



# HHS Public Access

Author manuscript

*J Exp Psychol Gen.* Author manuscript; available in PMC 2020 October 01.

Published in final edited form as:

*J Exp Psychol Gen.* 2019 October ; 148(10): 1814–1827. doi:10.1037/xge0000529.

## Probabilistic Learning of Emotion Categories

Rista C. Plate<sup>#</sup>, Adrienne Wood<sup>#</sup>, Kristina Woodard, Seth D. Pollak

University of Wisconsin-Madison

<sup>#</sup> These authors contributed equally to this work.

### Abstract

Although the configurations of facial muscles that humans perceive vary continuously, we often represent emotions as categories. This suggests that, as in other domains of categorical perception such as speech and color perception, humans become attuned to features of emotion cues that map onto meaningful thresholds for these signals given their environments. However, little is known about the learning processes underlying the representation of these salient social signals. In Experiment 1 we test the role of statistical distributions of facial cues in the maintenance of an emotion category in both children (6-8-years-old) and adults (18-22-years-old). Children and adults learned the boundary between neutral and angry when provided with explicit feedback (supervised learning). However, after we exposed participants to different statistical distributions of facial cues, they rapidly shifted their category boundaries for each emotion during a testing phase. In Experiments 2 and 3, we replicated this finding and also tested the extent to which learners are able to track statistical distributions for multiple actors. Not only did participants form actor-specific categories, but the distributions of facial cues also influenced participants' trait judgments about the actors. Taken together, these data are consistent with the view that the way humans construe emotion (in this case, anger) is not only flexible, but reflects complex learning about the distributions of the myriad cues individuals experience in their social environments.

### Keywords

emotion categorization; supervised learning; unsupervised learning

---

Faces offer rich, salient sources of information for guiding judgments and behaviors. Human faces contain over 40 muscles that contract and relax in patterns, producing configurations that individuals use to infer the mental states of others. In principle, a human face can generate approximately 16 million muscular combinations. Because of this staggering amount of information, individuals must learn to attend to and ignore an extraordinary amount of information from their environments in order to successfully and rapidly

---

Correspondence concerning this article should be addressed to Seth D. Pollak, Department of Psychology, University of Wisconsin-Madison, 1202 West Johnson Street, Madison, WI 53706. seth.pollak@wisc.edu.

All authors developed the study concept and contributed to study design. R.C.P. and A.W. collected the data and performed data analysis and interpretation. R.C.P. and A.W. drafted the manuscript. K.W. contributed to the manuscript for Experiments 2 and 3. S.D.P. provided critical revisions. All authors approved the final version of the manuscript for submission.

The computerized tasks, stimuli, data, analysis scripts and output (in R) are available online (<https://osf.io/ycb3q/>). Ideas and data from Study 1 were presented at Society for Personality and Social Psychology (2018, poster), Society for Affective Science (2018, flash talk). Studies 2 and 3 were presented at the meeting of the Cognitive Science Society (2018, poster).

understand the communicative signals in emotion displays. The classical view in human development holds that initial perceptual categories of facial expression are evolutionarily preserved (Ekman, 1992; Grossmann, 2015; Shariff & Tracy, 2011). The contrasting hypothesis is that humans can detect and track variations in the distribution of signals in environments, and use this probabilistic information to construct and flexibly update facial expression categories based upon their sensory and social experiences (Barrett, 2017; Clore & Ortony, 2013; Pollak & Kistler, 2002). Yet, little data exists about how children could come to synthesize the vast range of facial movements in a coherent way that represents functionally meaningful patterns from their environments.

The current experiments test whether probabilistic information in environments influences how individuals make judgments about other people's emotions. In addition, we examine whether learning exerts differential levels of influence earlier versus later in development by contrasting the performance of children and adults. Plasticity might be expected in children, but we reasoned that if such learning effects exist in mature individuals, it would suggest that emotion categories remain fluid.

From infancy, faces capture human attention, and facial configurations are often represented as reflecting emotion categories (Cong et al., 2018; Diamond & Carey, 1986; Frank, Vul, & Johnson, 2009; Pollak, Messner, Kistler, & Cohn, 2009; Pourtois, Schettino, & Vuilleumier, 2012; Russell & Bullock, 1986; Susskind et al., 2007). These categories allow observers to respond to faces quickly by ignoring variability in facial movements to make general judgments about emotion signaling (Campanella, Quinet, Bruyer, Crommelinck, & Guerit, 2002; Etcoff & Magee, 1992). For example, observers will not perceive a gradual transition from low to high muscle activation as a linear, continuous change, but will instead perceive a qualitative shift in the facial configuration at the point where a person appears to have become "angry" (Calder, Young, Perrett, Etcoff, & Rowland, 1996; Campanella et al., 2002; Pollak & Kistler, 2002; Wood, Lupyan, Sherrin, & Niedenthal, 2016). The location of this shift is known as the *category boundary*, and it has important implications for how the observer will respond to another person's emotion signaling.

Despite ample behavioral (Etcoff & Magee, 1992) and neural (Campanella et al., 2002) evidence that emotions can be represented categorically, it is not well understood how these categories are acquired, whether these categories are stable, and whether these categories reflect the statistical distributions of the emotion cues that individuals observe. The amount of emotion-related information to be attended to and ignored is so vast that rudimentary categories may be innately preserved in neural architecture. However, recent research on domain-general categorization suggests that category representations emerge "ad hoc," and shift to meet the demands of the environment (Casasanto & Lupyan, 2011; Levari et al., 2018). Emotion categories may need to be malleable to allow individuals to adjust flexibly to variations in cultural and situational norms (Aviezer, Trope, & Todorov, 2012; Marchi & Newen, 2015; Niedenthal, Rychlowska, & Wood, 2017). While all emotions may require such flexibility, the focus of the present investigation is on anger. Categorical representations of facial anger may be particularly dependent on the social environment given the potential costs of failing to detect threat signals (Pollak & Kistler, 2002).

To date, the role of learning during repeated exposure to varying intensities of facial emotions has not been examined. Additionally, it is unknown whether children and adults would differ in emotion category flexibility to such exposure. On the one hand, children are characterized by flexibility in their learning, while adults tend to rely more heavily on top-down processes (e.g., Gopnik et al., 2017). This reasoning might suggest that children would be more likely to shift category boundaries in a brief learning episode. On the other hand, adults have greater experience with interpreting and responding to facial configurations across multiple contexts. This expertise may allow them to be more successful at integrating contextual cues that they should shift their category boundaries. Therefore, it is unknown whether statistical distributions of emotion cues can influence category boundaries, and, if so, whether such flexibility operates robustly in both childhood and adulthood.

## How Might People Update Emotion Categories?

Research on categorization in children and adults—and in machine learning applications—identifies two overarching types of learning, namely *supervised* and *unsupervised learning* (Love, 2002). One way that individuals might learn about emotion categories and category boundaries is through explicit instruction. Though not immediately obvious, some societies do explicitly teach children how to categorize emotions. For example, many North American preschools display posters depicting what different “feelings” ought to look like. In other forms of explicit teaching, adults may label emotions for children during nonverbal expressions (“Look, Johnny is crying, you made him feel sad”; Ahn, 2005; Gordon, 1991; Pollak & Thoits, 1989). We can think of this as *supervised learning*, as it involves incorporating feedback about whether the observer’s initial interpretation of a facial display was correct. Supervised learning likely plays a role in emotion understanding across development. As one example, the extent and circumstances under which children experience shame versus guilt are tied to overt aspects of parenting practices and socialization (Eisenberg, 2000).

However, much of human learning is thought to be *unsupervised* (Fisher, Pazzani, & Langley, 2014). Whereas supervised learning relies on information being directly provided to the learner, *unsupervised learning* happens in the absence of explicit information or feedback. During unsupervised learning, the learner extracts statistical distributions from their environment to acquire meaningful categories. Attention to statistical distributions is essential in domains in which it would be infeasible to acquire all needed information through explicit, supervised learning. For example, unsupervised learning aids in aspects of language acquisition such as phonemic discrimination and word segmentation (Maye, Werker, & Gerken, 2002; Saffran, Newport, & Aslin, 1996), where children could not be provided with sufficient explicit labeling of relevant distinctions between stimuli.

The relative contribution of supervised and unsupervised learning has been examined for artificial category formation of novel stimuli in children and adults (Kalish, Rogers, Lang, & Zhu, 2011; Kalish, Zhu, & Rogers, 2015). For example, in one study, 4-8-year-old children updated their category boundaries via unsupervised learning even when their original category boundaries were established through supervised learning (Kalish, Zhu, & Rogers, 2015). In a similar experiment, adults changed their previously-formed category boundaries

based on exposure to statistical distributions that differed from the original, labeled (i.e., supervised) distribution (Kalish, Rogers, Lang, & Zhu, 2011). However, these previous studies, by design, created completely novel stimuli, so that participants do not enter the experiment with a priori knowledge or expectations about how to categorize each exemplar. Scientists understand less about how these general learning mechanisms might operate on purportedly privileged biological stimuli, such as human facial configurations that convey emotion. While there is evidence to suggest that unsupervised exposure to facial expressions sharpens already-existing category representations of emotions (Huelle, Sack, Broer, Komlewa, & Anders, 2014), it is unclear if unsupervised learning can also shift those category representations within a feature space.

