faces can be said to differ along a complex dimension. During testing, the participants were presented with additional stimuli that fell between the S- and S+ and on the S+ side of the distribution. The peak of responding was shifted away from the S+ in the direction opposite to the S-. The range of stimuli presented during testing was not manipulated, so that the contribution of adaptation level to this peak shift could not be assessed. These results spurred Spetch, Cheng, and Clifford (2004) to investigate whether range effects could be found with human face stimuli. In their first experiment, participants were trained to discriminate between a pair of faces and to press a "Yes" button for the specified target face (S+), and a "No" button for the other face (S-). The face stimuli were derived from a series of photographs, produced through morphing, that ranged from a photo of a unique face to an averaged photo. Following discrimination training, the participants were given either blocked tests or probe tests in which they were presented with a range of stimuli biased toward either the S+ or the S- end of the series. For either method of testing, the results showed an area shift (more responding on the S+ side than on the S- side), but no range effects were found. Spetch et al. conducted additional experiments with modified designs, but they still failed to demonstrate range effects on face discrimination. Moreover, they demonstrated that the procedure was not at fault for the failure to see a range effect; with a procedure used by Thomas et al. (1991), Spetch et al. found range effects with line tilt stimuli, but not with face stimuli.

Spetch et al. (2004) suggested that "the more complex and multidimensional the stimuli, the more likely area shift rather than range effects will be found" (p. 239). In Experiment 1, we tested whether face stimuli, as complex (multielemental) stimuli, are resistant to range effects even when they vary in only a single dimension. Experiment 1A replicated Spetch et al.'s failure to find adaptation effects in discrimination of average to unique morphed faces, whereas Experiments 1B and 1C, respectively, assessed whether range effects would occur if the face stimuli varied along the brightness dimension (an example of manipulation along an intensity dimension) or in orientation (an example of an arrangement dimension; Ghirlanda & Enquist, 2003). In Experiments 2 and 3, we assessed the role of instruction, discrimination difficulty, and width of the training and testing ranges on the occurrence of range effects with face stimuli. Because previous studies have found the classic peak shift effect with faces, our investigations focused exclusively on range effects. Our studies were designed to determine whether there are any conditions under which range effects can be found with complex stimuli.

EXPERIMENT 1A

In this experiment, we used a series of morphed faces that ranged from an average face to a unique face. We tested for adaptation effects after varying the range of faces experienced during testing. The test range for one group was biased in the region beyond the positive training stimulus, and test range for the other group was biased toward the negative stimulus end. Six faces that were common to both ranges allowed us to assess whether judgments of these faces were influenced by the range experienced. Specifically, a shift in responding toward the range experienced in testing would provide evidence for an adaptation level effect.

Method

Participants. The participants in all of the experiments were University of Alberta undergraduate students enrolled in introductory psychology courses. They received course credit for their participation. Forty participants were included in Experiment 1A (20 female and 20 male).

Design. Participants were randomly assigned in blocks to four conditions ($n = 10$ per condition). Half of the participants in each condition viewed faces from the female stimulus set, and the other half viewed faces from the male stimulus set. The experimental factors consisted of position of the positive stimulus within the stimulus series (toward the unique end or the average end of the series), and test stimulus range (biased toward the side of the positive stimulus or the negative stimulus). The 10 test stimuli presented to each participant formed a within-subjects factor. For statistical analyses, only the stimuli common to both range conditions (the middle 6 stimuli) were included. Thus, the design used for statistical purposes was a mixed design, with range, positive stimulus, and face gender as between-subjects factors (each with two levels) and test stimulus as a within-subjects factor (with six levels).

Stimuli. The stimuli were drawn from those used in Spetch et al. (2004). Briefly, these stimuli had been created by photographing 20 people's faces under controlled conditions and then averaging them

into a single photograph using techniques specified in the original article (p. 225). One original photograph and the average photograph were then combined through weighted averaging to create 41 stimuli between them, resulting in a morphed series. Weighting began at 0% average and 100% unique, and progressed in increments of 2.5%.

For the present experiment, one set of each gender was used. Of the 41 stimuli in each set, every 2nd stimulus, beginning at the 8th stimulus and ending at the 34th, was selected. This resulted in 14 stimuli per set, each 5% apart, that ranged from 82.5% unique to 17.5% unique. The pictures were grayscale bitmaps with on-screen dimensions of 4.3 × 5.0 cm (distance between upper jaw tips × distance from top of forehead to tip of chin). The monitors were 17-in. CRT monitors with a 5:4 aspect ratio. Display resolution was 600 × 480 pixels, and stimuli were displayed on a black background. The participants, seated, viewed the screen from a distance of 30–60 cm.

Procedure. The stimuli were presented on 17-in. CRT monitors, and auditory feedback during training was presented through headphones. The participants were told they would learn to discriminate between a pair of similar-looking photographs of faces. On-screen instructions informed the participants that the face would appear for only 1.5 sec, that the first face would be the target face, and that they should press the space bar whenever they saw the target face and refrain from doing so for all others.

When training commenced, the participants received auditory feedback after each trial, indicating whether responses were correct or wrong. The face stimuli were presented for a maximum of 1.5 sec or until the space bar was pressed. The interstimulus interval was 3 sec, during which time feedback was given, and the screen was completely black.

During training, the participants learned to discriminate between the 42.5% and 57.5% morphed faces (6th and 9th stimuli of the 14 in testing; see Table 1). Depending on the condition, one of these

served as the target (positive) stimulus and the other served as the negative stimulus.

The training phase included at least two blocks of eight stimulus presentations, the first of which had the same order of positive (+) and negative (-) stimuli across all participants: +, -, +, -, +, +, -, -. The second block presented four "+" and four "-" stimuli in a random order. Testing commenced if participants made six or more correct responses within the second block. If fewer than six responses were correct, additional eight-trial blocks of randomly ordered trials were presented until six or more correct trials occurred within a block.

At the onset of testing, on-screen instructions informed the participants that they were to continue the same task of recognizing the target face, except that additional face stimuli would be presented and there would no longer be auditory feedback on performance. During testing, each participant was presented with 10 stimuli, which were each presented 8 times in random order. For all participants, these included both training stimuli (6th and 9th positions), 2 stimuli between the training stimuli (7th and 8th positions), and 1 outside of each training stimulus (5th and 10th positions). Participants in the positive range condition received 4 additional stimuli from beyond the positive end of their training range, whereas participants in the negative range condition received 4 stimuli from beyond the negative training stimulus. Common test stimuli thus ranged from 37.5% to 62.5%, and the entire range from which test stimuli were selected was from 17.5% to 82.5% morphed. See Table 1 for further details, and Figure 1 for examples.

