The Representation of Emotion Knowledge Across Development

AUTHOR/ACKNOWLEDGEMENTS REMOVED

Abstract

Much is known about how adults perceive and categorize emotions, but do these findings similarly hold for young children? Here we investigate how children spontaneously represent facial cues associated with emotion. Children sorted emotion cues using a spatial arrangement method that circumvents common limitations in the study of children’s emotion knowledge, such as reliance on vocabulary. We examined children’s data using both top-down, supervised and bottom-up, unsupervised analyses. Children organize their emotion knowledge in ways that are dissimilar from adults, using broad valence-based distinctions rather than common emotion categories. Over development, children’s emotion judgments gradually align with adults as broader dimensions become refined into emotion categories. These data suggest that even among adults, valence forms the basis for how people structure perceptual information about emotions.

*Keywords:* emotion knowledge; categorization; face processing; free sorting; development

# Running Head: Representing emotion knowledge

Emotions are a rich source of information that children learn to use when formulating predictions about what is likely to occur in their environments. For example, facial movements from others, in combination with other contextual information, help children understand whether their actions are approved of by their social partners or caregivers, whether they should approach or avoid persons, objects, or situations, and whether an environment is safe. A frequently cited example of this later idea is social referencing, a phenomenon in which infants adaptively use social signals from caregivers to guide their movements and environmental exploration (Walle et al., 2017). Children’s acquisition of the ability to make use of emotion cues develops so rapidly that the considerable learning involved can appear seamless, masking potentially important changes across early development (Pollak, Camras, & Cole, 2019; Ruba & Pollak, 2020). The present study addresses how children think about and organize the perceptual input of facial configurations associated with emotions.

The longest standing theory about the structure of emotion from early infancy was proposed by Katharine Bridges (1932), who observed that children begin by fluctuating between a resting state of calm with punctuated states of distress. This view was the basis of contemporary theories that human understanding of emotions begins with differentiation between distress/lack of distress or unpleasant/pleasant, and becomes elaborated over time into fine-grained emotion categories (Nook & Somerville, 2019; Widen, 2013). These theories leave unresolved how children organize and represent the range of perceptual features they encounter to understand emotions and how this becomes elaborated over development.

The concepts most frequently used to refer to the initial building blocks of emotion are valence and arousal (Russell, 2017). Valence (positivity/negativity) can be conceptualized either as bipolar (a single scale from positive to negative with a neutral midpoint) or bivariate (two orthogonal scales of positivity and negativity; Larsen et al., 2009; Mattek et al., 2020). The dimension of arousal captures low to high activity or engagement. Other theories propose that key physical features such as open or closed mouths or the high contrasts of eyes form not only the basis of face perception, but also emotion reasoning (Caron et al., 1985). And still other views maintain that children have a rudimentary sense of a limited set of emotion categories that they use to understand facial configurations (Izard, 2007; Leppänen & Nelson, 2009).

The primary challenge to understanding how children think about various emotion cues concerns the difficulty in accurately assessing what children are perceiving when they are exposed to stimuli such as facial configurations (Barrett et al., 2019). Much of the data used to understand the structure of children’s emotion knowledge relies upon children’s production and comprehension of emotion labels (Ruba & Pollak, 2020). The most commonly used approaches in this field involve asking children to generate a verbal label to describe a facial stimulus (What is this person feeling?), confirm whether labels match an image displayed (Is this person feeling sad?), or select a stimulus from an array of predetermined response options. In the latter, children are asked to either select a label to match a face (Is this face angry, happy, sad, or scared?) or pick a face to match a label (Choose the face that looks happy). However, these approaches share three key limitations.

First, these methods are constrained by the emotion categories determined by the researcher: the researcher selects stimuli they believe represent “happy” or “sad” and accept only happy and sad as correct answers for those stimulus items. This approach can reveal the degree to which children successfully align their responses with the (adult) researcher’s view of emotion, but provide limited insight about a child’s own construal of the faces, which might not map onto any of the labels or categories that researcher selected.

Second, verbal-response methods equate knowledge of an emotion vocabulary word with a child’s use of perceptual information. This assumption can underestimate what children actually know about emotion. Many emotion words are not learned until later in development (Baron-Cohen et al., 2010), word comprehension often precedes word production (Bergelson & Swingley, 2012), and social referencing paradigms indicate that infants are adaptively using facial movements to guide their behavior long before expressive emotion vocabulary is present (Walden & Ogan, 1988). For these reasons, it is unsound to assume that a child who cannot produce, comprehend or use a word such as “scared” does not know something about the concept of fear or threat. Furthermore, seemingly simple emotion words change in abstraction across development (Nook et al., 2020), making it difficult to interpret whether children and adults even mean the same thing when using a labels such as “mad,” let alone complex ideas such as love or shame.

Third, most extant procedures were not designed to provide information about how children think about the relationships among emotion cues. Some kinds of relationships can be inferred through patterns of errors observed in verbal-response paradigms—such as the consistency of children’s confusion about anger versus disgust. Yet, for the most part, information about how children perceive and think about underlying relationships among emotion cues is limited. This limitation also reflects a broader problem in emotion research: interpretations of children’s “errors” are often predicated on the assumption that deviations from the researcher’s pre-determined label for an emotion stimulus are incorrect—that is, if the researcher has labelled a stimulus face as “sad” or a stimulus as eliciting “fear,” other interpretations or reactions to those stimuli are coded as errors.