## Present Experiments

In Experiment 1, we tested whether probabilistic information influences an individual's representation of emotions. We predicted that category boundaries between emotions would be malleable and would reflect the distribution of facial configurations encountered. We expect that statistical learning plays a role in the continual updating of category boundaries for all facial expressions. In the present work we test this idea using anger, because prior evidence suggests long-term social environments contribute to individual differences in category boundaries for anger (Pollak & Kistler, 2002). If participants are able to use distributional information of perceptual cues to alter their anger category boundary, even in a brief timeframe, this would suggest that the functional boundaries between facial configurations are flexible, and that this category can update according to social contexts. Testing both children and adults, as has been done in research on supervised and statistical learning of novel categories (Kalish, Rogers, Lang, & Zhu, 2011; Kalish, Zhu, & Rogers, 2015), allowed us to test for developmental differences in emotion category malleability. If both children and adults update their categories based on extant cues, then we will have evidence that emotion representations remain flexible into maturity. Experiments 2 and 3 extend Experiment 1 by examining the robustness of the learning mechanisms in adults, and begin to assess whether shifts in these learned categories have functional significance in terms of participants' interpretation of facial signals.

## Experiment 1

For all experiments, we report how we determined our sample size, data exclusions, all manipulations, and all measures. The experiment, dataset, R code and stimuli files for all experiments are available online (<https://osf.io/ycb3q>).

## Method

**Participants.**—Ninety-one children (44 female, 47 male; age range = 6-8 years,  $M_{\text{age}} = 7.52$  years,  $SD_{\text{age}} = 0.92$  years) and 105 adults (75 female, 23 male, 7 unreported gender; age range = 18-22 years,  $M_{\text{age}} = 18.61$ ,  $SD_{\text{age}} = 0.95$ ) participated in this experiment. Children ages 6-8-years-old were chosen because they have shown sensitivity to supervised and unsupervised learning in previous research using novel stimuli (Kalish, Zhu, & Rogers, 2015); additionally we aimed for 30 participants per condition per age group (three between-subject conditions), which exceeded sample sizes of this previous research. Two additional

children were excluded due to a program error. Children were recruited from the community in a large Midwestern city (9% African American, 3% Asian American, 1% Hispanic, 5% Multiracial, 79% White, 2% did not report race). Adults were undergraduates at a large university in the same city who participated for course credit. Adult participants and parents of child participants gave informed consent; children gave verbal assent. Parents received \$20 for their time and children chose a prize for their participation. The Institutional Review Board approved the research.

**Face stimuli.**—Stimuli included images of facial emotions for one female model (Model #10) selected from the MacArthur Network Face Stimuli Set (Tottenham et al., 2009). The selected model’s facial expressions were morphed in increments of 5% from a 100% neutral expression (i.e., 0% angry) to a 100% angry expression of emotion, creating 21 equally-spaced images (stimuli from Gao & Maurer, 2009; Figure 1). Stimuli were presented with PsychoPy (v1.83.04).

**Procedure.**—Participants completed the task with an experimenter in a laboratory testing room. Instructions were presented on the screen and, for children, read by the experimenter. The experiment included three phases. The goal of the introductory phrase was to expose participants to the model and procedure, and to allow participants to practice the categorization task. The goal of the supervised learning phase was to train participants to a common category boundary. The goal of the testing phase was to assess whether participants shift the category boundary acquired during the supervised learning phase based on the statistical distribution of stimuli.

**Introductory phase:** Participants saw an image of the model (“Jane”) and were told, “Just like everyone, sometimes Jane feels upset and sometimes Jane feels calm. Today we need your help figuring out if Jane is upset or calm.” Participants were then presented with an image of the model and a computerized slider that allowed them to see all morphed facial expressions across the entire range from neutral to angry. Of note, the emotion labels, “angry” or “neutral” were never provided to participants in this task.

Next, participants saw an image of a red-colored room containing boxing equipment and a 100% angry morphed image of Jane and were told, “When Jane is feeling upset she likes to go to the red room and practice boxing.” Participants then saw an image of a blue-colored room containing a chair and books and a 0% angry (100% neutral) morphed image of Jane, with the instructions, “When Jane is feeling calm she likes to go to the blue room and read a book. We need you to help us figure out what room Jane should go to. If you think Jane is feeling upset, click the red button (left arrow) to put her in the red room. If you think Jane is feeling calm, click the blue button (right arrow) to put her in the blue room.” Adults used the arrow keys to respond; children used the arrow keys with red/blue stickers corresponding to the location of the room. The side of the screen where each room appeared was counterbalanced between participants.

Finally, to ensure participants understood the task, they completed six practice trials with labeled faces (e.g., “Jane is feeling upset. Which room should she go to?”), and the response options (i.e., red room and blue room) appeared in the bottom right and left corners. If the

participant made an error, they received feedback (“Incorrect! Please try again”) and the trial was repeated until the participant responded correctly (when the participant responded correctly, they received the feedback, “Correct!”). The morphs presented during the introductory phase that were 0%, 10%, 20% angry were labeled as “calm,” and the morphs that were 80%, 90%, and 100% angry were labeled as “upset.” All participants saw the same morphs, but the order was randomized between participants.

**Supervised learning phase.:** All participants then completed the same 24 trials with “correct” / “incorrect” feedback in randomized order. Stimuli consisted of two repetitions of morphs ranging from 20% angry to 80% angry in 5% increments. The 50% morph was omitted in order to create a category boundary at the midpoint (Figure 1). Correct responses were defined for stimuli less than 50% angry as “blue room” and stimuli greater than 50% angry as “red room.” Note that in this phase and the testing phase, the words “calm” and “upset” did not appear.

**Testing phase.:** The testing phase directly followed the supervised phase and included 72 trials. Participants were randomly assigned to one of three between-subjects conditions (Figure 1). In the unshifted condition, participants ( $n = 31$  children, 35 adults) saw six repetitions of each morph ranging from 20% angry to 80% angry (50% morph omitted to create category boundary). In the anger shifted condition ( $n = 30$  children, 37 adults), participants saw six repetitions of morphs ranging from 40% angry to 100% angry (70% morph omitted to create category boundary). In the neutral shifted condition ( $n = 30$  children, 33 adults), participants saw six repetitions of morphs ranging from 0% angry to 60% angry (30% morph omitted to create category boundary). Participants did not receive feedback during this phase. Stimuli appeared in random order, however each individual morph was seen once before repeating.

## Results

**Analytic plan.**—In the results that follow, we first report analyses of the supervised learning phase. A logistic generalized linear mixed-effect modeling approach allowed us to examine the relative steepness of children’s and adults’ category boundaries, which indicated how precise and categorical their representations of “calm” and “upset” became during the brief supervised learning phase. To ensure that children and adults were able to do the task equally well, we also looked at overall accuracy. Next, we used the same logistic regression approach to analyze participant judgments in the testing phase. In particular, we examined whether participants’ category boundaries shifted as a function of experimental condition (unshifted, anger shifted, and neutral shifted), and whether children and adults differed in their sensitivity to the shifted distributions. Results from the logistic regression are reported with odds ratios (OR), which indicate by how much the odds of an “upset” response increase ( $OR > 1$ ) or decrease ( $OR < 1$ ) with a unit increase in the predictor variable (e.g.,  $OR = 2$  means the odds of an “upset” response double). All analyses were conducted in the R environment (R Development Core Team, 2008) with the lme4 package, version 1.1-15 (Bates, Mächler, Bolker, & Walker, 2015).

### Supervised learning phase.

**Individuals established a category boundary via supervised learning.:** First, we assessed whether children and adults were able to learn the category boundary that we taught during the supervised phase. Since the responses were dichotomous, we used logistic generalized linear mixed-effects models, regressing response (“calm” = 0, “upset” = 1) on the interaction between centered Age Group and mean-centered Percent Anger (ranging from 20 for mostly neutral facial expression to 80 for mostly angry facial expression, at 5-unit intervals), with a by-participant random slope for Percent Anger. The main effect of Percent Anger was significant, with the odds of an image being categorized as “upset” increasing by a factor of 6.77 with every 10% increase in anger intensity,  $b = 1.91$ ,  $\chi^2(1) = 640.53$ ,  $p < .001$ .

**Children and adults did not differ in the formation of categories via supervised learning.:** The main effect of Age Group was not significant, indicating that age did not alter the overall probability of someone categorizing a face as “upset”,  $b = .10$ ,  $\chi^2(1) = 0.63$ ,  $p = .427$ , odds ratio ( $OR$ ) = 1.11. The interaction between Age Group and Percent Anger trended towards statistical significance: Shifting 10% in percent anger present in a face increased the odds of it being categorized as “upset” by a factor of 1.25 *more* for adults compared to children,  $b = 0.22$ ,  $\chi^2(1) = 3.32$ ,  $p = .069$ . In other words, the slope of the category boundary on supervised learning trials was marginally steeper (more categorical) for adults compared to children (Figure 2). This marginal difference in category boundary steepness notwithstanding, children and adults seemed to learn the explicitly-taught “calm” and “upset” categories and performed comparably.

We next determined whether children and adults differed in their overall performance in the supervised learning phase. We regressed participants’ proportion of correct responses on Age Group (coded as children =  $-.5$ , adults =  $.5$ ). Children (accuracy  $M = .89$ ,  $SD = .10$ ) and adults (accuracy  $M = .90$ ,  $SD = .08$ ) did not differ significantly in their overall proportion of correct responses,  $b = 0.01$ ,  $F(1, 390) = 2.37$ ,  $p = 0.125$ , *partial*  $\eta^2 = 0.006$ .

