



Published in final edited form as:

*Child Dev.* 2023 May ; 94(3): 585–602. doi:10.1111/cdev.13897.

## Caregiver speech predicts the emergence of children’s emotion vocabulary

Mira L. Nencheva,

Diana I. Tamir,

Casey Lew-Williams

Princeton University, Department of Psychology

### Abstract

Learning about emotions is an important part of children’s social and communicative development. How does children’s emotion-related vocabulary emerge over development? How may emotion-related information in caregiver input support learning of emotion labels and other emotion-related words? This investigation examined language production and input among English-speaking toddlers (16–30 months) using two datasets: Wordbank ( $N=5520$ ; 36% Female, 38% Male, 26% unknown gender; 1% Asian, 4% Black, 2% Hispanic, 40% White, 2% Other, 50% unknown ethnicity; collected in North America; dates of data collection unknown) and CHILDES ( $N=587$ ; 46% Female, 44% Male, 9% unknown gender, all unknown ethnicity; collected in North America and the United Kingdom; data collection dates, where available were between 1962 and 2009). First, we show that toddlers develop the vocabulary to express increasingly wide ranges of emotional information during the first two years of life. Computational measures of word valence showed that emotion labels are embedded in a rich network of words with related valence. Second, we show that caregivers leverage these semantic connections in ways that may scaffold children’s learning of emotion and mental state labels. This research suggests that young children use the dynamics of language input to construct emotion word meanings, and provides new techniques for defining the quality of infant-directed speech.

---

From infancy through adulthood, humans experience and communicate rich and dynamic mental experiences. Words can capture these experiences, allowing us to peer into others’ minds and share our own internal experiences with others. Children’s ability to verbally label their emotions is an important force in their socio-emotional development (Hoemann et al., 2019). Children who can label their emotions have more effective emotion regulation, more peer integration in kindergarten, and greater socio-emotional school readiness (Cole et al., 2010; Eisenberg et al., 2005; Fabes et al., 2001; Roben et al., 2013; Trentacosta & Izard, 2007). How do children learn these important labels?

Children’s earliest social interactions include information about emotion experiences and emotion language simultaneously. For example, as a child observes a facial expression associated with happiness – a smile, or a positive disposition – they may hear *emotion labels*, such as “happy”, which directly name an internal emotional state. These labels may occur alongside *valenced words*, i.e., positive, or negative words such as “smile” and “good,” which contain emotional information but do not explicitly name the experienced

internal state. That is, a natural conceptual structure emerges over time because emotion experiences co-occur with valenced words and emotion labels.

Young children have a propensity to track the statistics of their input, allowing them to group co-occurring inputs together into emotion concepts. Caregivers may help scaffold children's learning about emotion and language simultaneously. In the current investigation, we explored how children develop emotion labels, and how this relates to caregivers' dynamic use of valenced language.

## Emotion labels in context

Emotion labels are a subset of valenced words that refer to a highly positive or negative internal mental state, such as "happy" or "sad". Cabanac (2002) defines emotions as "any mental experience with high intensity and high hedonic content (pleasure/displeasure)." They are on average less concrete than early-learned neutral words (e.g., "truck"; Vigliocco et al., 2014), meaning that it is hard to highlight a concrete object, action, or sensation in the physical environment of the child as the referent of the emotion label. One cannot point to, look at, or hand "happiness" or "sadness" to a child. In order to learn emotion labels, speakers have to abstract across variable physical contexts (Barrett et al., 2011; Hoemann et al., 2019, 2020). For example, a person may experience joy when their favorite soccer player scores a goal or when the villain in a cartoon slips on a banana peel. Speakers also need to learn that multiple emotion labels can apply to the same physical context. For example, tears can signal joy or grief. Or, the arrival of a visitor may evoke excitement or dread. Despite this variability, developing abstract concepts is core to human cognition (Borghi, 2022).