Here, we sought to understand how children represent emotions, without introducing verbal labels or assumptions about the accuracy of participant’s responses. To do so, we adapted the Spatial Arrangement Method (SpAM) developed by Goldstone (1994), which has been used successfully to measure knowledge concepts (Richie, White, Bhatia, & Hout, 2020). Through this method, participants freely sorted facial images according to the extent to which they perceive stimuli as semantically related without imposing the use or primacy of any specific dimension, category, or label. SpAM has been successfully used with adults (Hout et al., 2013; Hout & Goldinger, 2016), and has captured changes in children’s semantic knowledge in domains such as plants and animals (Unger et al., 2016; Unger & Fisher, 2019), as well as sub-category distinctions including clothes, foods, and tools (Vales et al., 2020).

We tested predictions that follow from extant theories about the emergence of human emotion, including the possibilities that (a) children use emotion categories (Izard 1997; Keltner et al., 2019), (b) children use continuous dimensions including bipolar valence and arousal (Russell, 2003), (c) children use valence in a bivariate manner (positivity and negativity are orthogonal; Larsen et al., 2009), and (d) children use a combination of these aforementioned features (Mattek et al., 2020). It is also likely that with learning and maturation, representation of emotions changes. To explore this possibility, we tested children as young as age 3 years (the earliest age we conjectured children would reliably use this method) through age 7 years (when children label many emotions similarly to adults) and compared children’s behaviors to those of adults. We approached the data in two distinct ways: (1) a top-down, supervised approach to test the extent to which predefined emotion categories and dimensions predict sorting behavior, and (2) a bottom-up, unsupervised approach examining participants’ behavior without prescribing primacy to any given theory or weight to any specific (purportedly predictive) dimension.

# Method

## Participants. We recruited 107 children (ages 3;0-6;11; mean age = 5.0, sd age = 1.10; 48 M, 59 F; race: 6.5% more than one race, 84.1% White, 9.3% Asian) and 40 adults (18-21; mean age = 18.83, sd age = 0.73; 10 M, 30 F; race 10% Hispanic, 30% Asian, 2.5% Black, 57.5% White). Two children did not complete all sorts and were excluded from the sorts they did not complete. We aimed to have 30 children in each age bin but had to stop data collection early because of the COVID-19 outbreak; the sample reported here includes 21 3-year-olds, 35 4-year-olds, 28 5-year-olds, and 23 6-year-olds. 20 participants per subgroup has provided sufficient power for most cluster analysis techniques (Dalmaijer et al., 2020), and our sample size is comparable to those used in past studies using the spatial arrangement method with children (n = 18 per group, Unger et al., 2016).

## Stimuli. Stimuli were drawn from the Interdisciplinary Affective Science Laboratory (IASLab) Facial Stimuli Set.[[1]](#footnote-1) We selected actors with the highest average accuracy ratings and no facial hair. The stimuli were designated by IASLab as open and closed mouth versions of anger, calm, disgust, excitement, fear, happiness, neutral, sadness, and surprise for a total of 18 images in each sorting condition. To test for the robustness of any possible effects, each participant completed two sorting conditions. One sorting condition consisted of 18 different facial configurations posed by the same individual; the other sorting condition consisted of 18 different individuals (half male and half female, with a male and female for each emotion). In this manner, the Same Individual condition reveals how participants construe different facial configurations from one individual, whereas the Different Individual condition reflects a generalization across individual actors, allowing examination of whether similar sorting patterns emerge when a variety of different perceptual features are changing (facial cue, identity, race, and gender).

*Ratings of stimuli.* Fifty undergraduates who did not participate in the sorting task completed 7-point Likert ratings of bipolar valence and arousal (Warriner et al., 2013) for each image. We also used the Evaluative Space Grid (ESG; Larsen, et al., 2009) method to collect ratings of bivariate valence (i.e., ratings of positivity and negativity). Valence is often treated as a bipolar measure ranging from negative at one pole to positive at the other with a neutral midpoint. However, bivariate valence—representing positivity and negativity in a two-dimensional space—has been found to more accurately capture emotional experience (Larsen & McGraw, 2011; Watson et al., 1999). Traditional bipolar valence scales pose interpretive challenges: scores in the middle of the scale could indicate that the individual perceives the stimulus as neither positive nor negative (indifference, neutrality), that the individual perceives a mix of positivity and negativity (ambivalence, multiple emotions), or that the perceiver is uncertain (a stimulus could be either positive or negative depending upon the context). The ESG method disentangles these possibilities by presenting participants with a square depicting a 5-point positivity scale on one axis and a 5-point negativity scale on the other, allowing participants to select where the stimuli fall along both dimensions. Additional details on stimuli ratings are available in the Supplemental Material.

### Design & Procedure. Images were presented on a Dell 24: P2418HT touchscreen monitor using PsychoPy [version v1.83.04; Peirce et al., 2019]. At the outset of each sorting condition, participants saw all the images to be sorted. The images then disappeared, and each image was presented one at a time in the center of the screen. Participants were instructed to arrange the images that go together, or are the same kind of thing, by touching the images they wished to move and dragging them to any location they wished on the grid. To introduce participants to the task, they saw 4 images (soccer ball, basketball, rabbit, and chair) and practiced moving them around on the screen. The grid had no labels or axes, so participants were not sorting onto a predefined space. In order to ensure that images were clearly visible to participants, images would expand in size (from 140x140 to 315x315) while participants touched them to move the image, and then returned to their original size once placed in the grid. When participants were no longer moving images, the experimenter asked if they were ready for the next picture. Participants could continue to move each image as often as they wanted throughout the task.