**Testing phase.**—Next, we examined the effect of shifting the statistical distribution of facial images on participants’ categorization in the testing phase. We used the same generalized linear mixed-effect model as with the supervised phase data, with the addition of two dummy variables to compare responses in the neutral shifted and anger shifted conditions to the unshifted condition. The full model regressed participant responses on a three-way interaction between Percent Anger\* Age Groups\* Experimental Condition dummy variables, plus all lower-order fixed effects and a by-participant random slope for Percent Anger.

**Exposure to shifted distributions altered participants’ category boundaries.:** Our primary question of interest was whether participants would override explicitly learned emotion category boundaries based upon mere passive exposure to a new statistical distribution of facial input. Participants did shift their emotion categories during the testing phase in response to the intensities of emotions they encountered. The main effect of Experimental Condition was significant,  $\chi^2(2) = 453.66$ ,  $p < .001$  (Figure 3). Dummy coded parameters indicated that the anger-response threshold was significantly earlier (i.e.,

occurred at a lower percentage morph) in the morph continuum for participants in the neutral shifted condition,  $b = 2.07$ ,  $z = 10.32$ ,  $p < .001$ ,  $OR = 7.93$  and significantly later (i.e., occurred at a higher percentage morph) in the morph continuum for the anger shifted participants,  $b = -2.44$ ,  $z = -12.36$ ,  $p < .001$ ,  $OR = 0.09$ , in comparison with the participants in the unshifted condition. These estimates indicate that both children and adults adapted their categories about which faces constituted “anger” in accord with the facial configurations encountered in the shifted experimental conditions. Furthermore, the Experimental Condition\*Age Group interaction was not significant,  $\chi^2(2) = 3.09$ ,  $p = .214$ . In other words, children and adults shifted their boundaries to the same extent.

Finally, the Experimental Condition\*Percent Anger interaction was not significant, indicating that participants’ category boundaries were no less steep if they were in a shifted condition,  $\chi^2(2) = 1.09$ ,  $p = .580$ . In other words, to the extent that the sharpness of a category boundary indicates the precision of the participant’s category representation, participants’ representations were equally precise even when their category boundaries moved substantially between the supervised learning and testing phases.

**Adults more readily adapted to the range of emotion expression intensity.:** Children and adults did not differ in their overall use of “calm” and “upset” responses in the unshifted condition (controlling for all other variables,  $b = -0.01$ ,  $\chi^2(1) = .001$ ,  $p = .970$ ). However, the interaction between Age Group and Percent Anger was significant, with the effect of 10% increases in morphed anger being greater by a factor of 2.816 for adults compared to children,  $b = 1.04$ ,  $\chi^2(1) = 27.52$ ,  $p < .001$ . This effect can be observed by the slightly steeper curves for adults compared to children in Figure 3. Thus, adults have more precise categories across experimental conditions compared with children. The 3-way interaction between Age Group, Percent Anger, and Experimental Condition was not significant,  $\chi^2(2) = 4.21$ ,  $p = .122$ .

## Discussion

After just a brief exposure to a new statistical distributions of facial emotions, children and adults changed their thresholds for what they considered to be someone feeling “upset.” Yet despite updating their category boundaries, participants in the neutral and anger shifted conditions showed comparable categorical precision to the perceptual judgments of participants in the unshifted condition. In other words, the new categorical representations were as distinct as those for which participants could continue to rely on the explicitly-taught category boundary. While these data do not necessarily address how children initially acquire emotion categories, the data provide evidence that these perceptual categories are flexible. This evidence supports the view that statistical distributions of facial configurations influence people’s representations of emotion categories. And, in turn, Experiment 1 suggests that categories for, and interpretations of, anger are malleable and can be adjusted according to the patterns in the environment.

Adults established (during the supervised and testing phases) somewhat steeper category boundaries compared to children, likely reflecting that adults had more precise category representations of facial anger overall. It may be that categorization becomes more robust



and efficient with experience, particularly as adults prioritize facial cues, whereas children are learning how to interpret faces and divide their attention equally between faces and other contextual information (Leitzke & Pollak, 2016). However, children's performance was nevertheless characterized by the high degree of flexibility in learning that has been observed across cognitive domains (e.g., Gopnik et al., 2017).

## Experiment 2

In Experiment 1, participants were able to rapidly update their category boundary between neutral and angry for a single model. We suggest such emotion perception flexibility allows perceivers to adjust to the expressive tendencies of the people in their current environmental context. But some social environments contain people with diverging expressive styles—for instance, one friend may be highly expressive, with unambiguous anger displays, and another friend may have expressions of anger that are much more subtle. When a perceiver encounters multiple people with distinct ranges of expression, does their category boundary reflect an average of all the individuals' expressive ranges? Or do perceivers track the probabilistic distributions of individual expressers and establish a unique category boundary for each social actor? To address this question, we trained adult participants on a common category boundary for three different actors (within-subject), and then exposed participants to a testing phase, where each actor's expressive distribution was shifted to a different distribution of intensity of facial displays. In addition to assessing whether perceivers are sensitive to, and can track, intra-individual variation in emotion expression, we asked whether perceivers make functional use of these differences in forming judgments of the expressers. The influence of distributional information on perceivers' explicit judgments about expressers has implications for unpacking the relation between perceptual cues and social behavior. Participants were constrained to adults for Experiments 2 and 3 for two reasons: 1) there were no developmental differences in the extent to which children and adults shifted boundaries in Experiment 1 and 2) moving to a within-subjects design required a greater number of trials, which we thought would be demanding for children.

## Method

**Participants.**—Participants were 55 adults (34 female, 18 male, 3 unreported gender; age range = 18 - 21 years,  $M_{\text{age}} = 18.77$ ,  $SD_{\text{age}} = 0.70$ ). We aimed for 40 participants because 30 participants per cell was sufficient in Experiment 1. Because of over-scheduling, we collected data on 55 participants. One participant was excluded for inattention (as defined by only using one response option throughout the task). Adults were undergraduates at the same university as Experiment 1 and participated for course credit (7% African American, 16% Asian American, 4% Hispanic, 4% Multiracial, 67% White, 4% unreported). Participants gave informed consent. The Institutional Review Board approved the research.

**Face stimuli.**—Stimuli included images of facial expressions of emotion for two male models (Model #24 and Model #42) and one female model (Model #10; same as in Experiment 1) selected from the MacArthur Network Face Stimuli Set (Tottenham et al., 2009). As in Experiment 1, all models' facial expressions were morphed in increments of 5% from a 100% neutral expression (i.e., 0% angry) to a 100% angry expression of emotion,

creating 21 equally-spaced images (Gao & Maurer, 2009). Stimuli were presented with PsychoPy (v1.83.04).

**Procedure.**—The procedure was identical to Experiment 1 except for the following minor changes to the instructions, the number of trials in the introductory, supervised, and testing phases, and the addition of actor trait judgments at the end of the task. In the introductory phase, participants were introduced to three models (“Jane,” “Tom,” and “Brian”). All other instructions were the same, with “Jane, Tom and Brian” replacing “Jane.” Eighteen practice trials were completed (six per actor). In the supervised phase, the number of trials was adjusted to account for the new actors. Stimuli consisted of three repetitions per actor, using the same morphing increments as before, for a total of 108 trials. In the testing phase, the unshifted, neutral shifted, and anger shifted conditions all occurred within-participants. Participants saw 72 trials of each model for a total of 216 trials. As the number of trials per model was the same as in Experiment 1, we used the same distributions of morphs. The actor assigned to each distribution (unshifted, anger shifted, or neutral shifted) was randomized across participants.

At the end of the task, participants completed an assessment of trait judgments about the actors whose faces they viewed during the experiment. For each actor, they were shown the actor’s 100% happy expression and were asked to, “Rate your impression of Jane / Brian / Tom: How likeable / approachable / irritable / angry / friendly is this person?”. Each actor was judged on each trait, for a total of 15 questions; the order of the actors was randomized between participants. Because of a programming error, trait judgment data was lost on five participants.

## Results

**Analytic plan.**—Here we report results from the testing phase. Results from the supervised phase are available online. We followed the same data processing and analysis procedure as Experiment 1, making changes only as demanded by the modified design of Experiment 2. Specifically, there was no Age variable (because we only included adults in this study) and the random effects structure for the testing phase model was adjusted to reflect the within-subject nature of the Actor Condition variable (dummy variable that coded for the neutral shifted, unshifted, or anger shifted actor). We then made a cross-study analysis of the current data and the adults’ data from Experiment 1 to compare participant efficiency in learning the shifted category boundaries when participants needed to track the distributions of multiple people’s facial expressions. Last, we report analyses of the actor trait judgments.

**Testing phase.**—We regressed participants’ responses (0 = “calm”, 1 = “upset”) on the interaction between Actor Condition and Percent Anger. The maximal model with by-participant random slopes for the interaction term and lower order terms failed to converge, so we removed the lower order random slopes (Brauer & Curtin, 2017). The main effect of Actor Condition was significant,  $\chi^2(2) = 251.53, p < .001$ , indicating that participants established distinct category boundaries for each of the three actors (see solid lines in Figure 5). The dummy coded parameters indicated that the anger-response threshold was significantly earlier (i.e., at a lower anger percentage) in the morph continuum when the

actors were neutral shifted,  $b = 0.82$ ,  $z = 9.62$ ,  $p < .001$ ,  $OR = 2.27$ , and significantly later (i.e., at a higher anger percentage) in the morph continuum when the actors were anger shifted,  $b = -0.62$ ,  $z = -7.45$ ,  $p < .001$ ,  $OR = 0.54$ , compared to when the actors were unshifted. The intercept was not significantly different from 0, indicating that participants' category boundaries for the unshifted actor remained at the 50% anger point (the boundary trained during the supervised phase),  $b = -0.06$ ,  $\chi^2(1) = 0.25$ ,  $p = .619$ ,  $OR = 0.94$ .