Analysis. Responses to the test stimuli were recoded on the basis of their serial relation to the positive stimulus used in training. The dependent measure was the proportion of positive responses (trials on which the space bar was pressed) to each test stimulus. To assess whether the distribution of responses shifted as a result of our range manipulation, a weighted peak statistic was calculated for each par-

Table 1
Design Used in All Three Experiments

Experiment	12	17	22	27	32	37	42	47	52	57				
Experiment 1C (degrees from vertical picture plane orientation)														
Lower range	12	17	22	27	32	37	42	47	52	57				
Upper range					32	37	42	47	52	57	62	67	72	77
Experiment 1B (relative brightness difference in CorelDraw brightness units)														
Lower range	-14	-12	-10	-8	-6	-4	-2	0	2	4				
Upper range					-6	-4	-2	0	2	4	6	8	10	12
Experiment 2B (relative brightness difference in CorelDraw brightness units)														
Lower range	-6	-5	-4	-3	-2	-1	0	1	2	3				
Upper range					-2	-1	0	1	2	3	4	5	6	7
Experiment 1A (% morph between unique and average face)														
Lower range	17.5	22.5	27.5	32.5	37.5	42.5	47.5	52.5	57.5	62.5				
Upper range					37.5	42.5	47.5	52.5	57.5	62.5	67.5	72.5	77.5	82.5
Experiment 2A (% morph between unique and average face)														
Lower range	2.5	10.0	17.5	25.0	32.5	40.0	47.5	55.0	62.5	70.0				
Upper range					32.5	40.0	47.5	55.0	62.5	70.0	77.5	85.0	92.5	100
Experiment 3, Testing Width Compressed, Training Width Narrow (% morph between two unique faces)														
Lower range	23	26	29	32	36	41	50	55	59	64	68			
Upper range				32	36	41	45	50	59	64	68	72	76	80
Experiment 3, Testing Width Compressed, Training Width Wide (% morph between two unique faces)														
Lower range	23	26	29	32	36	41	50	55	59	64	68			
Upper range				32	36	41	45	50	59	64	68	72	76	80
Experiment 3, Testing Width Extended, Training Width Narrow (% morph between two unique faces)														
Lower range	0	8	16	24	32	36	41	50	59	64	68			
Upper range					32	36	41	50	59	64	68	76	84	92
Experiment 3, Testing Width Extended, Training Width Wide (% morph between two unique faces)														
Lower range	0	8	16	24	32	36	41	50	59	64	68			
Upper range					32	36	41	50	59	64	68	76	84	92

Note—There were 14 stimuli in each set; the values in the table represent absolute stimulus steps. Values in boldface acted as training values (S+ and S- counterbalanced). Those highlighted in gray were in the common testing range and were used for analysis. "Lower range" and "Upper range" correspond to the range conditions (positive or negative) depending on the value of the positive stimulus for a given participant.

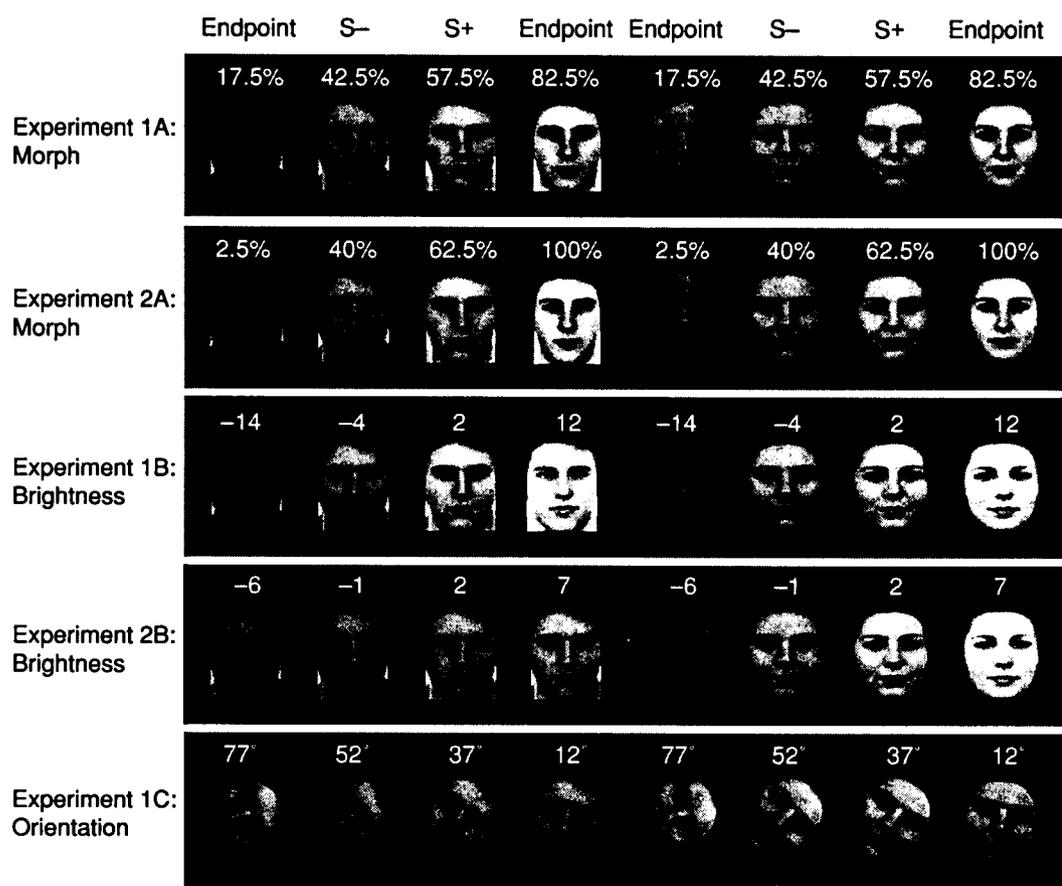


Figure 1. Sample face stimuli used in Experiments 1 and 2. Rows correspond to experiments: (A) morph; (B) brightness; (C) orientation. Stimulus identity is designated by column, as labeled. Only one endpoint would be seen per participant, as described in the experimental designs. Stimulus values appear above each face. Positive and negative stimulus values were counterbalanced across participants.

participant using only the six stimuli common to both range groups (essentially a weighted mean for frequencies of interval data; see Thomas, 1962, and Hays, 1994, p. 173). The peak was calculated by multiplying the morph percentage of each test stimulus within the common range by the proportion of responses given to that stimulus, and dividing by the sum of all six proportions.

A univariate ANOVA was performed on this weighted peak statistic, with range, positive stimulus, and face gender as factors, each with two levels. An adaptation effect would be expected to shift the response gradient in the direction of the padded range. Therefore, a significant range effect in which the peak is smaller for the negative range group than for the positive range group would provide evidence of an adaptation effect.