### The experiment began with a practice in which participants arranged 5 images (car, bus, squirrel, bird, table). The practice was not scored, because the principle of the method is that there are not wrong answers (see Supplemental Material for more details); however, this allowed us to assess whether participants understood the task (by, for example, grouping the animals together or, as one child explained, grouping the squirrel and the table together because “they both have legs”). For the next two conditions, participants saw faces with instructions to think about how the person might be feeling, and that people feeling the same kind of thing should go together. Participants then completed a *Same Individual Condition* in which they sorted 18 facial cues of emotion for one actor (Male # 7). Next, participants completed a *Different Individual Condition*, this time sorting 18 facial configurations posed by 18 different actors (Females: # 1, 4, 7, 10, 13, 14, 15, 17, 22; Males: # 2, 3, 4, 5, 8, 12, 14, 15, 17).

# Results

Analyses were conducted in R (version 4.0.3; R Development Core Team, 2020), fitting linear mixed-effects models using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015; version 1.1-26). Following the recommendations of Luke (2017), *F*-values and *p*-values for linear mixed-effects models were obtained using the Satterthwaite approximation of the degrees of freedom (Kuznetsova, Brockhoff, & Christensen, 2017). Participant’s patterns of sorting behavior were characterized by calculating the Euclidean distance between images, which were then normalized for each participant by scaling distances based on the maximum distance for each participant. All analysis code and analytic details can be found on the project’s OSF page, including a walkthrough of each analysis (https://rpubs.com/zcm/GRD\_main) and supplementary details (https://rpubs.com/zcm/GRD\_supplemental). We analyzed data using top-down, supervised approaches followed by bottom-up, unsupervised approaches.

**Comparing dimensions of affect and categories in sorting behaviors.**

We began with top-down, supervised methods to examine whether emotion category and dimensions of affect account for how closely different facial cues are placed to one another. We examined these features separately, and then compare how well the various dimensions and categories account for sorting behaviors.

## *Emotion Category.* We first investigated developmental change in the use of common English language emotion categories (e.g., sad, happy, anger, disgust, fear, surprise, neutral, calm, excitement[[2]](#footnote-2)) as a structure for emotion cues. To do so, we computed the average distance between images that shared the same category label (e.g., the distance of one happy face to another happy face) versus images that had differing category labels (e.g., the distance of one happy face to a sad face) for each participant (see also Unger et al., 2016 for a similar approach). To do so, we fit a linear mixed-effects model estimating the average distance between item pairs for adults versus children (coded .5,-.5), the category match for an image pair (same category pair vs. different category pair; centered), and their interaction with a by-participant random intercept and a by-participant random slope for category match. We analyzed results collapsing across sorting conditions, as there was no evidence that results differed between the same and different individual sorts (*p* = .43). Adults were more likely than children to place images belonging to the same emotion categories closer together than images belonging to different emotion categories, *b* = -0.15, Wald 95% CI = [-0.17,-0.12], *F*(1,172.04) = 122.02, *p* < 0.001.

To understand how children’s use of emotion categories changed across development, we next fit a linear mixed-effects model on the child data with age (in years; centered) as a continuous predictor with an otherwise identical model structure. Children were more likely to sort facial configurations based upon emotion category labels with increasing age, *b* = -0.03, Wald 95% CI = [-0.04, -0.02], *F*(1,131.25) = 26.49, *p* < 0.001. This developmental increase in use of category labels is shown in Figure 1 (see Supplementary Materials for plots representing age as a continuous variable). Follow-up analyses of each age group separately reveals that neither 3-year-olds (*p* = .45) nor 4-year-olds (*p* = .33) showed evidence of sorting based upon emotion categories, while 5-year-olds (*b* = -0.06, Wald 95% CI = [-0.09, -0.04], *F*(1,31.81) = 22.27, *p* < .001) and 6-year-olds (*b* = -0.11, Wald 95% CI = [-0.14, -0.08], *F*(1,33.94) = 65.47, *p* < .001) began using category information, though much less than adults.

Chart, box and whisker chart

Description automatically generated

*Figure 1*. Use of emotion categories in sorting behavior. Y-axis represents the difference in average distance for items belonging to the same vs. different emotion categories. An average value of zero represents no distinction by emotion category as faces from the same versus different emotion categories were equally far apart. Error bars represent 95% confidence intervals

*Dimensions of Affect.* Next, we tested whether bipolar valence, bivariate valence (separate ratings of positivity and negativity), and arousal predicted participant’s sorting behavior. To do so, we fit a series of linear mixed-effects models regressing the average distance between item pairs on the similarity of the dimension of interest (bipolar valence, arousal, positivity, and negativity) – measured in terms of the difference in average stimulus rating between image pairs. This analysis included age group (adults: .5; children: -.5), and interactions including a by-participant random intercept, a by-participant random slope for the dimension of interest, and a by-item random intercept. Adults were more likely than children to use each of the four dimensions to guide their sorting behaviors (bipolar valence: *b* = 0.07, Wald 95% CI =[0.06, 0.08], *F(*1, 145.44)= 148.24, *p* < .001; arousal: *b* = 0.03, Wald 95% CI =[0.02, 0.04], *F*(1, 142.97) = 42.19, *p* < .001; positivity: *b* = 0.10, Wald 95% CI =[0.08, 0.11], *F*(1, 145.50) = 146.91, *p* < .001; negativity: *b* = 0.11, Wald 95% CI =[0.09, 0.13] *F*(1, 144.36) = 110.77, *p* < .001). To further understand the developmental change in children’s use of each dimension, we fit linear mixed-effects models on the child data with age (in years; centered) as a continuous predictor and an otherwise identical model structure. Children increasingly used each feature across development (valence: *b* = 0.01 , Wald 95% CI =[0.01, 0.02], *F*(1, 103.81) = 30.91, *p* < .001; positivity: *b* = 0.01, Wald 95% CI =[0.01, 0.02], *F*(1, 103.87) = 21.36, *p* < .001; negativity: *b* = 0.03 , Wald 95% CI =[0.02, 0.03], *F*(1, 102.88) = 34.69, *p* < .001)— with the exception of arousal, *b* = 0.003, Wald 95% CI =[-0.001, 0.01], *F*(1, 103.62) = 2.02, *p* = .16. The pattern for arousal highlights how children’s development may not always occur as straightforward linear differentiation (see Supplemental Material section 3 for additional details).