The variability surrounding emotion labels poses a challenge for learning this important class of words. Indeed, infants learn words later if they lack a concrete referent (Bergelson & Swingley, 2013). However, when comparing across different words that similarly lack a concrete referent, emotionally charged abstract words, such as emotion labels, are learned earlier than abstract words that are more neutral (e.g., "premise" or "status"; Ponari et al., 2018, 2020). The emotional content of these abstract words, specifically how positive or negative they are, may bootstrap children's learning. Valence – the relative positivity or negativity of an experience – is the dimension that most strongly organizes adult emotion representations (Russell, 1980; Tamir et al., 2016; Thornton & Tamir, 2020) and is present in young children (Nook et al., 2017). Early childhood may be an especially important time for young children's development of valence-based representations. Valence gradually emerges in the first 5 years of life as a dimension that organizes children's representations of non-linguistic emotion cues and emotional experiences, in tandem with children's expanding vocabularies (Nencheva et al., 2021; Woodard, Zettersten & Pollak, 2021). Children's ability to label emotions with specificity emerges during this same span in early childhood (Widen, 2013; Wu et al., 2022). What allows young children to extract information about the valence of these words?

One important source of information that may support children's learning of emotion labels is the context in which these words occur (Thornton et al., 2020). Emotions are more likely to precede and follow other similarly positive or similarly negative emotions (Thornton &

Tamir, 2017). For example, if someone experiences a negative emotion (e.g., sadness), they are more likely to experience another negative emotion next (anger) rather than a positive emotion (happiness). When adults observe sequences of novel emotion labels (Thornton et al., 2020), they later judge emotion labels as more similar if they had consistently preceded or followed each other, suggesting that context offers a useful cue for the learning of concepts. Children's early input likely reflects this valence-consistent context for emotion experiences. This context may serve as a key source of information during learning. When children observe that a new emotion co-occurs with another positive emotion, or within a positive context, this offers key information about the meaning of the new emotion. Infants may likewise track how often emotion words co-occur in order to construct the meaning of a particular emotion concept (Shablack et al., 2020). That is, nearby emotion labels and other related words may cue the valence of a novel emotion label. In order for young children to learn these relations, they must simultaneously track statistical information across time and form semantic connections between words that occur close to each other in time.

### Using statistical information to learn novel labels

There is considerable evidence that young children track the statistics of temporal sequences. How may this domain-general mechanism support children's learning of emotion labels? Infants implicitly track syllables, words, objects, and events in their environment and detect when their statistics are altered (Lew-Williams et al., 2011; Saffran et al., 1996; Saffran & Kirkham, 2018). Infants are sensitive to how stimuli in their environment are distributed across time, a skill necessary for learning the semantic relations between words, and more specifically, for linking emotion labels with their valenced contexts. Research has already shown that infants do this for non-valenced terms. For example, young children can infer what a label refers to based on information in surrounding sentences (Horowitz & Frank, 2015; Schwab & Lew-Williams, 2020). By two years of age, toddlers reliably infer that two words are similar if both co-occur with the same related words (Lany & Saffran, 2010; Wojcik & Saffran, 2015). When semantically related words precede a label, they guide young learners' attention and facilitate processing of the label (Peters et al., 2021). These findings offer further evidence that toddlers should be able to extract the valence of a novel emotion label from the valence of its context.

Research on children's semantic networks shows that children's vocabulary composition is associated with their ability to use semantic context to learn new words. That is, children are more likely to learn novel labels that are (vs. are not) semantically related to words they already know (Hills et al., 2009). Young children are also faster to recognize a novel label that belongs to a category that is well (vs. poorly) represented in their vocabulary (Borovsky et al., 2016). These findings provide further reason to believe that children may be more able to capitalize on valence cues in the context of a novel emotion label if they already know many positive or negative words. This valenced knowledge should bootstrap a child's emotion label learning, compared to children who mostly know neutral words.