Predicted peaks were found by first getting the weighted average of the test stimuli, to find the final adaptation level, and then adding the distance between S+ and the training AL. This method is useful for determining ordinal predictions to compare conditions, but it is not intended to provide accurate quantitative predictions, because we do not have subjective scales for our stimuli (see Thomas, 1993). In addition, the predicted values are very likely to yield overestimations of the range effects, since the adaptation levels likely do not shift fully by the end of testing, and the predicted values refer to the location of peak responding rather than the weighted peak statistic used in the analysis.

Results and Discussion

The ANOVA on the weighted mean position scores (positive range = 53.3%, and negative range = 51.8%) did not reveal a significant range effect [$F(1,32) = 0.452$, $p > .05$]. The predicted values were 67.5% and 47.5%, respectively. No other effects in the model were significant. Thus, this experiment failed to reveal evidence of an adaptation level effect, and it therefore replicated the results found by Spetch et al. (2004). In Figure 2, one can see that the six common values are virtually parallel for the two range groups, and the greater overall responding in the negative group is not indicative of a range effect.

EXPERIMENT 1B

The lack of range effect with morphed faces was consistent with previous results (Spetch et al., 2004) and led to a consideration of what might block the effect. The morphed dimension is complex, in that the stimuli change in configuration, at both an elemental level and in their entirety. In Experiment 1B, we varied the brightness of

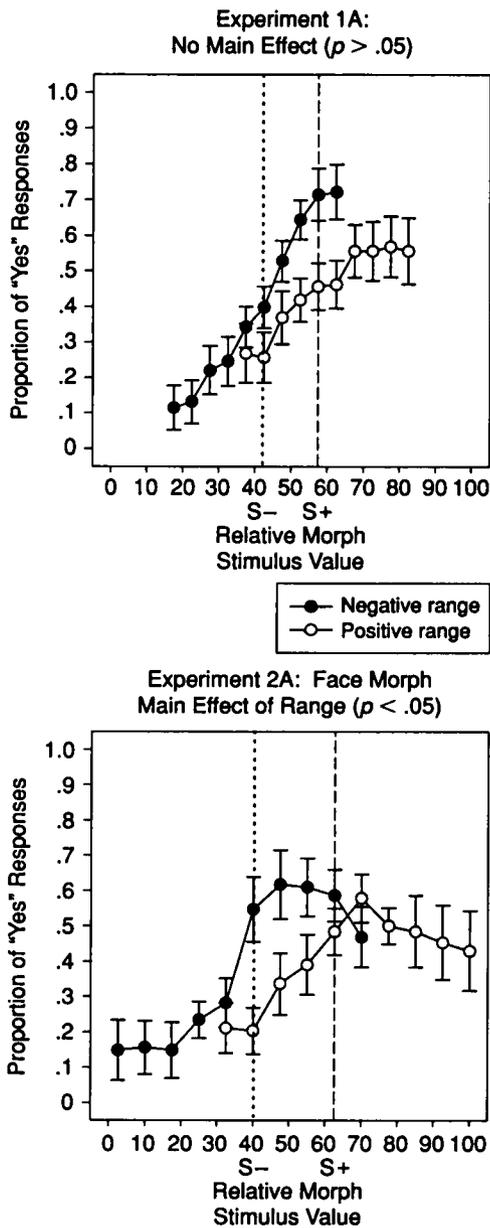


Figure 2. Proportion of "Yes" responses to test stimuli in morphed face experiments, broken down by stimulus range condition (means \pm standard errors). Face gender and positive stimulus conditions have each been collapsed across. Only stimuli common to both ranges were used in analysis.

the face images to produce a stimulus series that instead varied along a simple intensity dimension (Ghirlanda & Enquist, 2003). If the absence of an adaptation effect was due to the complexity of the changes in morphed faces, then adaptation effects might occur if the faces were varied along one simple dimension, such as brightness. The effect should fail to occur, however, if faces themselves were somehow resistant to adaptation effects, perhaps due to stimulus complexity, as hypothesized by Spetch et al. (p. 239).

Method

Participants. Forty University of Alberta undergraduate students were included in Experiment 2 (23 female, 17 male).

Design. The design of this experiment was identical to that of Experiment 1A, except that the stimulus series consisted of faces that varied along the brightness dimension.

Stimuli. From the 41 photographs available from the original morph set, 1 was selected from each gender (the 18th face stimulus of the originals). This original was edited with CorelDraw to create 14 faces that ranged in steps of 2 from +12 to -14 brightness units away from the original. The face at +12 units was the lightest face, and the -14 face was the darkest. The positive and negative training values were -4 and 2; the common testing range was from -6 to 4.

Procedure. The procedure was the same as in Experiment 1A, except that the instructions informed participants that the pictures would vary in brightness, and that they should pay attention to brightness in order to respond correctly.

Results and Discussion

As predicted by the adaptation level hypothesis, participants in the positive range condition had a higher mean response position score (1.45) than did participants in the negative range condition (-1.0) [$F(1,32) = 19.269, p < .05$]. Predicted values were 6 and -2, respectively. No other effects were significant. In Figure 3, one can see a distinct shift in the two gradients, which otherwise appear to have an identical overall shape. Each group thus demonstrated a shift in responding toward the side that was padded with more test stimuli, as was to be expected if the adaptation level shifted during testing and responding continued to be based on the relation between the training adaptation level and the target stimulus. This result is very interesting, because, despite an identical procedure, Experiment 1A did not yield this result. This implicates the complex changes occurring in morphed faces in the lack of a range effect.

EXPERIMENT 1C

In Experiment 1C, faces were varied in their orientation. Orientation is an example of an arrangement dimension (Ghirlanda & Enquist, 2003), but, like brightness, it is a much simpler dimension than morphed faces. Because the faces differed only in this one dimension, we expected that the results would be similar to those found in Experiment 1B for variations in brightness.

Method

Participants. Forty University of Alberta undergraduate students were included in Experiment 1C (21 female, 19 male).

Design. This design was the same as in the previous pair of experiments, except that the stimulus series consisted of faces that varied along the orientation dimension.

Stimuli. The faces were the same as those used in Experiments 1A and 1B; but they were cropped to fit within an ellipse, and then a stimulus series was created by manipulating the orientation of the faces in the picture plane using CorelDraw. The series ranged from 12° (i.e., close to vertical) to 77° (i.e., close to horizontal) in 5° increments. This placed the positive and negative training stimuli at 37° and 52°, centered on 44.5°, and the common test range was from 32° to 57°.