*Comparing dimensions of affect and categories in sorting behaviors.* Finally, we examined how well emotion category predicted participant’s sorting behavior compared to valence and arousal. To do so, we computed the average distance between all stimulus pairs (n = 306 unique pairs) for each age group and predicted these distances from a pair’s similarity on each dimension of interest simultaneously. This general linear model revealed how much each dimension aided in explaining variance in each age group’s sorting behavior. First, we estimated the use of bipolar valence, arousal, and whether image pairs shared the same discrete emotion category (0 = different category pair; 1 = same category pair). Second, we estimated the effects of bivariate valence with positivity and negativity as two orthogonal dimensions.

*Bipolar valence, arousal, and shared emotion category.*Valence emerged as (by far) the strongest predictor (Figure 2) of how participants grouped facial images, an effect that increased steadily with age. Arousal was a significant predictor for 4-year-olds, but declined as children grew older. Consistent with the results from the previous section, emotion category did not emerge as a predictor until age 5 years. The total variance explained by this model increased steadily across age (Table 1), accounting for a significant amount of the error variance for all age groups: *F*(3, 302) > 15, *p* < .001) with the exception of the youngest age group (3-year-olds: *F*(3, 302) = 1.33, *p* = .26.

Chart, bar chart

Description automatically generated

*Figure 2*. Delta R-squared for each predictor of sorting behavior. Error bars represent bootstrapped 95% confidence intervals.

Table 1: Predicting Sorting Distance from Valence, Arousal, and Shared Emotion Category

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Predictor** | **Estimate** | ***t*-value** | ***p*** | **ΔR2** | **Overall R2** |
| ***3-year-olds*** |  |  |  |  | .01 |
| **Valence** | 0.004 | 1.67 | .10 | .01 |  |
| **Arousal** | -0.004 | -1.12 | .27 | .00 |  |
| **Emotion Category** | -0.006 | -0.55 | .58 | .00 |  |
| ***4-year-olds*** |  |  |  |  | .13 |
| **Valence\*\*\*** | 0.01 | 5.73 | <.001 | .09 |  |
| **Arousal\*\*\*** | -0.02 | -4.76 | <.001 | .07 |  |
| **Emotion Category** | -0.002 | -0.16 | .87 | .00 |  |
| ***5-year-olds*** |  |  |  |  | .31 |
| **Valence\*\*\*** | 0.03 | 10.76 | <.001 | .26 |  |
| **Arousal\*\*** | -0.02 | -3.16 | .002 | .02 |  |
| **Emotion Category\*** | -0.03 | -2.22 | .027 | .01 |  |
| ***6-year-olds*** |  |  |  |  | .46 |
| **Valence\*\*\*** | 0.05 | 14.49 | <.001 | .37 |  |
| **Arousal\*\*\*** | -0.02 | -3.82 | <.001 | .03 |  |
| **Emotion Category\*\*\*** | -0.06 | -3.63 | <.001 | .02 |  |
| ***Adults*** |  |  |  |  | .78 |
| **Valence\*\*\*** | 0.09 | 30.04 | <.001 | .65 |  |
| **Arousal\*\*\*** | -0.02 | -3.83 | <.001 | .01 |  |
| **Emotion Category\*\*\*** | -0.08 | -5.26 | <.001 | .02 |  |

*Note.* Asterisks denote significance level, \**p* < .05; \*\**p* < .01; \*\*\**p* < .001.

*Bivariate valence, arousal, and shared emotion category.*We repeated the previous analysis replacing bipolar valence with bivariate valence (positivity and negativity as independent predictors). As expected, ratings of positivity and negativity were highly correlated with bipolar ratings, precluding us from including all five predictors in the same model. The dimension of negativity emerged as the strongest predictor of sorting behavior across all age ranges, even 3-year-olds, and explained substantially more unique variance than positivity, arousal and emotion category (Table 2).

Table 2: Predicting Sorting Distance from Positivity, Negativity, Arousal, and Shared Emotion Category