The interaction between Percent Anger and Actor Condition was also significant,  $\chi^2(2) = 7.40$ ,  $p = .025$ , indicating that the steepness of the category boundary slopes was not uniform across the three shifted Conditions. The dummy coded interaction terms suggest that the slope of the category boundary for the neutral shifted actors was steeper than the slope for the unshifted actors,  $b = 0.35$ ,  $z = 2.67$ ,  $p = .008$ ,  $OR = 1.42$ , while the slope for the anger shifted actors only trended towards being steeper than the unshifted actors slope,  $b = 0.18$ ,  $z = 1.65$ ,  $p = .10$ ,  $OR = 1.20$ .

**Adapting emotion representations for multiple expressers.**—The above analyses indicate that participants did not maintain a single emotion category for each of the different actors in the testing phase. Rather, participants were able to encode each actor's particular distribution of facial displays and formed distinct category boundaries for each actor.

It is possible that the multiple actor context caused participants to draw comparisons between the actors. For instance, perhaps viewing “Tom’s” distributions colored participants’ interpretations of “Jane’s” expressions. To test this possibility, we analyzed the adult data from Experiment 1 along with trials from Experiment 2 in which Jane was the target (since Jane was the only target in Experiment 1) across all possible shifts (neutral shifted, unshifted, anger shifted). We regressed participant responses on the Actor Condition\*Percent Anger interaction (as above), plus each of those variables’ interactions with Experiment (Experiment 1 =  $-.5$ , Experiment 2 =  $.5$ ). Next, we included by-participant random slopes for the Actor Condition\*Percent Anger interaction. The interaction between Experiment and Actor Condition was significant,  $\chi^2(2) = 45.98$ ,  $p < .001$ ,  $ORs = 0.31$  and  $9.41$ , demonstrating that participants did not shift their boundaries for Jane as far when they were also seeing Tom and Brian’s competing expression distributions (Figure 4). The interaction between Experiment and Percent Anger was also significant, such that participants had steeper (more categorical) category boundaries in Experiment 1 compared to Experiment 2,  $\chi^2(1) = 63.53$ ,  $p < .001$ ,  $OR = 0.41$ .

**The effect of shifted distributions on trait judgments of actors.**—Participants’ ratings of each of the three actors’ anger, irritability, likability, friendliness, and approachability were strongly correlated ( $r's > .64$ ), so we computed an average score of Overall Evaluation (higher scores indicate more positive evaluations of the target; models analyzing each of the five judgments separately are online). Participants were expected to rate the actors in the neutral shifted condition most positively, followed by the unshifted actors, then the anger shifted actors, since the latter had the most “intense” negative facial movements. This prediction translated into a linear contrast (neutral shifted =  $-.5$ , unshifted =  $0$ , anger shifted =  $.5$ ), and created an orthogonal quadratic contrast variable (neutral and anger shifted =  $-.33$ , unshifted =  $.66$ ). We then regressed the Overall Evaluation scores on

the two contrast variables. After initially fitting a linear mixed-effect model with random slopes for both contrast variables, we removed the orthogonal quadratic contrast's random slope when the model failed to converge (Brauer & Curtin, 2017). The linear variable trended towards significance,  $b = -0.26$ ,  $t(71.47) = -1.85$ ,  $p = .06$ ,  $OR = 0.77$  (the orthogonal contrast was not significant). In other words, we found suggestive—but not statistically significant—evidence that participants made more negative trait attributions about actors with more extreme anger expressions compared to the actors with more subtle expressive distributions during the task (even though the actors displayed the same, 100% happy expression, when trait ratings were obtained). Thus, participants not only updated their category boundaries in response to the shifted distributions of anger expressions, but may have also formed impressions of the actors' emotional traits based on those distributions.

## Discussion

Experiment 2 demonstrates that, when faced with multiple people differing in anger expressivity, perceivers formed individual-specific category boundaries. This effect indicates robustness of statistical learning, allowing perceivers to both track and integrate multiple distributions simultaneously. Additionally, a trend in the data suggests that perceivers formed trait judgments that varied based on the statistical distribution of each individual's expressivity. While the effect on actor trait judgments was small, if valid, it would indicate that perceptual categories have a functional implication in how individuals engage with social partners.

One notable limitation for Experiment 2 (and Experiment 1) is the presence of the supervised phase that occurred before participants engaged in the testing phase. In the supervised phase, participants were trained to a midpoint category boundary. Perhaps, as a result of this training, participants acquired an assumption that they should respond “red room” for 50% of the trials and respond “blue room” for 50% of the trials for each actor. Participants acting in accordance with this assumption would demonstrate the same pattern of results observed in Experiment 2. To address this potential confound, we attempted to replicate Experiment 2, but removed the supervised learning phase. Therefore, the testing phase of Experiment 3 fully constitutes unsupervised learning in that participants are exposed to the stimuli without any prior feedback regarding categorization.

## Experiment 3

### Method

**Participants.**—Participants were 40 adults (21 female, 19 male; age range = 18 - 21 years,  $M_{\text{age}} = 18.88$ ,  $SD_{\text{age}} = 0.88$ ). We aimed for 40 participants as in Experiment 2. Adults were undergraduates at the same university as in Experiments 1 and 2, who participated for course credit (5% African American, 10% Asian American, 8% Hispanic, 3% Multiracial, 75% White). Participants gave informed consent. The Institutional Review Board approved the research.

**Face stimuli.**—Stimuli used were identical to those in Experiment 2.

**Procedure.**—The procedure was identical to Experiment 2 except that the introductory phase was shortened and the supervised learning phase was omitted. The introductory phase consisted of six trials in which each actor’s 0% angry and 100% angry morphs were presented. Participants also completed the survey of actor judgments at the end of task, though because of program errors, only 37 of the participants had complete survey data.

## Results

**Analytic plan.**—We followed the same data processing and analysis procedure as Experiment 2.

**Testing phase.**—The model was identical to Experiment 2. The intercept term was significantly different from 0, suggesting that in the absence of the supervised phase, participants’ category boundary for the unshifted actors was not at 50% anger,  $b = 1.88$ ,  $\chi^2(1) = 68.85$ ,  $p < .001$ ,  $OR = 6.55$ . Participants who have never received feedback on their responses tended to categorize an expression as “upset” at a lower anger percentage (Figure 5). As in the previous experiments, the effect of Actor Condition was significant,  $\chi^2(2) = 32.10$ ,  $p < .001$ . The dummy coded parameters indicated that the category boundaries for the neutral shifted actors were at a lower anger percentage than for the unshifted actors,  $b = 0.52$ ,  $z = 3.92$ ,  $p < .001$ ,  $OR = 1.68$ . Participants’ boundaries for anger shifted actors trended toward being at a higher anger percentage than for the unshifted actors,  $b = -0.21$ ,  $z = -1.69$ ,  $p = .09$ ,  $OR = 0.81$ . Thus we replicated the overall effect from Experiment 2, even after removing the potential confound of the supervised phase.

**The effect of initial supervised learning: Comparing Experiments 2 and 3.**—We combined the testing phase data from Experiments 2 and 3 (coded as  $-.5$  and  $.5$ , respectively) to examine whether the initial supervised phase (which was only present in Experiment 2, and trained participants to an initial boundary location of 50% for all actors) changed participants’ response patterns in the subsequent testing phase. The model structure was identical to that of the model comparing Experiments 1 and 2.

The main effect of Experiment was significant, such that participants in Experiment 3 made significantly more “upset” categorizations than “calm”,  $b = 1.95$ ,  $\chi^2(1) = 63.42$ ,  $p < .001$ ,  $OR = 7.03$ . This is illustrated in Figure 5 by the overall leftward shift of the Experiment 3 lines. The interaction between Actor Condition and Experiment was also significant,  $\chi^2(2) = 16.14$ ,  $p < .001$ . The dummy coded interaction terms indicate that the Experiment 2 participants shifted their responses on anger shifted trials significantly further from their unshifted trial responses, when compared to Experiment 3 participants,  $b = 0.38$ ,  $z = 2.56$ ,  $p = .01$ ,  $OR = 1.46$ . The moderating effect of Experiment on the distance between the neutral shifted boundary and the unshifted boundary was not significant,  $b = -.23$ ,  $z = -1.55$ ,  $p = .12$ ,  $OR = 0.79$ .

**The effect of shifted distributions on actor judgments.**—We took an identical approach to analyzing participants’ ratings of the three actors’ anger, irritability, likability, friendliness, and approachability, which we averaged into an Overall Evaluation score. The linear contrast for Actor Condition was significant, such that participants’ evaluations of the

three actors became increasingly negative as their expressive distributions shifted towards the more angry end of the continuum,  $b = -0.39$ ,  $t(48.63) = -2.15$ ,  $p = .04$ ,  $OR = 0.68$ .