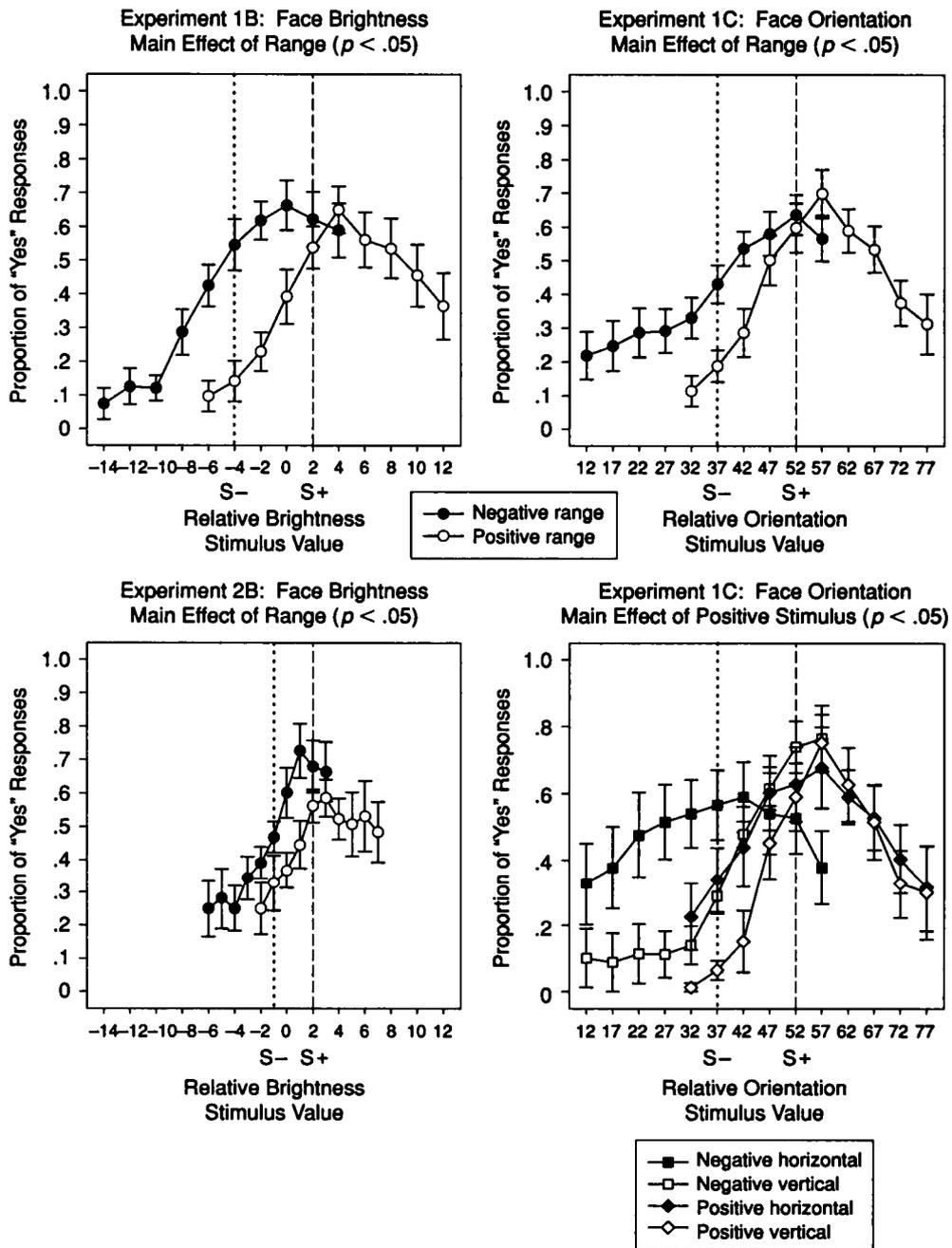


Figure 3. Proportion of "Yes" responses to test stimuli in brightness (1B and 2B) and orientation (1C) experiments, broken down by stimulus range condition (means \pm standard errors). Face gender and positive stimulus conditions have each been collapsed across in the first three graphs. The fourth plot collapses across stimulus range, showing gradients for positive stimulus conditions. Only stimuli common to both ranges were used in analysis.

Procedure. This procedure was identical to that in the previous experiments, except that the participants were told that the photographs would vary in orientation and that they were to identify the target stimulus on the basis of this property. The participants were also instructed to not touch the screen in any way. This prevented them from using a mark on the screen or their finger as a cue. An experimenter was in the room at all times, and no transgression of this rule was observed.

RESULTS AND DISCUSSION

The mean response position was higher in the positive range condition (50.7°) than in the negative range condition (45.3°), which indicates an adaptation effect [$F(1,32) = 16.112, p < .05$]. The predicted values were 62° and 42° ,

respectively. The plot in Figure 3 shows the two range gradients, and it is clear that the response gradients differed depending on which side was padded with more stimuli, as predicted by the adaptation level hypothesis.

There was also a significant effect of which stimulus was positive during training (more vertical or horizontal) on the weighted peak location (vertical condition at 50.0°, horizontal condition at 45.9° [$F(1,32) = 9.29, p < .05$]). Recall that for analysis and plots, stimulus values were re-coded relative to the positive stimulus value during testing, to place expected peaks for each group in the same region. This effect of positive orientation can be seen in Figure 3, which shows the gradients of four groups, composed by crossing the range and positive stimulus conditions. Discrimination between stimuli was better (i.e., a higher peak and a steeper slope away from this peak) in the groups for which the positive stimulus was closer to vertical than in the group for which the positive stimulus was closer to horizontal. The curves for the horizontal positive groups (particularly the negative range subset), relative to those for the vertical positive groups, are flatter and shifted toward the negative side. One plausible explanation for this result is that humans might be more sensitive to changes in the vertical orientation of faces because of greater experience in viewing vertically oriented faces than in viewing horizontally orientated faces. Indeed, Collishaw and Hole (2002) found that when faces were blurred to disrupt their featural information and then were subsequently tilted away from vertical, recognition became poorer (see also Appelle, 1972). This suggests that the configural processing is more difficult for faces that are tilted away from vertical, which might cause participants in the horizontal positive condition to be less accurate in discriminating the target orientation from the other orientations. Extensive prior experience with vertically oriented faces might set up a strong preexperimental adaptation level of vertically oriented faces, which is not completely overcome by a brief experimental experience. A near vertical S+ works in concert with the preexperimental adaptation level, whereas a near horizontal S+ works against the preexperimental adaptation level.