There is evidence that caregivers may adapt their speech in ways that scaffold young children's ability to track the co-occurrences of related words. When addressing young children, caregivers immediately precede and follow a given word with related words

more so than when they are addressing adults (Hills, 2013). This adaptation may support learning, as children learn novel labels better when they are surrounded by related rather than unrelated discourse (Fernald et al., 2009; Schwab & Lew-Williams, 2017, 2020). In the domain of emotion words, this suggests that caregivers could likewise provide context for emotion labels that support children's learning of these important words. Caregivers may curate similarly valenced words around the introduction of unfamiliar emotion labels. For example, when introducing the label "happy", a caregiver can provide information about the situation or actions surrounding the emotion (e.g., "Rosa got a wonderful present for her birthday! She was so happy!"). By introducing other positive words concurrently, such as "wonderful", "present", and "birthday", caregivers can cue children to cluster these concepts together, and to think about them separately from concepts that rarely show up concurrently, like "yucky", or "sad". The emotion labels can then take on meaning, as young children cue into the valence (i.e., the positivity vs. negativity) of the cluster, and thus the valence of the emotion label "happy". Therefore, caregiver cues to the valence of an emotion label in the discourse surrounding the labeling event may support learning both in the moment (by helping children infer the meaning of the novel emotion label) as well as over the course of development (by building up a network of semantic similarity between the emotion label and other valenced words).

## Current investigation

In the current investigation, we explore (1) the emergence of valenced words in children's production, and (2) how caregiver discourse may support young children's learning of emotion labels.

In Study 1, we characterize the development of 1- to 2-year-old children's semantic network of valenced words and examine the rate at which young children learn emotional and neutral words. We hypothesized that children's early vocabularies will consist primarily of concrete neutral words, and would later expand to include more abstract positive and negative words. We tested this using data from Wordbank, a large database of the productive vocabularies of young children (Fenson, 2007; Frank et al., 2017). As children learn more valenced words, they should be able to use this rich semantic network to bootstrap their understanding of new emotion labels and vice versa. We test this possibility in Study 2, using the same database, hypothesizing that knowing highly emotional words makes it easier for children to learn emotion labels, and vice versa.

Next, we test how children form these semantic connections between emotion labels and other valenced words. Specifically, in Studies 3, 4 and 5, we examined how the structure of caregiver input may leverage the valenced, semantic connections to facilitate emotion label learning. If the co-occurrence of similarly-valenced words bootstraps emotion label learning, then young children should learn emotion labels best when caregivers provide useful context about the valence of the labels. In Study 3, we hypothesized that caregivers use emotion labels within contexts that match in valence. We characterized the temporal dynamics of sentiment in caregiver speech surrounding emotion labels in English corpora (including North American, British) from the Child Language Data Exchange System (MacWhinney, 2000, 2014; Sanchez et al., 2019). In Studies 4 and 5, we provide correlational evidence

that co-occurrence between emotion labels and similarly-valenced words in children's input may support children's production of emotion labels. Study 4 examined whether variability in the extent to which different emotion labels lend themselves to such co-occurrence in child-directed speech predicts earlier or later production. Study 5 then examined the longitudinal hypothesis that children produce emotion labels in more accurate contexts when their caregivers surround emotion labels with similarly valenced words. Together, the five studies provide insight into how young children may use dynamic language input to construct complex meanings.

Analysis code and materials for all studies are available at: [https://osf.io/n689q/?view\\_only=f8cbc3ad7281495da0ff02be935c7e6c](https://osf.io/n689q/?view_only=f8cbc3ad7281495da0ff02be935c7e6c)

## STUDY 1: The development of children's emotion vocabulary

How does the representation of valence in young children's vocabularies develop over time? In Study 1, we characterized the development of young children's semantic networks of emotion-related words. Our exploratory analyses arbitrate between two possibilities. On the one hand, young children may learn new words uniformly across the valence continuum by learning emotional and neutral words simultaneously. Alternatively, young children may learn some parts of this emotional continuum before others.