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Predictor** | | | **Estimate** | | ***t*-value** | ***p*** | | **ΔR2** | **Overall R2** | |
| ***3-year-olds*** |  |  | |  | | |  | | | .02 |
| **Positivity** | | | -0.005 | | -1.01 | .31 | | .00 |  | |
| **Negativity\*** | | | 0.01 | | 2.23 | .027 | | .02 |  | |
| **Arousal** | | | -0.00004 | | -0.01 | .99 | | .00 |  | |
| **Emotion Category** | | | -0.004 | | -0.33 | .74 | | .00 |  | |
| ***4-year-olds*** | | |  | |  |  | |  | .18 | |
| **Positivity** | | | -0.003 | | -0.70 | .49 | | .00 |  | |
| **Negativity\*\*\*** | | | 0.02 | | 5.66 | <.001 | | .09 |  | |
| **Arousal\*** | | | -0.009 | | -2.43 | .02 | | .02 |  | |
| **Emotion Category** | | | 0.004 | | .42 | .68 | | .00 |  | |
| ***5-year-olds*** | | |  | |  |  | |  | .45 | |
| **Positivity\*** | | | -0.01 | | -2.48 | .014 | | .01 |  | |
| **Negativity\*\*\*** | | | 0.07 | | 12.02 | <.001 | | .26 |  | |
| **Arousal** | | | 0.006 | | 1.22 | .22 | | .00 |  | |
| **Emotion Category** | | | -0.02 | | -1.31 | .19 | | .00 |  | |
| ***6-year-olds*** | | |  | |  |  | |  | .57 | |
| **Positivity** | | | -0.01 | | -1.27 | .21 | | .00 |  | |
| **Negativity\*\*\*** | | | 0.09 | | 13.87 | <.001 | | .27 |  | |
| **Arousal** | | | 0.01 | | 0.92 | .36 | | .00 |  | |
| **Emotion Category\*\*** | | | -0.04 | | -2.89 | .004 | | .01 |  | |
| ***Adults*** | | |  | |  |  | |  | .80 | |
| **Positivity\*\*\*** | | | 0.05 | | 8.36 | <.001 | | .05 |  | |
| **Negativity\*\*\*** | | | 0.10 | | 15.39 | <.001 | | .16 |  | |
| **Arousal** | | | -0.003 | | -0.54 | .59 | | .00 |  | |
| **Emotion Category\*\*\*** | | | -0.08 | | -4.92 | <.001 | | .02 |  | |

*Note.* Asterisks denote significance level, \**p* < .05; \*\**p* < .01; \*\*\**p* < .001.

*Does bivariate valence predict sorting behavior better than bipolar valence?* To determine whether separate dimensions of positivity and negativity were better predictors than bipolar valence, we replaced the valence predictor with separate predictors of positivity and negativity. Bivariate dimensions of valence were a better predictor of sorting behavior in all but the youngest age group, with the most substantial gains in predictor among the 5- and 6-year-olds (3-year-olds: *F*(1, 301) = 2.82, *p* = .09; 4-year-olds: *F*(1, 301) = 19.46, *p* < .001; 5-year-olds: *F*(1, 301) = 74.58, *p* < .001; 6-year-olds: *F*(1, 301) = 77.51, *p* < .001; adults: *F*(1, 301) = 21.88, *p* < .001).

**Bottom-up assessment of emotion knowledge**

Our last sets of analyses used unsupervised methods to provide a complementary perspective on how emotions might be represented. These analyses extract patterns from the sorting data without relying on any predetermined dimensions or categories. We considered Same and Individual Sorts separately because the following analyses require pairwise distances between all items, which are only available within a given sorting block .

First, we used 2-dimensional multidimensional scaling (MDS) to visually represent participant’s classifications (Figure 3). To better understand the dimensions, we fit vectors of image ratings for bipolar valence, arousal, positivity, and negativity onto our MDS solution over 1,000 permutations to derive the squared correlation coefficient of each vector (envfit in Vegan package; Oksanen, 2019). This analysis reveals that stimuli ratings of valence, positivity, and negativity consistently correlate with the MDS dimensions (*r*2 > 0.84 and *p* < .001 across all sort conditions for both adults and children). Arousal only correlated with the dimensions in the Same Individual Sort (Adults: *r*2 = 0.40, *p* < .05; Children: *r*2 = 0.44 , *p* < .05) but not in the Different Individual Sort (Adults: *r*2 = 0.10, *p* = .46; Children: *r*2 = 0.22, *p* = .15). Additional MDS visualizations are provided in Supplemental Material section 5.

**Chart

Description automatically generated**

Figure 3: Classical multidimensional scaling solution (2 dimensions) for average sorting distances across all children and adults in the Same Individual (A, B) and Different Individual (C, D) Sorts. Vectors show squared correlation coefficients between image ratings and the MDS dimensions.

Second,we used hierarchical clustering to examine age-related changes in how participants organized emotion cues (Ward's method; Ward, 1963). Clustering was performed on distance matrices calculated for each age group in each sorting condition using the pairwise distances between all sorted images (see Supplemental Material section 6 for further detail). As expected, children’s clustering structures become increasingly adult-like (Table 3, Figure 4). Children’s increasingly adult-like structures appear to be driven by changes in emotion knowledge and not improvement on the task generally, as the practice structure is highly similar to adults for all age groups except 3-year-olds. Changes in children’s clusters also appear systematic, as children closer in age are more similar to one another. For example, the sorting behavior of 5-year-olds had a stronger correlation with 6-year-olds and 4-year-olds than with adults.

Table 3: Comparison of children’s hierarchical clustering solutions to adult’s clustering solutions

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Practice Sort* | | *Same Individual Sort* | | *Different Individuals Sort* | |
| *Age Group* | Adj. Rand (k= 3) | *c* | Adj. Rand (k= 3) | *c* | Adj. Rand (k= 3) | *c* |
| *3-year-olds* | 0.21 | -.03 | 0.02 | 0 | 0.16 | 0.21 |
| *4-year-olds* | 1.0 | 0.84 | 0.14 | 0.2 | 0.14 | 0.2 |
| *5-year-olds* | 1.0 | 0.99 | 0.49 | 0.41 | 0.49 | 0.38 |
| *6-year-olds* | 1.0 | 0.98 | 0.83 | 0.65 | 0.38 | 0.40 |

*Note.* Each value in the table represents the similarity between children’s clustering at a specific age group and adults’ clustering solution. An adjusted Rand index of 0 indicates two clusters have a Rand index that matches the expected value for random groupings, with higher and lower values indicating higher- or lower-than-chance level similarity between the two clusters. *c* is the cophenetic correlation coefficient between the two dendrograms. This value ranges from -1 to 1 with values near 0 suggesting that the two dendrograms are not statistically similar. Additional measurements of similarity for all values of k are available in the Supplemental Material section 6.