## Discussion

Experiment 3 was identical to Experiment 2 but lacked the initial supervised phase. We did this to determine whether participants' seeming sensitivity to each actor's distribution merely reflected a tendency to categorize half of each actor's expressions as "upset", a task demand that the supervised phase may have inadvertently enforced. In comparing Experiment 2 and Experiment 3, it appears that the supervised phase in Experiment 2 served to: 1) bring all participants away from their default anger detection threshold to a common 50% anger boundary, and 2) make participants somewhat more sensitive to the distinct expression distributions for each actor. The latter effect may be due Experiment 2's task demand to respond "calm" and "upset" equally for each actor. However the effect of shifting each actor's distribution in Experiment 3 did not go away when the supervised phase was removed. This suggests people can and do encode the unique expressive ranges of individuals and categorize their expressions accordingly, regardless of task demands.

## General Discussion

People represent facial configurations as members of categories (de Gelder, Teunisse, & Benson, 1997; Cong et al., 2018). But it was unknown whether and how people flexibly change them to reflect their social contexts, and whether this tendency changes with age. The present data provide support for the view that the statistical distribution of observed expressions operates on people's representations of emotions, highlighting the malleability of emotion categories across ages, and (though not directly tested here) suggesting a potential mechanism through which emotion categories might be formed. Further, the specificity and efficiency with which the boundaries shifted during exposure to the distribution of faces—even with multiple actors—suggests that these learning processes are both robust and flexible.

### Exploring a Mechanism for Adjusting to the Expressive Styles of Others

The current work suggests that unsupervised statistical learning is one way in which perceivers quickly adjust their interpretation of facial emotions according to individual differences in expressivity. People often encounter shifts in how emotions are expressed, both in the short-term (e.g., with a particularly expressive social partner) and in the long-term (e.g., when visiting a new culture). Substantial variability in expressivity exists (Friedman, Prince, Riggio, & DiMatteo, 1980), some resulting from personality (Friedman, DiMatteo, & Taranta, 1980) and gender (Kring & Gordon, 1998), such that some individuals are more expressive facially than others. To successfully engage with others, observers must be able to track and adjust to such individual differences. At the same time, and consistent with the data reported here, patterns of individual variation are not independent, as perceivers also integrate exemplars with reference to each other.

On a larger scale, entire cultures vary in production of facial expressions (Niedenthal, Rychlowska, & Wood, 2017). Cross-cultural differences in expressivity are thought to be

cultural adaptations to social and ecological pressures, and interactions with people from other cultures partly rely on an observer's ability to quickly adjust to the expressive style of their new social partner (Girard & McDuff, 2017; Rychlowska et al., 2015; Wood, Rychlowska, & Niedenthal, 2016). The current work elucidates one way in which perceivers quickly adjust their interpretation of facial emotions, that is, depending upon statistical properties of their social partner's expressivity. To successfully engage with others, observers must be able to learn from and adjust to these variations. Future research can explore the possibility that this process of tracking the distributional properties of emotional displays contributes to children's initial acquisition of emotion representations and expressivity norms as they navigate the social world.

Finally, the current work suggests that emotion perception researchers should proceed with caution whenever using emotion categorization or detection tasks to measure individual differences they assume to be stable (e.g., Baron-Cohen et al., 2001). Here we found that people's emotion category representations are flexible and responsive to the distribution of cues in the social context.

### Limitations and Future Directions

In one sense, the laboratory paradigm is a very simplified version of emotion categorization. Yet, it still provides insights into real-world behavior. Rather than having participants simply label a face with emotion words (something people arguably rarely do in everyday life, but often are asked to do in the laboratory), participants predicted the actor's likeliest next behavior. No emotion labels were encountered in the supervised and testing phases of the experiment. We suggest the current task bears closer resemblance to what people do with the information they extract from the faces of people around them (Martin, Rychlowska, Wood, & Niedenthal, 2017). However, while we avoided labels that imply an internal state, such as "anger," and did not use labels as the forced-choice options, our use of the terms "calm" and "upset" during the instructions and practice trials may have influenced participants' representations of the stimuli (see Doyle & Lindquist, 2018; Lupyan, Rakison, & McClelland, 2007).

Future research can also consider how this learning and updating process generalizes across emotions. The current study explored the effects of supervised and unsupervised learning on anger; future research can confirm that similar processes operate across other emotion categories. Anger indicates a potential threat, and is an especially salient and attention-grabbing stimulus. It could be that adults and children are more attuned to the patterns of anger expressions than other expressions. Additionally, an interesting next step will be to examine how these learning processes operate when embedded in richer contexts that mimic the social world. For example, antecedent events and behavioral consequences that precede and follow a facial expression can shape or reinforce a perceiver's current and future interpretations of the expression (Barrett, Mesquita, & Gendron, 2011).

### Conclusion

The current experiment is the first known demonstration that exposure to a particular statistical distribution of facial emotion changed people's emotion category boundaries. That

emotion categories are malleable and responsive to environmental statistics raises new possibilities for understanding human emotion perception. Such flexibility also allows individuals to adjust emotion concepts across contexts and organize appropriate behavioral responses based on the available socio-cultural cues.

## Acknowledgements

This research was supported by the National Institute of Mental Health through grant R01MH61285 to S.D.P. and by a core grant to the Waisman Center from the National Institute of Child Health and Human Development (U54 HD090256). R.C.P. was supported by a National Science Foundation Graduate Research Fellowship (DGE-1256259) and the Richard L. and Jeanette A. Hoffman Wisconsin Distinguished Graduate Fellowship. A.W. was supported by an Emotion Research Training Grant (T32MH018931-24) from the National Institute of Mental Health. We thank the individuals who participated in these experiments, and the research assistants, particularly Teresa Turco, who helped with data collection. We also thank Chuck Kalish, Paula Niedenthal, and Jenny Saffran for their feedback on an earlier version of this manuscript.

## References

- Ahn HJ (2005). Teachers' discussions of emotion in child care centers. *Early Childhood Education Journal*, 32(4), 237–242.
- Aviezer H, Trope Y, & Todorov A (2012). Holistic Person Processing: Faces With Bodies Tell the Whole Story. *Journal of Personality and Social Psychology*, 103(1), 20–37. 10.1037/a0027411 [PubMed: 22352325]
- Baron-Cohen S, Wheelwright S, Hill J, Raste Y, & Plumb I (2001). The “Reading the Mind in the Eyes” Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42(2), 241–251. 10.1017/S0021963001006643
- Barrett LF (2017). Functionalism cannot save the classical view of emotion. *Social Cognitive and Affective Neuroscience*, 12(1), 34–36. 10.1093/scan/nsw156 [PubMed: 27798259]
- Barrett LF, Mesquita B, & Gendron M (2011). Context in emotion perception. *Current Directions in Psychological Science*, 20(5), 286–290. 10.1177/0963721411422522
- Bates D, Mächler M, Bolker B, & Walker S (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. 10.18637/jss.v067.i01
- Brauer M, & Curtin JJ (2017). Linear Mixed-Effects Models and the Analysis of Nonindependent Data: A Unified Framework to Analyze Categorical and Continuous Independent Variables that Vary Within-Subjects and/or Within-Items. *Psychological Methods*. 10.1037/met0000159
- Calder AJ, Young AW, Perrett DI, Etcoff NL, & Rowland D (1996). Categorical perception of morphed facial expressions. *Visual Cognition*, 3(2), 81–118. 10.1080/713756735
- Campanella S, Quinet P, Bruyer R, Crommelinck M, & Guerit J-M (2002). Categorical perception of happiness and fear facial expressions: an ERP study. *Journal of Cognitive Neuroscience*, 14(2), 210–227. 10.1162/089892902317236858 [PubMed: 11970787]
- Casasanto D, & Lupyan G (2011). Ad hoc cognition. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, No. 33).
- Clore GL, & Ortony A (2013). Psychological construction in the OCC model of emotion. *Emotion Review*, 5(4), 335–343. 10.1177/1754073913489751 [PubMed: 25431620]
- Cong Y-Q, Junge C, Aktar E, Raijmakers M, Franklin A, & Sauter D (2018). Pre-verbal infants perceive emotional facial expressions categorically. *Cognition and Emotion*, 0(0), 1–13. 10.1080/02699931.2018.1455640
- de Gelder B, Teunisse J-P, & Benson PJ (1997). Categorical Perception of Facial Expressions: Categories and their Internal Structure. *Cognition and Emotion*, 11(1), 1–23. 10.1080/026999397380005
- Diamond R, & Carey S (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General*, 115(2), 107–117. 10.1037/0096-3445.115.2.107 [PubMed: 2940312]