EXPERIMENT 2

The results of Experiments 1B and 1C clearly indicate that faces per se are not resistant to range effects. Robust range effects were found when people discriminated between faces that varied along the single dimensions of brightness and orientation. At the same time, we replicated the findings of Spetch et al. (2004) in failing to find a range effect when people discriminated among faces that varied along the complex dimension created by morphing faces. However, in addition to the complexity of the dimensions, there were two other potentially important differences between the morph manipulation and the manipulations of brightness and orientation. First, the instructions differed because participants were alerted to the dimension varied (different-looking faces in Experiment 1A, photographs

varying in brightness or orientation in Experiments 1B and 1C, respectively). Second, the morph discrimination may have been more difficult than the brightness and orientation discriminations. Indeed, the average number of training trials required in order to meet the criterion to move to testing was higher in Experiment 1A (50.0) than in Experiment 1B (37.1) or Experiment 1C (23.0). Therefore, in Experiment 2A, we replicated the conditions of Experiments 1A and 1B, but with two changes. First, we used identical instructions for both the morph and the brightness discrimination. Second, we made the morph discrimination easier by selecting values during both training and testing that were spaced farther apart, and the brightness discrimination harder by spacing the stimuli closer together.

Experiment 2A

This experiment was a replication of Experiment 1A, with training stimulus discriminability increased, and with greater distance between adjacent test stimuli, and thus a greater distance between the test series endpoints.

Method

Participants. Thirty-two University of Alberta undergraduate students were included in Experiment 2A (22 female, 10 male).

Design. This design was the same as in the previous experiments. The stimulus series consisted of morphed faces selected from the same large stimulus set as in Experiment 1A.

Stimuli. The total range of the 14 stimuli was from 2.5% to 100% morphed from unique to average, with 7.5% steps between stimuli. Training stimuli were 40% and 62.5%, and the common test range was 32.5% to 70%. The comparison between Experiment 1A and Experiment 2A can be seen in Table 1.

Procedure. Procedures were the same as in Experiment 1A, including the instructions, which specified that the participants would need to discriminate between faces that looked very similar.

Results and Discussion

The stimulus changes did indeed allow for a range effect; the mean response position was 51.3% for the negative range condition and 58.9% for the positive range condition [$F(1,32) = 10.23, p < .05$]. The predicted values were 47.5% and 77.5%, respectively. The results are plotted in Figure 2, wherein one can see a distinctly different center for each response gradient. The effect contrasts with that in Experiment 1A, in which there was a greater overall rate of responding in the negative group, but no obvious difference in gradient location. The average number of training trials to criterion was 42.75; it was 50.2 for Experiment 1A.

Experiment 2B

This experiment was a replication of Experiment 1B, with training stimulus discriminability decreased, and with a test stimulus range that had less distance between adjacent stimuli. The instructions were also identical to those in Experiments 1A and 2A.

Method

Participants. Thirty-two University of Alberta undergraduate students were included in Experiment 2B (23 female, 9 male).

Design. This design was the same as in the previous experiments. The stimulus series consisted of brightness modified faces produced in the same manner as described in Experiment 1B.

Stimuli. The 14 stimuli ranged from -6 to 7 brightness units, in steps of 1 unit. Training stimuli were -1 and 2, and thus the common test range was from -2 to 3. Comparisons between Experiment 1B and 2B may be made by using Table 1.

Procedure. The only change in procedures was the neutralization of the instructions, which now told participants that they needed to discriminate between faces that looked quite similar, rather than between faces that differed in brightness (as in Experiment 1B).

Results and Discussion

Despite increasing training difficulty (mean 48.0 trials to criterion in Experiment 2B; 31.2 in Experiment 1B) and decreasing test range, a range effect still developed; the negative range peak (0.63) and positive range peak (1.27) were significantly different [$F(1,32) = 5.39, p < .05$]. The predicted values were 0 and 4, respectively. The results are plotted in Figure 3. Thus, range effects with faces that varied in the brightness dimension were robust across variations in instruction and in the extent of the training and test range.

EXPERIMENT 3

The contrast between the results of Experiments 1A and 2A suggested that either the training discriminability or the extent of the test range might be important factors in the occurrence of range effects with morphed faces. Previous studies suggested that both of these factors could influence the magnitude of range or peak shift effects with other stimulus dimensions (Baron, 1973; Thomas et al., 1991). Experiment 3 was designed to determine the effect of training discriminability and of the magnitude of range manipulation on range effects with morphed faces.

Method

Participants. Ninety-six University of Alberta undergraduate students participated in Experiment 3; data for 85 were kept (53 female, 32 male), and 11 participants failed to pass training (unrecorded gender). Five were omitted from the statistical analysis because they failed to respond to the common stimulus range, thus making it impossible to calculate a peak within that range. These participants did respond to stimuli outside that range, and are thus included in Figure 5.

Design. Building from the previous designs, this experiment includes two new between-subjects variables: training width and testing width. Training width was the distance between the S+ and S-, and it was either wide (36% and 64%) or narrow (41% and 59%). Testing width corresponded to the total test stimulus range and was either compressed (23%–80%) or extended (0%–100%). The test stimulus intervals were uneven within a series. This allowed us to produce particular differences between the training and testing adaptation level: the difference was approximately 7% of morph for the compressed range, and 14% of morph for the extended range. Table 1 details the exact setup of stimulus values within conditions. The complete set of factors includes testing width (compressed or extended), training width (narrow or wide), range (positive or negative), positive stimulus (above or below 50% morph stimulus), and face gender (male or female).

Stimuli. New face stimuli were used for this experiment, both to generalize to a new set of faces, and to increase the difference be-

tween the two ends of the series. Whereas the previous stimuli were morphed between a single unique face and an averaged face, the faces for Experiment 3 were morphed between two very different-looking unique faces. The original faces were acquired from the Max Planck Institute for Biological Cybernetics (Department Bülthoff URL, www.kyb.mpg.de/bu/index.html; Face Database URL, faces.kyb.tuebingen.mpg.de/). These faces were produced from 3-D scans of people's heads; the scans were then strategically morphed together to eliminate individual identities from the face database. We selected particular face pairs from the 200 available front-facing faces, and manipulated them with the morphing procedure used by Spetch et al. (2004). One female and one male set were used. New 17-in. monitors were used for this study. The monitors had an aspect ratio of 4:3; the resolution was set to 1,024 × 768 pixels, and the faces were approximately 4 × 6 cm (distance between upper jaw tips × crown top to chin tip). The faces were presented in color rather than in grayscale as in the previous studies. Sample stimuli appear in Figure 4.

Procedure. The procedure was a slightly modified version of that in the previous experiments. The instructions were the same as in Experiment 1A. The training criterion was changed to 20 out of 24 trials correct, over three blocks of training trials. Eleven participants who did not meet this criterion within 30 min of training were discontinued prior to testing. For the 85 participants who met criterion, seven blocks of the 11 randomly ordered test stimuli were presented. Each subject's response proportions were thus made up of 7 responses at each stimulus value. All other procedural details were the same as in previous experiments.