Our predictions are grounded in findings from three prior studies that we aimed to replicate in our dataset; namely, that children learn abstract words later than concrete words (Bergelson & Swingley, 2013), that valenced words are, on average, less concrete than neutral words (Vigliocco et al., 2014), and that older children learn highly positive or negative abstract words (e.g., "delight" or "tyranny") earlier than relatively neutral abstract words (e.g., "scheme" or "premise"; Ponari et al., 2018, 2020)

Based on these findings, we predict that (1) toddlers would first learn 'neutral' concrete words and then learn the relatively abstract valenced words. Toddlers should learn neutral abstract words even later (2) this would result in a gradually increasing emotional range in children's vocabularies over the first few years of life. The present study served to motivate our main research questions addressed in Studies 2–5.

### Method

**Dataset**—We tested these predictions on the vocabularies of 5,520 toddlers between the ages of 16 and 30 months using parent-report data from the MacArthur-Bates Communicative Development Inventory MCDI (Fenson, 2007), Wordbank version 0.3.0 (Frank et al., 2017). This measure asks caregivers to report which of a set of 680 words their child understands and says.

Of the toddlers included in the Wordbank database, 1,989 were female, 2,105 were male; 2202 identified as White, 67 as Asian, 222 as Black, 131 as Hispanic, and 93 as other. Gender information was missing for 1,426 of the participants and ethnicity information was missing for 2,805 of the participants. For the last analysis in Study 1, which computed the

emotional range of children's vocabularies, we excluded 9 participants who did not produce any words.

The dates of data collection were not included in the Wordbank database; however, a list of contributing publications and researchers can be found on the Wordbank website (<http://wordbank.stanford.edu/contributors>). Papers contributing to the English database range in publication dates between the years 2000 and 2016 and were collected in North America.

## Variables

**Word emotionality and valence:** We performed a sentiment analysis using the *afectr* package (Thornton & Tamir, 2020) to compute the valence of all words in the MCDI form (680 words total). This tool computes the valence of a word based on its similarity to other, previously rated, emotion labels (Tamir et al., 2016; Thornton & Tamir, 2017, 2020). This dimension of valence distinguishes positive from negative words (hereafter referred to as *valence*) and valenced words (i.e., positive and negative) from non-valenced words (i.e., neutral; hereafter referred to as *emotionality*, computed as the absolute value of valence). We verified the valence values extracted via sentiment analysis by comparing them to normed human valence ratings for 492 of the MCDI words (Warriner et al., 2013). We found a robust correlation between sentiment analysis estimates and human ratings ( $r = 0.62$ ,  $p = 1.87 \times 10^{-53}$ ), suggesting that the computational sentiment analysis yielded comparable values to human raters. Therefore, our analyses exclusively used the continuous measures of valence from this sentiment analysis. We also computed the emotionality of each word, giving high values to highly positive *and* highly negative words, and low values to relatively more neutral words (see Figure 1a). The values for valence and emotionality were z-scored in all statistical analyses.

**Emotional range:** We assessed the emotional range of a child's vocabulary by computing the standard deviation of the valence of all words produced by the child. By definition of this metric, children who produced words spanning a wider range of valenced values (covering positive and negative extremes of the valence continuum) had a wider emotional range than children who produced a narrower range of valenced values (for example, covering primarily the 'neutral' middle of this continuum). This measure was then z-scored for statistical analyses.

**Concreteness:** We obtained concreteness ratings from Köper & im Walde (2017) for the words in the MCDI (available for 560 of the 680 words).