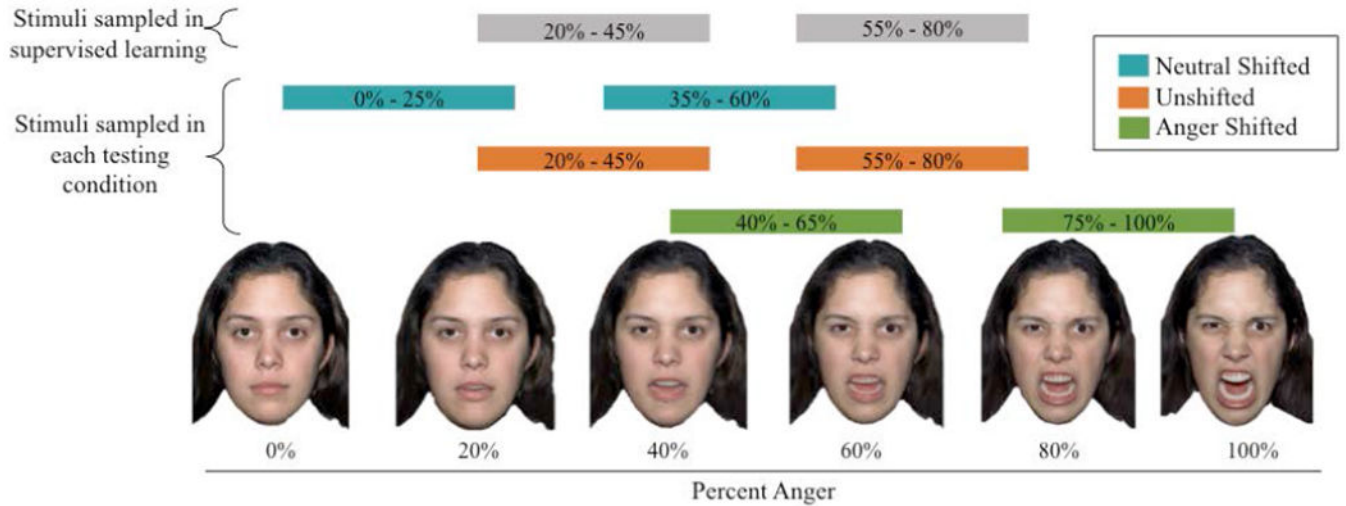


- Doyle CM, & Lindquist KA (2018). When a word is worth a thousand pictures: Language shapes perceptual memory for emotion. *Journal of Experimental Psychology: General*, 147(1), 62. [PubMed: 29309197]
- Ekman P (1992). An argument for basic emotions. *Cognition & emotion*, 6(3-4), 169–200. <https://doi-org/10.1080/02699939208411068>
- Eisenberg N (2000). Emotion, regulation, and moral development. *Annual review of psychology*, 57(1), 665–697. <https://doi-org/10.1146/annurev.psych.51.1.665>
- Etcoff NL, & Magee JJ (1992). Categorical perception of facial expressions. *Cognition*, 44(3), 227–240. 10.1016/0010-0277(92)90002-Y [PubMed: 1424493]
- Fisher DH, Pazzani MJ, & Langley P (2014). *Concept Formation: Knowledge and Experience in Unsupervised Learning*. San Mateo, CA: Morgan Kaufmann Publishers.
- Friedman HS, DiMatteo MR, & Taranta A (1980). A study of the relationship between individual differences in nonverbal expressiveness and factors of personality and social interaction. *Journal of Research in Personality*, 74(3), 351–364. 10.1016/0092-6566(80)90018-5
- Friedman HS, Prince LM, Riggio RE, & DiMatteo MR (1980). Understanding and assessing nonverbal expressiveness: The affective communication test. *Journal of Personality and Social Psychology*, 39(2), 333–351. 10.1037/0022-3514.39.2.333
- Frank MC, Vul E, & Johnson SP (2009). Development of infants' attention to faces during the first year. *Cognition*, 770(2), 160–170. 10.1016/j.cognition.2008.11.010
- Gao X, & Maurer D (2009). Influence of intensity on children's sensitivity to happy, sad, and fearful facial expressions. *Journal of experimental child psychology*, 702(4), 503–521. 10.1016/j.jecp.2008.11.002
- Girard JM, & McDuff D (2017). Historical heterogeneity predicts smiling: Evidence from large-scale observational analyses. In *Automatic Face & Gesture Recognition (FG 2017)*, 2017 12th IEEE International Conference on (pp. 719–726). 10.1109/FG.2017.135
- Gopnik A, O'Grady S, Lucas CG, Griffiths TL, Wente A, Bridgers S, & Dahl RE (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proceedings of the National Academy of Sciences*, 201700811 10.1073/pnas.1700811114
- Gordon SL (1991). 12 The socialization of children's emotions: emotional culture, competence, and exposure. *Children's understanding of emotion*, 319.
- Grossmann T (2015). The development of social brain functions in infancy. *Psychological Bulletin*, 141(6), 1266 10.1037/bul0000002 [PubMed: 25984728]
- Huelle JO, Sack B, Broer K, Komlewa I, & Anders S (2014). Unsupervised learning of facial emotion decoding skills. *Frontiers in Human Neuroscience*, 8 10.3389/fnhum.2014.00077 [PubMed: 24478674]
- Kalish CW, Rogers TT, Lang J, & Zhu X (2011). Can semi-supervised learning explain incorrect beliefs about categories?. *Cognition*, 120(1), 106–118. 10.1016/j.cognition.2011.03.002 [PubMed: 21474122]
- Kalish CW, Zhu X, & Rogers TT (2015). Drift in children's categories: when experienced distributions conflict with prior learning. *Developmental science*, 18(6), 940–956. <https://doi-org/10.1111/desc.12280> [PubMed: 25530185]
- Kring AM, & Gordon AH (1998). Sex differences in emotion: expression, experience, and physiology. *Journal of personality and social psychology*, 74(3), 686 10.1037/0022-3514.74.3.686
- Leitzke BT, & Pollak SD (2016). Developmental changes in the primacy of facial cues for emotion recognition. *Developmental psychology*, 52(4), 572 10.1037/a0040067 [PubMed: 26784383]
- Levari DE, Gilbert DT, Wilson TD, Sievers B, Amodio DM, & Wheatley T (2018). Prevalence-induced concept change in human judgment. *Science*, 360(6396), 1465–1467. 10.1126/science.aap8731 [PubMed: 29954981]
- Love BC (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review*, 9(4), 829–835. 10.3758/BF03196342 [PubMed: 12613690]
- Lupyan G, Rakison DH, & McClelland JL (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological science*, 18(12), 1077–1083. 10.1111/j.1467-9280.2007.02028.x [PubMed: 18031415]