Diagram

Description automatically generated

A

Diagram, histogram

Description automatically generated

B

*Figure 4*: Dendrograms and correlations between dendrogram structures for (A) Same Individual Sort and (B) Different Individual Sort. The numbers 1 and 2 indicate that the images had open and closed mouths, respectively. Colors specify the three cluster solutions for each age-bin and highlight commonalities across dendrograms. Red clusters contain anger and disgust images, green clusters contain certain fear and neutral images, and blue images contain certain happy, calm, and surprise images. 3-year-olds dendrograms are colored differently as they showed less differentiation.

These changes in organizational structure can also be seen in dendrograms, which visualize the hierarchical clustering solutions. Each facial image is a node on the dendrogram that forms another node (represented by a horizontal line) when it merges with another face. Clusters are determined by the vertical height of the branches in a dendrogram, not by which labels are closest to one another laterally. Thus, faces that were found to be the most similar would be connected as a node with a very low height. To ensure that the hierarchical clustering solutions captured meaningful groupings of emotion cues, we predicted the bipolar valence ratings of cues using cluster group (k = 3). As in the distance-based analyses above, bipolar valence was a strong predictor of both children’s (Same Individual Sort: F(2,15)=73.86, *p*<.001; Different Individual Sort: F(2,15)=26.03, *p*<.001), and adults’ (Same Individual Sort: F(2,15)=51.87, *p*<.001; Different Individual Sort: F(2,15)=61.71, *p*<.001) cluster groups.

# General Discussion

### This study reveals developmental changes in how humans represent perceptual information associated with emotions. By using a non-verbal, open-ended procedure, we circumvented the traditional limitations incurred in the assessment of emotion knowledge in young children. We found that children primarily rely upon dimensions such as valence. Adult-like reliance on common English language emotion categories (happy, sad, angry, etc.) only gradually emerged, with little evidence that children consistently used these categories until around five years of age. Even 6-year-olds still showed less emotion category-based sorting compared to adults. Similar patterns of incremental change in how children represent emotion emerged in unsupervised analyses that did not impose *a priori* assumptions about the structure of the stimuli, such as specific emotion categories or dimensions of affect.

*Nuances in the use of valence*

Valence accounted for a very large proportion of all participant’s emotion judgments, providing converging evidence with prior studies (e.g., Jackson et al., 2019; Nook et al., 2017). However, the present data reveal new insights about this dimension likely to inform future research. First, treating valence as bivariate (represented by separate unipolar scales of positivity and negativity) better accounted for behavior than treating it as a bipolar construct (a single continuum ranging from positive to negative). Second, positivity and negativity are not used equally early in development. Young children relied heavily negativity and did not consistently use positivity. Allowing positivity and negativity to exist separately might also better capture human experience: One can experience spicy food as both painful and delicious, or horror movies as both frightening and entertaining.

*The limits of arousal*

Though often discussed in tandem with valence, we found that arousal decreased in use across development after age four, explained a much smaller proportion of behavior than valence, and did not consistently correlate with any multidimensional scaling solution. This limited role of arousal is in contrast with many theories of emotion that posit emotions initially emerge from a 2-factor understanding of valence and arousal (for reviews see Barrett & Bliss-Moreau, 2009; Russell & Barrett, 1999), and is inconsistent with studies that rely upon arousal as a key aspect of emotion (Bliss-Moreau et al., 2020; Nook et al., 2017). There are a number of reasons for these divergent conclusions. First, arousal can be presented to research participants in different ways, including subjective feelings of activation, perceptions of calmness/deactivation or excitement/activation, or perceptions of intensity. These definitions could elicit a variety of interpretations from respondents. Second, arousal is often offered as a speculative explanation of the data without actually measuring arousal using independent ratings. Third, arousal may index natural covariation in positivity and negativity, rather than capturing unique variance in emotion (Haj-Ali et al., 2020; Kron et al., 2013). Our data support this third possibility, as the variance explained by arousal disappears when we use bivariate valence (positivity and negativity).

*What changes in the structure of emotion knowledge over development?*

Our data suggest that developmental changes in how emotions are represented are not simply reflecting children’s responses becoming more consistent or children becoming more competent at the task with increasing age; instead, the manner in which children prioritized and used different dimensions of emotion changed across age. Starting at 4 years of age, children systematically organized images according to broader dimensions that were not necessarily linear: children closer in age had clustering structures that were much more aligned with one another than with those of adults. These results suggest that children prioritize perceptual information about emotion in a systematic manner that is distinct from how adults organize this same information.

The patterns that we observed in the development of emotion knowledge appear similar to those discovered in other domains of development. For instance, the development of non-emotional categories (e.g., animals and other natural kinds) reveals that children first make broad distinctions (e.g., birds v. mammals) and later show finer differentiation of items based on their category membership (e.g., ostriches v. peacocks; Vales et al., 2020). The present data uncover a similar pattern of finer-grained differentiation across development for emotion knowledge. We found children first use broad, primarily valence-based distinctions, and with greater experience, draw more fine-grained distinctions that use emotion category information (Matthews et al., 2020; Widen, 2013). Rather than a distinct shift from using valence to using emotion categories, we found continued and refined use of valence and emotion categories across development.