**Word frequency:** In order to control for effects of the frequency of words in children's input, we computed the log frequency (number of occurrences of a target word per million words) of each word in CHILDES, a large database of transcribed caregiver-child interactions, using the *chldes-db* R package (Sanchez et al., 2019). For this estimate, we included all corpora of English speech by caregivers (including both mothers and fathers) addressing toddlers between the ages of 15 and 30 months (a total of 49 corpora of 522 children). Frequency in CHILDES could not be estimated for 53 of the words. These words were excluded only when reporting results controlling for frequency. There was a small

positive correlation between the emotionality of a word and its frequency in child input ( $r = 0.09$ ,  $p = 0.02$ ), suggesting that if highly emotional words are learned later, this is not because children hear them less often.

**Age of acquisition:** We computed the age at which at least 50% of children produced the word, also known as age of acquisition (AoA). To extract this number, we used the `fit_aoa` function in `wordbankr` (Braginsky et al., 2018; Frank et al., 2017) to fit a logistic curve over the proportion of children producing each word over time. In this analysis, 69 words were excluded because they did not reach a threshold of 50% of children producing the word within the study age range. All analyses of age of acquisition were performed on the remaining 608 words of the MCDI.

A table with the analyzed words and their age of acquisition, valence, concreteness, and frequency can be found on the project [osf page](#).

**Data analysis**—For all analyses in Study 1, we used linear regression predicting the age of acquisition of the word from its emotionality. Standardized regression coefficients ( $\beta$ ) were estimated using the `lm.beta` function (Behrendt, 2014). All analyses were performed in R, version 4.1.3 (R Core Team, 2022).

## Results and Discussion

**Emotionality shapes toddlers' early word productions**—We first explored whether the age at which toddlers produce a word is related to the word's emotionality, and tested our prediction that toddlers would produce valenced words later than neutral words. This regression showed a small but significant positive effect of emotionality on age of acquisition (AoA ~ emotionality;  $\beta_{emotionality} = 0.08$ ,  $t(606) = 2.06$ ,  $p = 0.04$ ), such that the more extremely positive or negative a word was, the later toddlers produced it (see Figure 1b). This effect held when controlling for word frequency (AoA ~ emotionality + frequency;  $\beta_{emotionality} \sim 0.1$ ,  $t(560) = 2.37$ ,  $p = 0.02$ ), and after removing the most influential data points (Supplementary Analysis 1) and the most extremely positive and negative words (Supplementary Analysis 2). We did not find an overall effect of valence (i.e., positive words were learned around the same time as negative words), replicating prior work from Braginsky et al. (2016). The effect of emotionality did not interact with that of valence (Supplementary Analysis 3).

**Emotionality and concreteness: A replication of prior findings**—To explore why toddlers in this dataset produced valenced words later, we replicated several findings in prior literature showing an interplay of valence and abstractness in children's word learning. First, using concreteness ratings from Köper & im Walde (2017), we replicated a significant negative association between the emotionality and concreteness of words in the MCDI (concreteness ~ emotionality;  $\beta_{emotionality} \sim -0.41$ ,  $t(558) = -10.6$ ,  $p = 5.02 \times 10^{-24}$ ), previously shown in Vigliocco et al. (2014). In short: the more emotional a word is, the less concrete it is. Second, we replicated the result from Bergelson and Swingley (2013) that children learn concrete words earlier than abstract words (AoA ~ concreteness;  $\beta_{concreteness} = -0.3$ ,  $t(561) = -7.49$ ,  $p = 2.74 \times 10^{-13}$ ).

Next, we performed exploratory analyses to probe whether concreteness explains the link between emotionality and later AoA. When controlling for concreteness, the effect of emotionality on age of acquisition was no longer significant (AoA ~ emotionality + concreteness;  $\beta_{\text{emotionality}} \sim -0.05$ ,  $t(557) = -1.25$ ,  $p = 0.21$ ). This result – along with the findings that emotionality predicts concreteness and that concreteness predicts AoA – suggests that in this sample, concreteness mediates the effect of emotionality on AoA. That is, toddlers are likely to learn highly emotional words later than neutral words because they are less concrete. However, it is important to note that the current study was not designed to probe the nuanced interplay of emotionality and concreteness. Future investigations that are specifically designed to hold concreteness constant and to vary emotionality (e.g., by creating matched pairs, as in Ponari et al., 2018, 2020) will enable better understanding of these effects in toddlers.