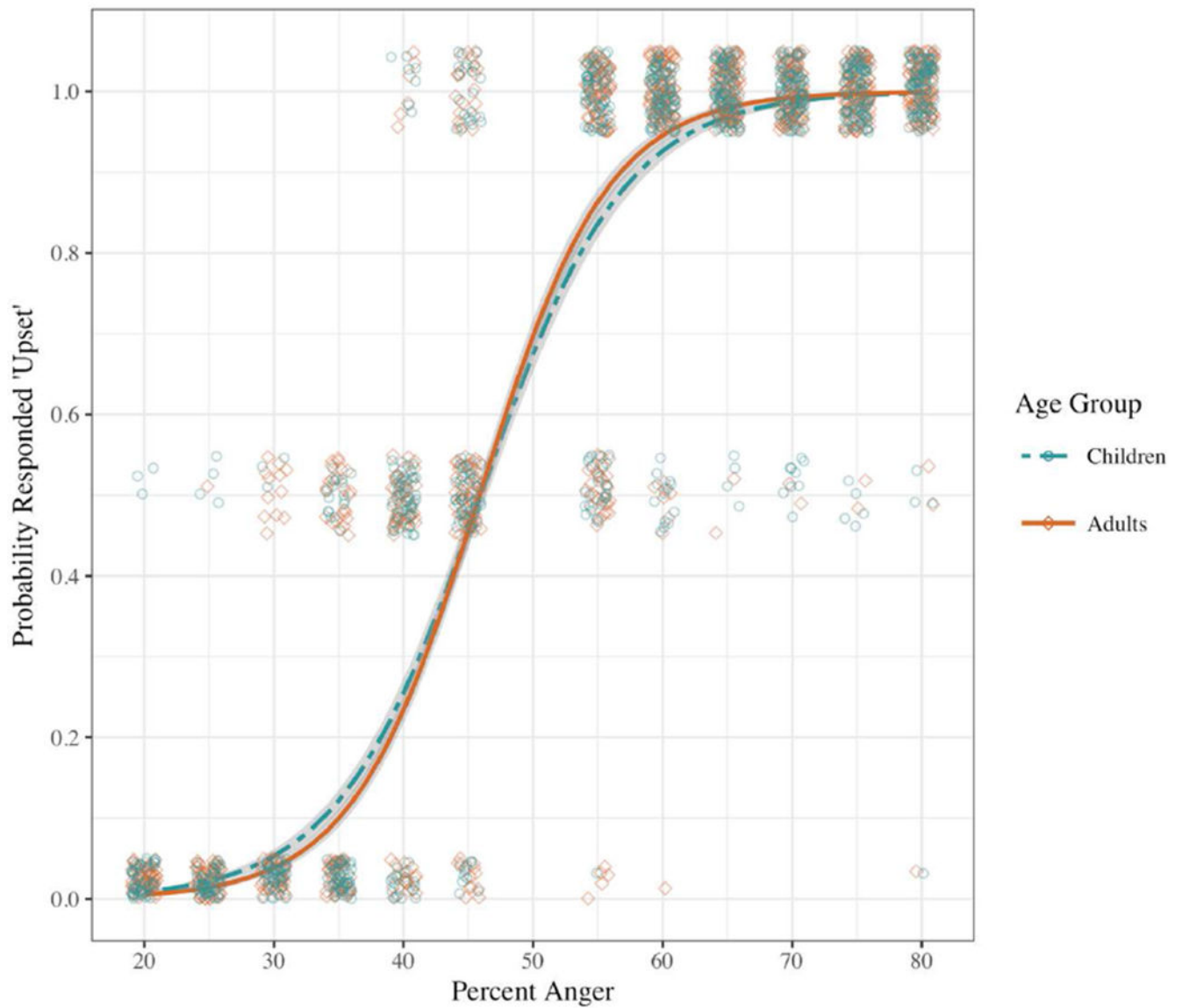
- Marchi F, & Newen A (2015). Cognitive penetrability and emotion recognition in human facial expressions. *Frontiers in Psychology*, 6, 1–12. 10.3389/fpsyg.2015.00828 [PubMed: 25688217]
- Martin J, Rychlowska M, Wood A, & Niedenthal P (2017). Smiles as multipurpose social signals. *Trends in Cognitive Sciences*, 0(0). 10.1016/j.tics.2017.08.007
- Maye J, Werker JF, & Gerken L (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101–B111. 10.1016/S0010-0277(01)00157-3 [PubMed: 11747867]
- Niedenthal PM, Rychlowska M, & Wood A (2017). Feelings and contexts: socioecological influences on the nonverbal expression of emotion. *Current Opinion in Psychology*, 17, 170–175. 10.1016/j.copsyc.2017.07.025 [PubMed: 28950965]
- Pollak LH, & Thoits PA (1989). Processes in emotional socialization. *Social Psychology Quarterly*, 22–34.
- Pollak SD, & Kistler DJ (2002). Early experience is associated with the development of categorical representations for facial expressions of emotion. *Proceedings of the National Academy of Sciences*, 99(13), 9072–9076. 10.1073/pnas.142165999
- Pollak SD, Messner M, Kistler DJ, & Cohn JF (2009). Development of perceptual expertise in emotion recognition. *Cognition*, 110(2), 242–247. 10.1016/j.cognition.2008.10.010 [PubMed: 19059585]
- Pourtois G, Schettino A, & Vuilleumier P (2012). Brain mechanisms for emotional influences on perception and attention: What is magic and what is not. *Biological Psychology*, 92(3), 492–512. 10.1016/j.biopsycho.2012.02.007 [PubMed: 22373657]
- R Development Core Team. (2008). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing Retrieved from <http://www.R-project.org>
- Russell JA, & Bullock M (1986). Fuzzy concepts and the perception of emotion in facial expressions. *Social Cognition*, 4(3), 309–341. 10.1521/soco.1986.4.3.309
- Rychlowska M, Miyamoto Y, Matsumoto D, Hess U, Gilboa-Schechtman E, Kamble S, ... Niedenthal PM (2015). Heterogeneity of long-history migration explains cultural differences in reports of emotional expressivity and the functions of smiles. *Proceedings of the National Academy of Sciences*, 112(19), E2429–E2436. 10.1073/pnas.1413661112
- Saffran JR, Newport EL, & Aslin RN (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35(4), 606–621. 10.1006/jmla.1996.0032
- Shariff AF, & Tracy JL (2011). Emotion expressions: On signals, symbols, and spandrels— A response to Barrett (2011). *Current Directions in Psychological Science*, 20(6), 407–408. 10.1177/0963721411429126
- Susskind JM, Littlewort G, Bartlett MS, Movellan J, & Anderson AK (2007). Human and computer recognition of facial expressions of emotion. *Neuropsychologia*, 45(1), 152–162. 10.1016/j.neuropsychologia.2006.05.001 [PubMed: 16765997]
- Tottenham N, Tanaka JW, Leon AC, McCarry T, Nurse M, Hare TA, & Nelson C (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry research*, 168(3), 242–249. 10.1016/j.psychres.2008.05.006 [PubMed: 19564050]
- Wood A, Lupyan G, Sherrin S, & Niedenthal P (2016). Altering sensorimotor feedback disrupts visual discrimination of facial expressions. *Psychonomic Bulletin & Review*, 23(4), 1150–1156. 10.3758/s13423-015-0974-5 [PubMed: 26542827]
- Wood A, Rychlowska M, & Niedenthal PM (2016). Heterogeneity of long-history migration predicts emotion recognition accuracy. *Emotion*, 16(4), 413–420. 10.1037/emo0000137 [PubMed: 27045686]

### Research in Context

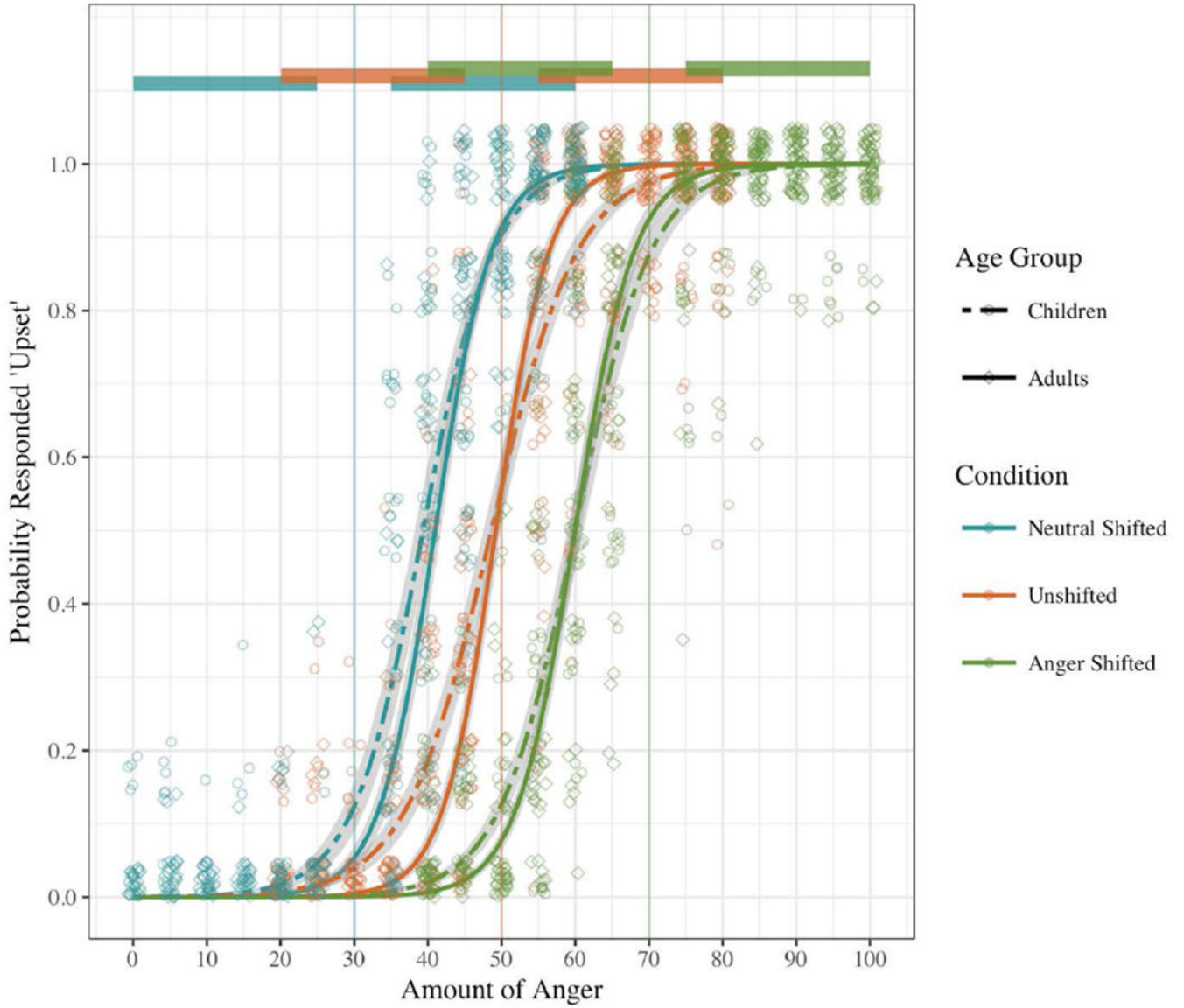
Evidence suggests that people perceive facial configurations as members of emotion categories, even though the facial muscles making up expressions can vary continuously across many feature dimensions. As with other perceptual categories, perceivers must learn the diagnostic features of emotion categories, and become sensitive to the boundaries between those categories (e.g., detecting when another person has become angry). Adding to this perceptual challenge, perceivers must also update category representations based on differences in expressivity across individuals and cultures. Given the authors' combined background in emotional development, cognitive science, and the influence of culture on emotion representations, we were interested in identifying potential learning processes that support this flexible emotion category representation. Inspired by recent interest in supervised and unsupervised learning mechanisms, we investigated these ideas in the domain of emotion. The robustness of the mechanism was probed through replication across three experiments, in environments with a single versus multiple expressers, and in the relation between statistical learning and explicit trait ratings. We are continuing this line of research to further the understanding of the influence of individual differences and contextual factors in emotion representations and the flexibility of emotion categories.