Results and Discussion

The data were analyzed with an ANOVA with five between-subjects factors: training width (narrow or wide), test width (compressed or extended), range (positive or negative), positive stimulus value, and face gender. This analysis revealed significant main effects for range [$F(1,48) = 20.79, p < .05$] and test width [$F(1,48) = 11.00, p < .05$], and significant interactions of training width × range [$F(1,48) = 6.22, p < .05$] and test width × range [$F(1,48) = 10.63, p < .05$]. Analysis of the simple effects on the training width × range interaction revealed that there were range effects only for the narrow training width condition [narrow, $F(1,48) = 22.66, p < .05$; wide, $F(1,48) = 2.36, p > .05$]. Simple effects analysis of the range × testing width interaction revealed that range effects occurred only for extended conditions [extended, $F(1,48) = 29.15, p < .05$; compressed, $F(1,48) = 0.89, p > .05$]. Plots of the results for the four combinations of training and test width are presented in Figure 5.

Additional planned ANOVAs on each of the four groups produced by crossing training width and testing width revealed a significant range effect in both narrow training groups. In the extended test width condition, the negative range peak was 44.8%, and the positive peak was 62.1% [$F(1,8) = 16.34, p < .05, \eta_p^2 = .67$]. The predicted values were 45.2% and 72.8%, respectively. The 5 participants who were excluded from the analysis because they did not respond to any of the common test stimuli were all from the narrow extended condition. All of these participants responded to the noncommon stimuli on the padded end of the test stimulus range and thus demonstrated extreme response shift; this means that their exclusion would lead to an underestimation of the response shift (see Figure 6). It is unlikely that these outliers exemplified a powerful ad-

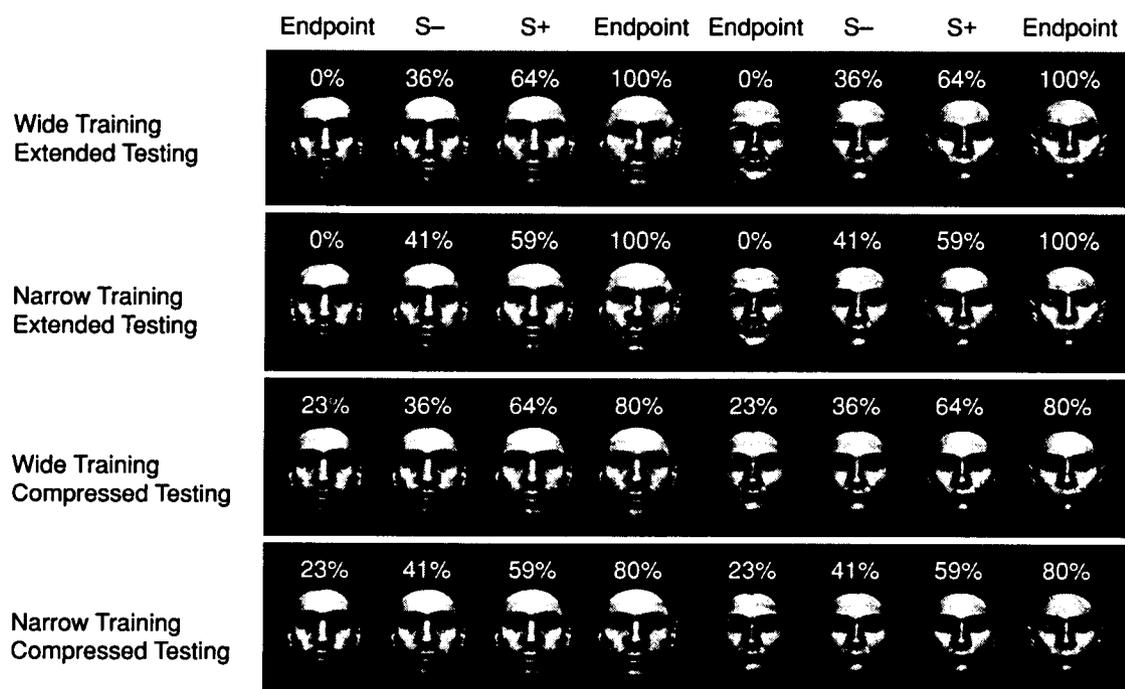


Figure 4. Sample face stimuli used in Experiment 3. Rows correspond to conditions as labeled, and stimulus identity is designated by column labels. Only one endpoint would be seen per participant, as described in the experimental designs. Stimulus values appear above each face. Positive and negative stimuli were counterbalanced across participants.

aptation effect. Instead, they may have forgotten the target face far more than other participants, or perhaps they used a different response strategy. In the narrow training and compressed testing width, the negative range peak was 57.9%, and the positive peak was 61.6% [$F(1,13) = 5.99$, $p < .05$, $\eta_p^2 = .32$]. The predicted values were 52.9% and 65.6%, respectively. The range effect just missed significance in the wide training and extended testing condition (the negative peak was 51.7%, and the positive peak was 58.5%) [$F(1,15) = 3.05$, $p > .05$, $\eta_p^2 = .17$]. The predicted values were 50.2% and 77.8%, respectively. The range effect did not approach significance in the wide training and compressed test range condition (a negative range peak of 58.2% and a positive peak of 57.8%) [$F(1,12) = 0.004$, $p > .05$, $\eta_p^2 = 0$]. The predicted values were 57.9% and 70.6%, respectively.

These results suggest that the occurrence of a range effect with morphed faces depends both on the discriminability of the training stimuli and on the extent of the range manipulation. Larger range effects occur with a large change in adaptation level, produced by an extensive range manipulation, and with a more difficult discrimination.

GENERAL DISCUSSION

These three experiments demonstrate that range effects can be found with face stimuli, but that only certain conditions allow it to emerge. First, range effects appear readily when faces are varied along the simple dimensions of brightness and orientation. Second, range effects appear

with morphed faces only when the training conditions are relatively difficult, and only when the range manipulation produces a sufficiently great change in adaptation level.

Although we did not find a range effect in the morphed faces of Experiment 1A, increasing the range manipulation and slightly increasing the discriminability of the training stimuli resulted in a range effect in Experiment 2A. Conversely, when faces varied in brightness, a range effect occurred in Experiment 1A and remained when we weakened the range manipulation and decreased discriminability in Experiment 2B. In Experiment 3, we varied both discriminability and extent of range manipulation for morphed faces, and we found that both factors were important. Specifically, a difficult training discrimination and a wide test range manipulation each contributed to the occurrence of a range effect. Indeed, when these conditions were both present (extended range, narrow discrimination), a very sizable range effect occurred, and when they were both absent (compressed range, wide discrimination), no range effect occurred. Thus, range effects appear to readily occur when faces are varied along a simple dimension, but when they are varied along the complex morphed dimension, training difficulty and extent of the range manipulation are important factors.