# *Limitations and Future Directions*

Although we used a diverse set of stimuli, including open and closed mouthed images of nine emotion categories, from Asian, Black, and White males and females, future research would benefit from even greater variety. For example, the emotion categories used here were those commonly used in English, and were posed rather than naturally occurring. Moreover, facial cues are not the only source of emotion information during human interactions. Future research is needed about children’s reasoning when more situational context is available (e.g., Srinivasan & Martinez, 2018), as well as with video and audio stimuli (Woodard, Plate, Morningstar, Wood, & Pollak, in press) to provide greater understanding of how emotions develop.

# Conclusions

Emotions are critical for human adaptation and survival, yet relatively little is understood about how humans come to understand and represent emotion signals. Several explanations commonly used to account for the early development of emotion find little or only partial support in the present data. Young children in our task did not begin to use basic emotion categories until around the age of 5, arguing against the theory that this knowledge plays a large role in young children’s emotion understanding. Children also did not rely equally on the dimensions of valence and arousal, instead using negative valence far more heavily. The picture of emotion development that emerges from our data is of an incremental learning process in which children change their representations of emotion using combinations of factors that are weighted differently across development. This insight opens the door for new investigations about how humans learn to navigate the complex communicative system of the social world.

# References

Baron-Cohen, S., Golan, O., Wheelwright, S., Granader, Y., & Hill, J. (2010). Emotion Word Comprehension from 4 to 16 Years Old: A Developmental Survey. *Frontiers in Evolutionary Neuroscience*, *2*(NOV), 25. https://doi.org/10.3389/fnevo.2010.00109

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Barrett, L. F., & Bliss-Moreau, E. (2009). Chapter 4 Affect as a Psychological Primitive. In *Advances in Experimental Social Psychology* (Vol. 41, pp. 167–218). Academic Press. https://doi.org/10.1016/S0065-2601(08)00404-8

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects

models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

<https://doi.org/10.18637/jss.v067.i01>

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

Bliss-Moreau, E., Williams, L. A., & Santistevan, A. C. (2020). The immutability of valence and arousal in the foundation of emotion. *Emotion*, *20*(6), 993–1004. https://doi.org/10.1037/emo0000606

Bridges, K. M. B. (1932). Emotional Development in Early Infancy. *Child Development*, *3*(4), 324. https://doi.org/10.2307/1125359

Caron, R. F., Caron, A. J., & Myers, R. S. (1985). Do Infants See Emotional Expressions in Static Faces? *Child Development*, *56*(6), 1552. https://doi.org/10.2307/1130474

Dalmaijer, E. S., Nord, C. L., & Astle, D. E. (2020). Statistical power for cluster analysis. *ArXiv*. http://arxiv.org/abs/2003.00381

Goldstone, R. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers*, *26*(4), 381–386. https://doi.org/10.3758/BF03204653

Haj-Ali, H., Anderson, A. K., & Kron, A. (2020). Comparing three models of arousal in the human brain. *Social Cognitive and Affective Neuroscience*, *15*(1), 1–11. https://doi.org/10.1093/scan/nsaa012

Hout, M. C., & Goldinger, S. D. (2016). SpAM is convenient but also satisfying: Reply to verheyen et al. (2016). *Journal of Experimental Psychology: General*, *145*(3), 383–387. https://doi.org/10.1037/xge0000144

Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient, spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General*, *142*(1), 256–281. https://doi.org/10.1037/a0028860

Izard, C. E. (1997). Emotions and facial expressions: A perspective from Differential Emotions

Theory. In J. A. Russell & J. M. Fernández-Dols (Eds.), *Studies in emotion and social*

*interaction, 2nd series. The psychology of facial expression* (p. 57–77). Cambridge

University Press; Editions de la Maison des Sciences de l'Homme.

<https://doi.org/10.1017/CBO9780511659911.005>

Izard, C. E. (2007). Basic Emotions, Natural Kinds, Emotion Schemas, and a New Paradigm. *Perspectives on Psychological Science*, *2*(3), 260–280. https://doi.org/10.1111/j.1745-6916.2007.00044.x

Jackson, J. C., Watts, J., Henry, T. R., List, J. M., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, *366*(6472), 1517–1522. https://doi.org/10.1126/science.aaw8160

Keltner, D., Tracy, J. L., Sauter, D., & Cowen, A. (2019). What Basic Emotion Theory Really Says for the Twenty-First Century Study of Emotion. In *Journal of Nonverbal Behavior* (Vol. 43, Issue 2, pp. 195–201). Springer New York LLC. https://doi.org/10.1007/s10919-019-00298-y

Kron, A., Goldstein, A., Lee, D. H. J., Gardhouse, K., & Anderson, A. K. (2013). How Are You Feeling? Revisiting the Quantification of Emotional Qualia. *Psychological Science*, *24*(8), 1503–1511. https://doi.org/10.1177/0956797613475456

Kuznetsova A., Brockhoff P.B. and Christensen R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software, 82(13), pp. 1–26. doi: 10.18637/jss.v082.i13.