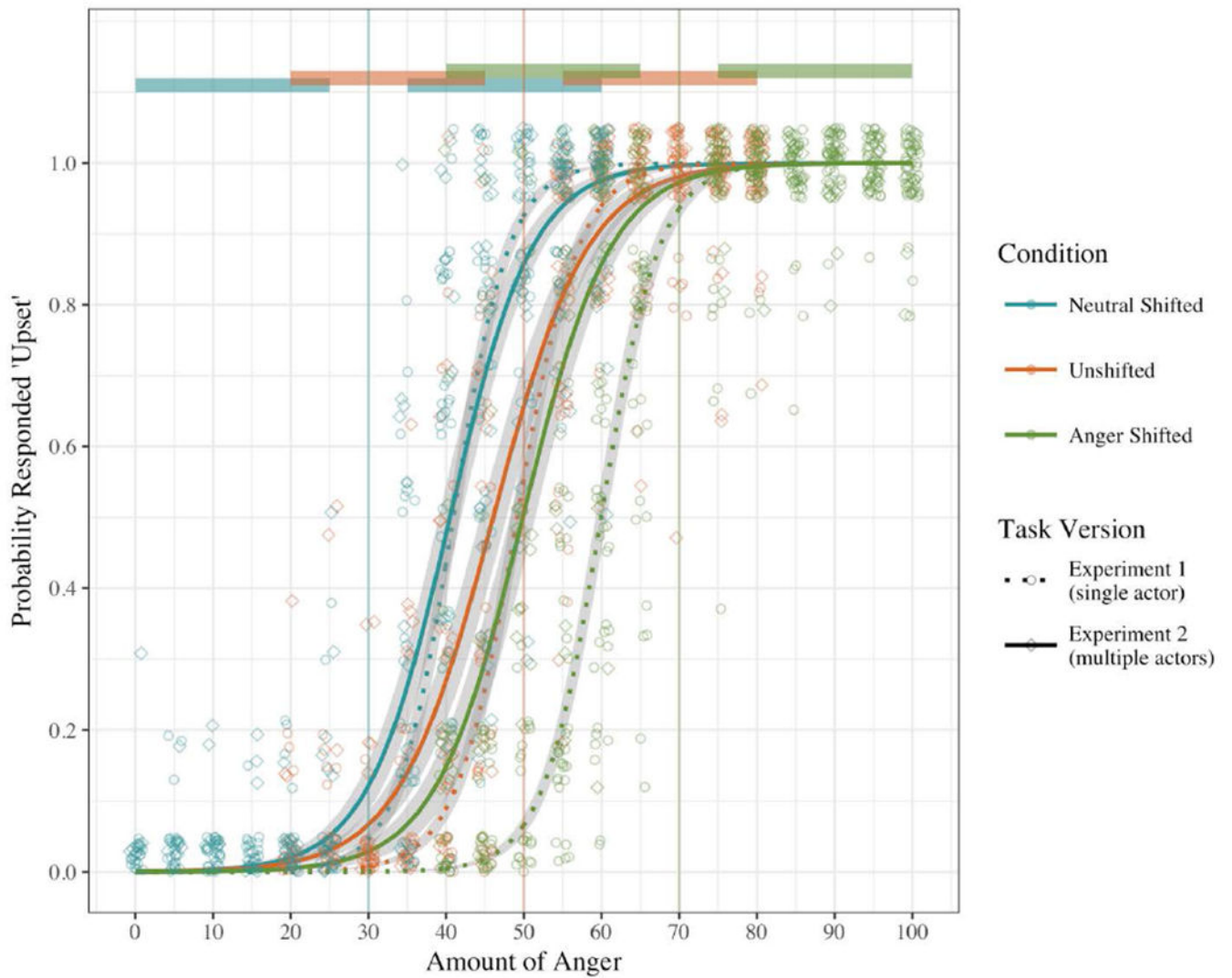
**Figure 1.** Experiment 1 stimuli and sampling strategies used in the supervised learning and testing phases. *Note.* Only a subset of the 21 morphed neutral-to-angry stimuli are represented in this figure. Percentages indicate the percent angry expression present in the morphed images: In other words, the “0%” image contained 100% neutral expression and 0% angry expression, while the “100%” image contained 0% neutral expression and 100% angry expression. The top gray rectangles illustrate the range of morphs (from 20% to 45% angry and 55% to 80% angry) used in the supervised learning phase, and the rectangles below illustrate the percent morphs (from 0% angry to 100% angry) used in each of the three conditions (representing shifted stimuli distributions) during the testing phase. Please note that the model depicted here is model #03 from Tottenham et al., 2009; model #10 was used in study (not depicted due to copyright regulations).



**Figure 2.** Experiment 1 supervised learning phase: Model predictions and participant-level data. *Note.* Lines are point estimates from logistic mixed-effects models with the interaction between Age Group and Percent Anger, and lower-order effects. Error bands represent standard errors of the point estimates. Points are individual participants' proportion of "upset" responses at a given facial expression morph value.

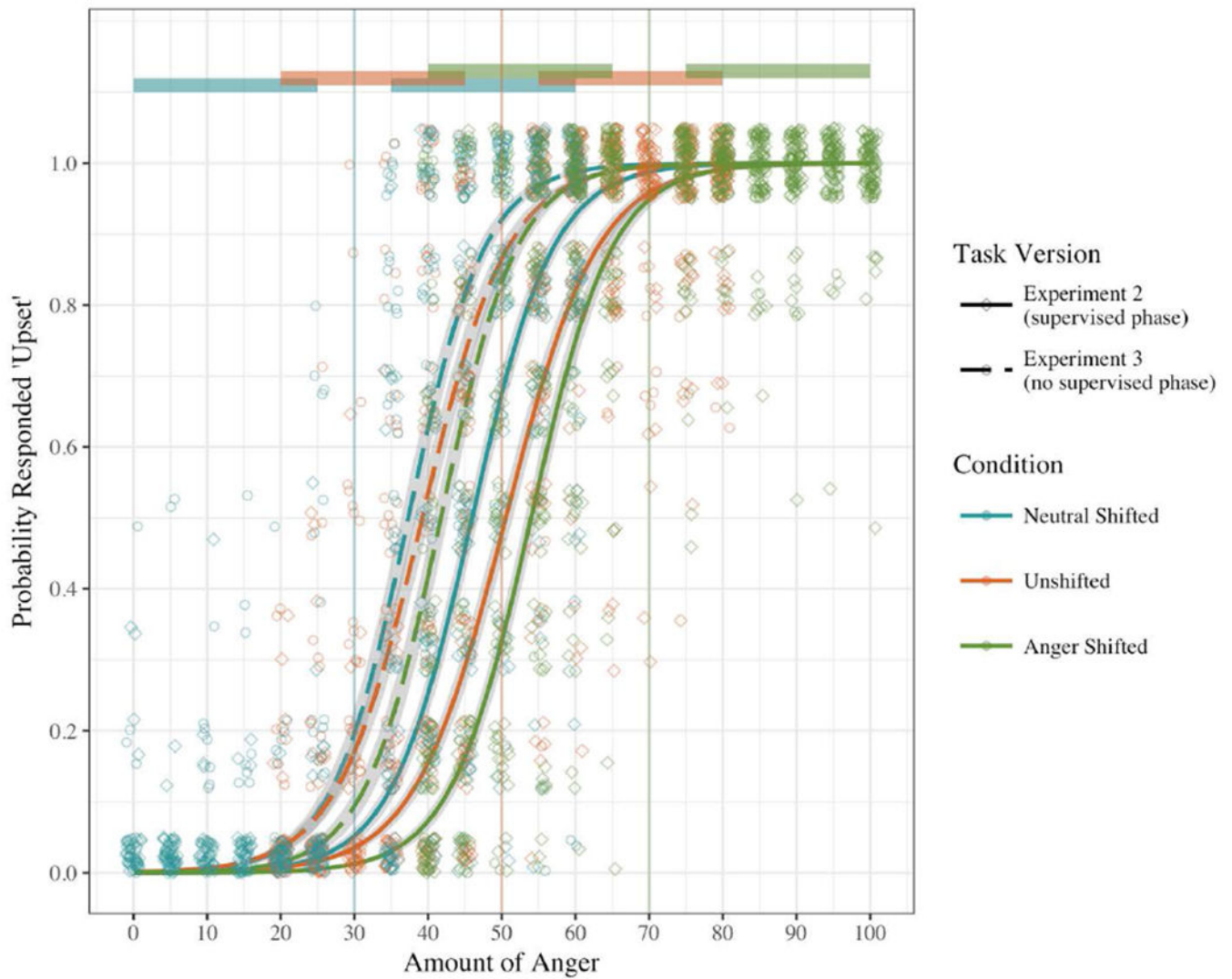


**Figure 3.** Experiment 1 testing phase: Model predictions and participant-level data. *Note.* Lines are point estimates from logistic mixed-effects models with the three-way interaction between Age Group, Condition (dummy-coded for the sake of graphing), and Percent Anger, and all lower-order effects. Error bands represent standard errors of the point estimates. Points are individual participants' proportion of "upset" responses at a given morph value. Vertical lines indicate the implied category boundary for each condition. Horizontal bars at the top indicate the range of morphed images to which participants in each of the three conditions were exposed.



**Figure 4.**

Comparison of the testing phase responses of Experiment 1 adult participants and Experiment 2 participants on all conditions (neutral shifted, unshifted, anger shifted) of “Jane” trials. While participants updated their category boundaries for Jane based on her facial expression distribution, the subsequent category shift was also influenced by the facial distributions of the other actors encountered in the environment of Experiment 2.



**Figure 5.**

Comparing Experiments 2 and 3 testing phase responses. Participants in Experiment 3 did not complete initial supervised trials. Therefore, they tended to start from a default category boundary that was much lower in anger as compared to the 50% boundary taught to participants in Experiment 2 in the supervised learning phase. Despite this main effect of removing the supervised trials (illustrated by the dashed Experiment 3 lines being further to the left), Experiment 3 participants nonetheless learned distinct category boundaries for each of the three actors.