Direct comparisons between Experiment 3 and the morphed face conditions of the first two experiments are problematic because of differences in the faces used, and the way in which the morphed range was created (i.e., morphing between a unique and an average face vs. morphing between two unique faces). However, given the

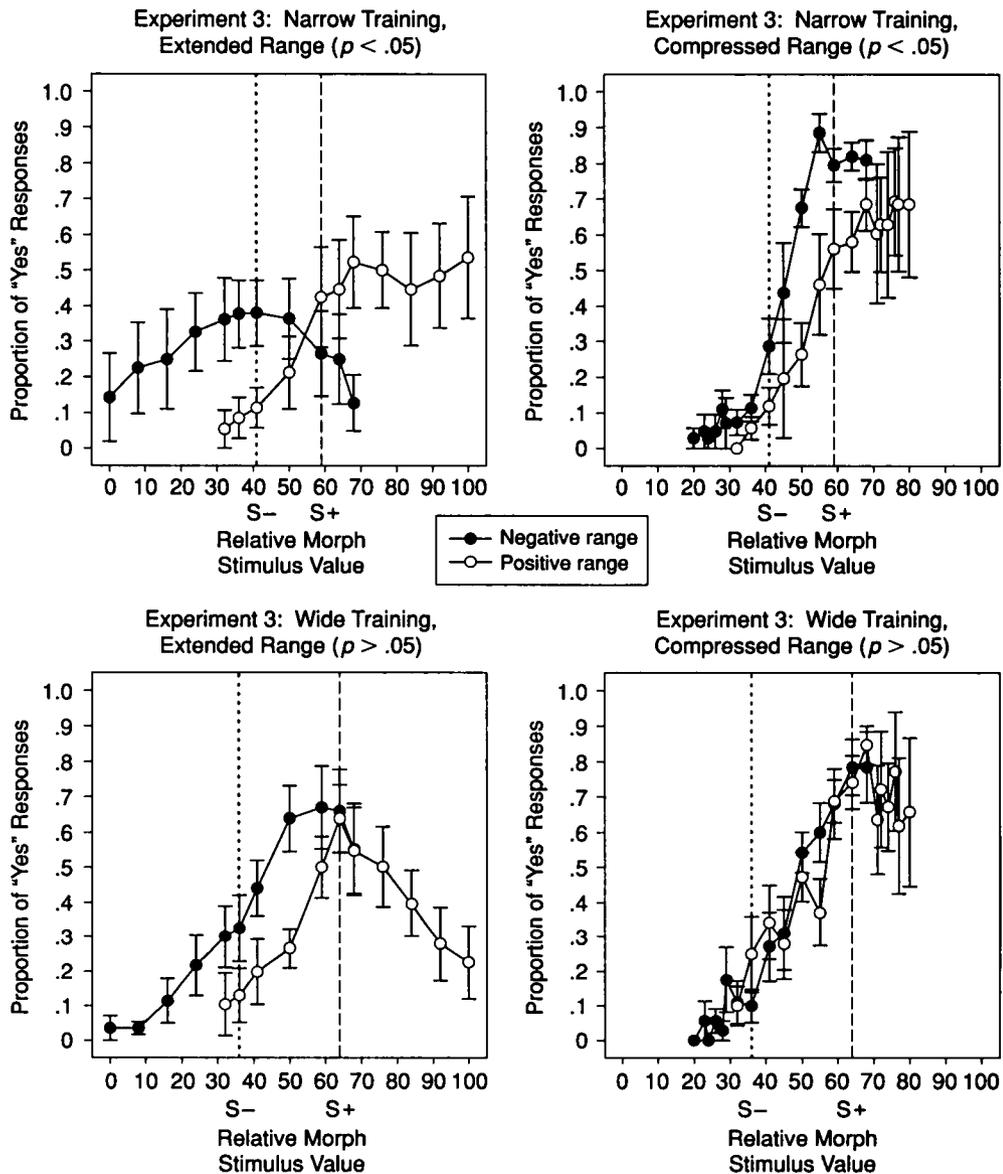


Figure 5. Proportion of "Yes" responses to test stimuli in Experiment 3 for crossings of training width and range width conditions, broken down by stimulus range condition (means \pm standard errors). Face gender and positive stimulus conditions have each been collapsed across. Only stimuli common to both ranges were used in analysis.

context of Experiment 3, it is interesting that no range effects were found in Experiment 1A, which had a relatively difficult discrimination, and a weak range manipulation, whereas range effects were found in Experiment 2A, which had an easier discrimination and a much stronger range manipulation. It seems likely that the increased range manipulation was responsible for the occurrence of range effects in Experiment 2A; but with morphed faces, there may also be an optimal level of discriminability for producing range effects, and this level is possibly dependent on the strength of the range manipulation. By contrast, the occurrence of range effects on the brightness dimension

does not appear to require a difficult discrimination: The participants in Experiment 1B learned the brightness discrimination readily yet showed strong range effects.

The impact of extent of range manipulation makes sense, according to the adaptation level hypothesis: if encoding and responding are dependent on the adaptation level, and if the adaptation level changes, then response rate to a given stimulus will change following the shift in adaptation level. The greater the change in adaptation level, the greater the change in the location of peak responding.

The relationship between training difficulty and range effects is less clear. Thomas and colleagues' previous pre-

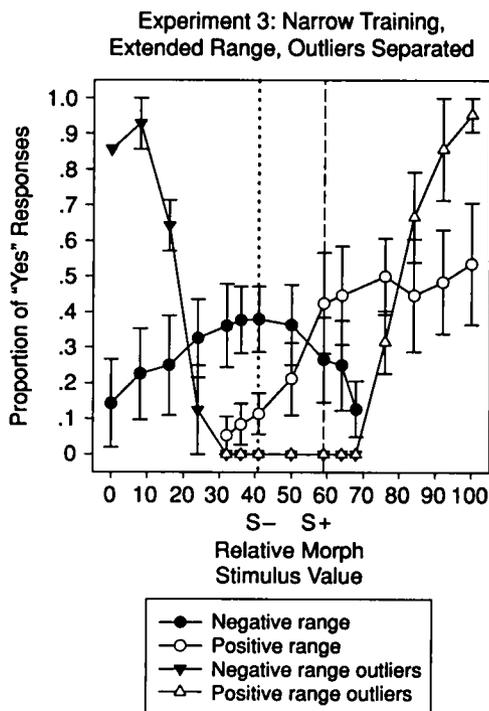


Figure 6. Proportion of “Yes” responses to test stimuli in Experiment 3 for the narrow training width and extended range width condition, broken down by stimulus range condition, and by outlier status (means \pm standard errors). Face gender and positive stimulus conditions have each been collapsed across. This shows the difference between participants used in the analysis and those who were necessary to leave out.

ditions and empirical results suggest that range effects should increase with greater discriminability between training stimuli (Thomas et al., 1991; Thomas, Svinicki, & Vogt, 1973). Although the range effects seen with the brightness dimension in Experiments 1B and 2B seem consistent with this prediction, the results of Experiment 3 are not, because range effects for morphed faces were stronger with the more difficult discrimination. Similarly, Baron (1973) found that humans showed a larger response shift during generalization testing for discriminations with a narrow difference (100 Hz) than during generalization for discriminations with a wider difference (200 Hz) in tone frequency. Clearly, more research is needed in order to determine why increases in discriminability of the training stimuli sometimes increase and sometimes decrease range effects.