Ponari et al. (2018, 2020) showed that older children learn highly positive and negative abstract words earlier than neutral abstract words. We also replicated this finding in our analyses. As an example, we found that the highly positive word “smile” has an AoA of 27 months, whereas the frequency-matched relatively neutral concrete words “tongue” and “bowl” have an earlier AoA of 23 months. In contrast, abstract neutral words like “pretend” and “think” have a later AoA of 30 months. There was variability in the size of this effect for different words; for example, “bucket”, another frequency-matched relatively neutral word, has an age of acquisition of 26 months. However, these results suggest that children, on average, (1) hear highly emotional words just as frequently as more neutral words, (2) learn highly valenced abstract words later than concrete neutral words, and (3) learn highly valenced words earlier than neutral abstract words (Supplementary Analysis 5).

**The emotional range of children’s vocabularies expands over time**—Regardless of the specific mechanism that results in toddlers’ later production of highly valenced words, we hypothesized that this trend would shape children’s growing vocabularies over the first few years of life. We examined how the range of positive and negative words in toddlers’ vocabularies changes over time. How do the learning trajectories of differently valenced words affect the overall range of valenced words that children produce? If toddlers learn highly positive and negative abstract words later than relatively neutral concrete words, and if they do not yet produce many neutral abstract words, then the emotional range of their vocabularies should also increase over the first few years of life. In this analysis, we assessed how the emotional range of children’s vocabularies changes over time in a linear regression predicting emotional range from age, controlling for vocabulary size. Indeed, emotional range increased with age (valence range ~ age;  $\beta_{\text{age}} \sim 0.46$ ,  $t(5498) = 37.96$ ,  $p = 2.78 \times 10^{-280}$ ). For example, at 16 months, toddlers’ vocabularies had an average emotion range of 0.15 (valence: mean = 0.04, min = -0.33, max = 0.41; Figure 1c), whereas at 30 months, they had an average range of 0.18 (valence: mean = 0.04, min = -0.64, max = 0.82; Figure 1c). The increase in emotion range remained significant when controlling for vocabulary size (valence range ~ age + vocabulary size  $\beta_{\text{age}} = 0.14$ ,  $t(5497) = 8.07$ ,  $p = 8.35 \times 10^{-16}$ ), suggesting that the expansion in emotion range is not merely a byproduct of larger vocabularies in older children, but rather are the result of specific expansions in toddlers’ productions of highly positive and negative words (Supplementary Analysis 5).



Together, these findings support the hypothesis that children's early semantic networks start off relatively sparse in terms of representing highly positive or negative words, and instead are more densely populated by neutral concrete words. Over time, as children learn more valenced words, their semantic networks represent an increasingly wide range of positive and negative words. This developmental change may have important consequences for children's ability to learn and express emotion-related information. In Study 2, we explored how the increasing emotional range of children's semantic networks interacts with their production of specific emotion labels.

## STUDY 2: The structure of children's emotion vocabulary

In Study 1, we characterized how young children's semantic networks include an increasingly wide emotional range between 16 and 30 months of age. In Study 2, we hypothesized that knowing valenced words makes it easier for children to learn emotion labels and vice versa. For example, knowing the words like "smile" and "nice" might bootstrap learning of labels like "happy"; similarly, knowing an emotion label like "sad" could bootstrap learning of related valenced words like "bad" and "cry". That is, there should be a relation between the network of valenced words that a child produces and the child's production of emotion labels. Here we tested this possibility. We predicted that children would be more likely to produce a specific emotion label if they know other similarly-valenced words. In an exploratory analysis, we examined two forms of similarity between an emotion label and words in the child's vocabulary: one looking at overall emotionality, and another assessing overall positivity vs. negativity (valence).