Larsen, J. T., & McGraw, A. P. (2011). Further evidence for mixed emotions. *Journal of Personality and Social Psychology*, *100*(6), 1095–1110. https://doi.org/10.1037/a0021846

Larsen, J. T., Norris, C. J., McGraw, A. P., Hawkley, L. C., & Cacioppo, J. T. (2009). The evaluative space grid: A single-item measure of positivity and negativity. *Cognition & Emotion*, *23*(3), 453–480. https://doi.org/10.1080/02699930801994054

Leppänen, J. M., & Nelson, C. A. (2009). Tuning the developing brain to social signals of emotions. In *Nature Reviews Neuroscience* (Vol. 10, Issue 1, pp. 37–47). NIH Public Access. https://doi.org/10.1038/nrn2554

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Mattek, A. M., Burr, D. A., Shin, J., Whicker, C. L., & Kim, M. J. (2020). Identifying the Representational Structure of Affect Using fMRI. *Affective Science*, *1*(1), 42–56. https://doi.org/10.1007/s42761-020-00007-9

Matthews, C. M., Thierry, S. M., & Mondloch, C. J. (2020). Recognizing, Discriminating, and Labeling Emotional Expressions in a Free-Sorting Task: A Developmental Story. *Emotion*. Advance online publication. https://doi.org/10.1037/emo0000851

Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A., & Somerville, L. H. (2017). Increasing verbal knowledge mediates development of multidimensional emotion representations. *Nature Human Behaviour*, *1*(12), 881–889. https://doi.org/10.1038/s41562-017-0238-7

Nook, E. C., & Somerville, L. H. (2019). Emotion concept development from childhood to adulthood. In *Nebraska Symposium on Motivation* (Vol. 66, pp. 11–41). Springer. https://doi.org/10.1007/978-3-030-27473-3\_2

Nook, E. C., Stavish, C. M., Sasse, S. F., Lambert, H. K., Mair, P., McLaughlin, K. A., & Somerville, L. H. (2020). Charting the development of emotion comprehension and abstraction from childhood to adulthood using observer-rated and linguistic measures. *Emotion*, *20*(5), 773–792. https://doi.org/10.1037/emo0000609

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... & Solymos, P. (2019). vegan: Community Ecology Package. R package version 2.5-6. 2019.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods.* 10.3758/s13428-018-01193-y

Pollak, S. D., Camras, L. A., & Cole, P. M. (2019). Progress in understanding the emergence of human emotion. *Developmental Psychology*, *55*(9), 1801–1811. https://doi.org/10.1037/dev0000789

R Development Core Team. (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Richie, R., White, B., Bhatia, S., & Hout, M. C. (2020). The spatial arrangement method of

measuring similarity can capture high-dimensional semantic structures. *Behavior*

*Research Methods*, 1-23.

Ruba, A. L., & Pollak, S. D. (2020). The development of emotion reasoning in infancy and early childhood. *Annual Review of Developmental Psychology*, *2*, 22.1-22.29. https://doi.org/10.1146/annurev-devpsych-060320-102556

Russell, J. A. (2017). Mixed Emotions Viewed from the Psychological Constructionist Perspective. *Emotion Review*, *9*(2), 111–117. https://doi.org/10.1177/1754073916639658

Russell, J. A., & Barrett, L. F. (1999). *Core Affect , Prototypical Emotional Episodes , and Other Things Called Emotion : Dissecting the Elephant*. *76*(5).

Srinivasan, R., & Martinez, A. M. (2018). Cross-Cultural and Cultural-Specific Production and Perception of Facial Expressions of Emotion in the Wild. *IEEE Transactions on Affective Computing*. https://doi.org/10.1109/TAFFC.2018.2887267

Unger, L., & Fisher, A. V. (2019). Rapid, experience-related changes in the organization of children’s semantic knowledge. *Journal of Experimental Child Psychology*, *179*, 1–22. https://doi.org/10.1016/j.jecp.2018.10.007

Unger, L., Fisher, A. V., Nugent, R., Ventura, S. L., & MacLellan, C. J. (2016). Developmental changes in semantic knowledge organization. *Journal of Experimental Child Psychology*, *146*, 202–222. https://doi.org/10.1016/j.jecp.2016.01.005

Vales, C., Stevens, P., & Fisher, A. V. (2020). Lumping and splitting: Developmental changes in the structure of children’s semantic networks. *Journal of Experimental Child Psychology*, *199*, 104914. https://doi.org/10.1016/j.jecp.2020.104914

Walden, T. A., & Ogan, T. A. (1988). The Development of Social Referencing. *Child Development*, *59*(5), 1230. https://doi.org/10.2307/1130486

Walle, E. A., Reschke, P. J., & Knothe, J. M. (2017). Social Referencing: Defining and Delineating a Basic Process of Emotion. *Emotion Review*, *9*(3), 245–252. https://doi.org/10.1177/1754073916669594

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Watson, D., Wiese, D., Vaidya, J., Tellegen, A., Watson, D., Tellegen, A., & David, W. (1999). The Two General Activation Systems of Affect: Structural Findings, Evolutionary Considerations, and Psychobiological Evidence The Affect Circumplex Historical Overview. In *Journal ol’ Personality and Social Psychology* (Vol. 76, Issue 5).

Widen, S. C. (2013). Children’s Interpretation of Facial Expressions: The Long Path from Valence-Based to Specific Discrete Categories. *Emotion Review*, *5*(1), 72–77. https://doi.org/10.1177/1754073912451492

Woodard, K., Plate, R. C., Morningstar, M., Wood, A., & Pollak, S.D. (in press).

Categorization of vocal emotion cues depends on distributions of input. *Affective Science*.

1. Development of the Interdisciplinary Affective Science Laboratory (IASLab) Face Set was supported by the National Institutes of Health Director’s Pioneer Award (DP1OD003312) to Lisa Feldman Barrett. Gendron, M., Lindquist, K. A., & Barrett, L. F. (unpublished data). More information available online at <https://www.affective-science.org/face-set.shtml>. [↑](#footnote-ref-1)
2. If analyses are limited to include only the most basic emotion categories (happy, sad, anger, disgust, fear, and sad), the same pattern of results are found (see Supplemental Material, section 2, “Basic Emotions”). [↑](#footnote-ref-2)