Overall, our results clearly show that range effects can occur with complex face stimuli. When face stimuli are varied along the complex dimension produced by morphing between individuals, the range effect is fragile and is sensitive to both the discriminability of the training stimuli and the extent of the range manipulation. Nevertheless, with a difficult discrimination and an extreme range manipulation, we found a large range effect with morphed faces, thus lending further support for Thomas’s (1993) contention that adaptation level is an important de-

terminant of stimulus generalization in humans. It is very likely that Spetch et al. (2004) did not find range effects in morphed faces because of smaller range manipulations; common stimulus regions were larger in some cases, and ranges differed in width rather than in direction, which has less power than the bidirectional approach used in the present study.

Although our results clearly show that range effects can occur when complex stimuli are varied along a complex dimension, the present results, together with those of Spetch et al. (2004), suggest that such effects may be more sensitive to procedural variations than range effects that occur with the use of simple stimuli. Moreover, the possibility exists that some participants may have attempted to simplify the complex stimuli into simple dimensions for discrimination—into, for example, head width or nose length. In future research, it would be interesting to include probe tests in which a single stimulus feature is varied in order to determine the strategy used, and then to compare generalization and range effects under conditions in which this strategy can be used and one in which it is prevented.

There is one provocative set of results demonstrating what appears to be a natural occurrence of range effects in face perception. Webster, Kaping, Mizokami, and Duhamel (2004) found that after adaptation to an exemplar from the “Asian face” category, the categorical boundary between Asian and Caucasian moved further into the Asian category: A face that was previously seen as being midway between Caucasian and Asian was seen as clearly Caucasian after adaptation to an Asian face. Although their experiment dealt with effects that are better termed adaptation aftereffects (due to the very short time frame involved; see Leopold, O’Toole, Vetter, & Blanz, 2001, and Webster & MacLin, 1999), some further correlational results appear to match our findings of range effects in faces. Asian students who had been in North America for at least 1 year set boundaries different from those set by their newly arrived exchange student counterparts, and they showed a negative correlation between how long they had been in North America and how Asian their categorical boundary was. In other words, students who had been in North America for a long time were more likely to judge a face in the middle of the spectrum as being Asian than were those who had been in North America for a shorter period of time. There was also a positive correlation between that boundary and the amount of time they reported spending with people of the same ethnicity. Thus, greater exposure to Caucasians caused the Asian students’ category boundary to become less Asian and more Caucasian. Although Thomas’s (1993) adaptation level hypothesis focuses on explaining peak shift in generalization testing, it might also have implications for the role of adaptation level in determining categorical boundaries for stimuli. If the encoding of categorical boundaries is dependent on adaptation level, these boundaries should move in the same direction as the adaptation level moves, and experimental results that parallel Webster et al.’s correlational findings might be possible.

REFERENCES

- APPELLE, S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. *Psychological Bulletin*, *78*, 266-278.
- BARON, A. (1973). Postdiscrimination gradients of human subjects on a tone continuum. *Journal of Experimental Psychology*, *101*, 337-342.
- COLLISHAW, S. M., & HOLE, G. J. (2002). Is there a linear or a nonlinear relationship between rotation and configural processing of faces? *Perception*, *31*, 287-296.
- GHIRLANDA, S., & ENQUIST, M. (2003). One century of generalization. *Animal Behaviour*, *66*, 15-36.
- HANSON, H. M. (1959). Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology*, *58*, 321-334.
- HAYS, W. L. (1994). *Statistics* (5th ed.). Belmont, CA: Wadsworth.
- HELSON, H. (1947). Adaptation-level as frame of reference for prediction of psychophysical data. *American Journal of Psychology*, *60*, 1-29.
- LEOPOLD, D. A., O'TOOLE, A. J., VETTER, T., & BLANZ, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, *4*, 89-94.
- LEWIS, M. B., & JOHNSTON, R. A. (1999). Are caricatures special? Evidence of peak shift in face recognition. *European Journal of Cognitive Psychology*, *11*, 105-117.
- MACKINNON, M. (1972). Adaptation-level theory, anchor theory, and the peak-shift phenomenon. *Journal of Motor Behavior*, *4*, 1-12.
- NALLAN, G. B., MCCOY, D. F., PACE, G. M., & WELCH, R. (1979). Generalization gradients following differential intradimensional auto-shaping. *Perceptual & Motor Skills*, *48*, 671-677.
- SHEPARD, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317-1323.
- SPENCE, K. W. (1937). The differential response of animals to stimuli differing within a single dimension. *Psychological Review*, *44*, 430-444.
- SPETCH, M. L., CHENG, K., & CLIFFORD, C. W. G. (2004). Peak shift but not range effects in recognition of faces. *Learning & Motivation*, *35*, 221-241.
- THOMAS, D. R. (1962). The effects of drive and discrimination training on stimulus generalization. *Journal of Experimental Psychology*, *64*, 24-28.
- THOMAS, D. R. (1993). A model for adaptation-level effects on stimulus generalization. *Psychological Review*, *100*, 658-673.
- THOMAS, D. R., & JONES, C. G. (1962). Stimulus generalization as a function of the frame of reference. *Journal of Experimental Psychology*, *64*, 77-80.
- THOMAS, D. R., MOOD, K., MORRISON, S., & WIERTELAK, E. (1991). Peak shift revisited: A test of alternative interpretations. *Journal of Experimental Psychology: Animal Behavior Processes*, *17*, 130-140.
- THOMAS, D. R., SVINICKI, M. D., & VOGT, J. (1973). Adaptation level as a factor in human discrimination learning and stimulus generalization. *Journal of Experimental Psychology*, *97*, 210-219.
- WEBSTER, M. A., KAPING, D., MIZOKAMI, Y., & DUHAMEL, P. (2004). Adaptation to natural facial categories. *Nature*, *428*, 557-561.
- WEBSTER, M. A., & MACLIN, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin & Review*, *6*, 647-653.

(Manuscript received September 24, 2004;
revision accepted for publication February 3, 2006.)