### Method

**Dataset**—In Study 2, we reanalyzed the vocabularies of 5,520 toddlers between the ages of 16 and 30 months from the MCDI (Fenson, 2007) in Wordbank version 0.3.0 (Frank et al., 2017).

#### Variables

**Emotion label production:** In order to assess emotion labels, we noted which (if any) of the eight emotion labels included in the MCDI ("hate", "love", "happy", "hurt", "mad", "sad", "scared", "tired") each child produced. The criterion for counting a word from the MCDI as an emotion label was that it had to refer to an internal mental state or experience that was highly valenced (positive or negative; Cabanac, 2002).

**Quantifying semantic connections:** We measured connections between each emotion label with all other words produced by the child as the similarity in valence or emotionality between the words. Similarity was computed as the difference in these measures between the emotion label and the word (scaled between 0 and 1), subtracted from 1. As a summary measure, for each emotion label, we computed its *strength centrality*, a measure commonly used to assess how strongly connected a given element of a network is to all other elements of that network (Candeloro et al., 2016). We did so by averaging the weight of the connections between the emotion label and each word produced by the child.

We quantified the strength of the semantic connection between a given emotion label and the remaining words in the child's productive vocabulary in terms of both valence and emotionality (for an example, see Figure 2a). The valence component quantifies similarity between two words based on the relative positivity or negativity of these words. This metric groups positive words with other positive words and negative words with negative words. For example, positive labels (e.g., "happy") have strong connections with other positive words (e.g., "smile") and negative labels (e.g., "sad") have strong connections with other negative words (e.g., "cry"). In contrast, the label would have weak connections with words with oppositely signed valence from the label (e.g., the connection between the negative label "sad" and the positive word "smile").

The emotionality component quantifies similarity between two words based on how extreme the valence of the words is, regardless of whether they are positive or negative. This metric groups both highly positive and negative words together as distinct from neutral words. For example, a positive emotion label (e.g., "happy") would share a strong connection with both positive and negative words (e.g., "smile" and "cry"), but a weak connection with relatively neutral words (e.g., "spoon").

For each emotion label and for each child, we computed how strongly the emotion label was connected to other words in the child's vocabulary for each of these two components, using the method described above (for an example see Figure 2a). Because in some cases the two strength centrality measures (emotionality and valence) were collinear, we included each in a separate model.

**Data analysis**—In Study 2, we used mixed-effects logistic regressions to predict whether each child produces each of the 8 emotion labels (Bates et al., 2015). For all models, we included random intercepts and effects by participant for the variable of interest (the average connection strengths of the labels, measured as strength centrality as described above). Additionally, we controlled for the child's vocabulary size and the label frequency. Full model formulae are included in the results section.

## Results and Discussion

**Semantic connections predict production of specific emotion labels**—We tested the prediction that children are more likely to produce a specific emotion label when they already know similarly valenced words. In this analysis, we quantify the connections between a specific label and other words in the child's vocabulary to predict production of individual labels. In a mixed-effect logistic regression, we predicted whether the child produces the specific emotion label depending on how strongly the label was connected to the rest of the child's vocabulary (its strength centrality). We controlled for vocabulary size and label frequency, and included random intercepts per child.

As expected, we found that children were more likely to produce emotion labels that were better connected in terms of emotionality (production of label (yes / no) ~ emotionality strength centrality + vocabulary size + label frequency + label concreteness + (1+emotionality strength centrality | participant);  $\beta = 0.8$ ,  $z = 21.95$ ,  $p = 9.42 \times 10^{-107}$ ). That is, children were more likely to produce an emotion label if they already knew words

with a similar degree of emotionality (both highly positive and negative words). Similarly, children were more likely to produce emotion labels with strong connections in terms of valence (production of label (yes / no) ~ valence strength centrality + vocabulary size + label frequency + label concreteness + (1+valence strength centrality | participant);  $\beta = 1.07$ ,  $z = 29.84$ ,  $p = 1.34 \times 10^{-195}$ ). That is, children were more likely to produce an emotion label if they already produced words that were similarly positive or negative (see Figure 2b). This analysis leaves open the possibility of a mutually reinforcing link between the production of emotion labels and other valenced words.

### STUDY 3: Caregiver cues to networks of emotion words

In Study 2, we observed that children are more likely to produce an emotion label when it is embedded in a rich network of known valenced words. How do children learn these important semantic connections between emotion labels and other valenced words? In Study 3, we hypothesized that caregivers may build these connections for their children through their language input. Specifically, we expected that when caregivers produced an emotion label, they would embed it within sentences that include other similarly valenced words, such that emotion labels and valenced words co-occur in time. We probed this question in a preliminary exploratory analysis of the 8 emotion labels used in Study 2, and a preregistered analysis of 94 mental state labels.

#### Method

**Dataset**—Our goal was to measure how parents contextualize emotion labels in real parent-child interactions. We thus examined a dataset that includes transcripts of parent-child conversations. We analyzed caregiver-child interactions from 2,936 CHILDES transcripts in English (including North American English, British English), containing over one million caregiver utterances. The dates of data collection were only available for ~60% of transcribed interactions. Dated transcribed interactions were collected between 1962 and 2009.

The transcripts contained speech from the caregivers of 587 children between the ages of 16 and 30 months, of which 292 were female, 261 male. Gender information was missing for 54 of the children, and the dataset did not include ethnicity information for any of the participants. Analyses were conducted using the *childevs* R package (Sanchez et al., 2019).

Only a subset of the transcribed interactions contained the mental state labels of interest. For the preliminary analysis of 8 emotion labels, there were 1,522 transcribed interactions of caregivers addressing 266 children. This resulted in 7,972 utterances containing emotion labels and 156,678 surrounding utterances. For the preregistered analysis of 94 additional mental state labels, there were 1,675 transcripts of caregivers addressing 366 children. This included 8,834 utterances containing this set of emotion labels and 174,094 surrounding utterances.

**Emotion labels**—In a preliminary analysis, we identified parent (both mothers' and fathers') utterances containing any of the eight emotion labels used in Study 2 (“hate”,

“love”, “happy”, “hurt”, “mad”, “sad”, “scared”, “tired”). Utterances containing negation (272 utterances) or multiple types of emotion labels (e.g., both a positive and a negative label; 59 utterances) were excluded. In a second preregistered analysis, we analyzed caregivers’ labeling of 94 mental states. These states followed a broader definition of any word labeling an internal mental experience, and were selected from a list of 166 validated mental states from Tamir et al. (2016). We excluded any states that did not occur in caregiver speech or that were only used in a different sense that did not label a mental state (e.g., “warm” and “gloomy” can be used to refer to temperature or weather). Because these labels were not specifically selected to reflect the words present in toddlers’ early communicative interactions, they were less frequent than the labels in the first analysis. For this study, we preregistered a primary analysis for the mental states that were rated the highest on emotion (normed rating of at least 0.75 out of 1,  $N = 47$  labels) from Tamir et al. (2016), as well as an exploratory analysis for the full set of 94 labels. The results between the two sets of states were similar, but stronger for the larger set of 94 labels. The preregistration can be found here: [https://aspredicted.org/Y7J\\_MT4](https://aspredicted.org/Y7J_MT4), and the list of states can be found on the project [osf page](#).

**Variables**—We quantified the valence of caregivers’ words surrounding an emotion label in two ways.

**Valenced context of preceding and following utterances:** First, we analyzed the utterances surrounding an utterance with an emotion label. For each labeling instance, we computed the valenced sentiment of the 10 preceding and 10 following utterances, using the *affectr* package (Thornton & Tamir, 2020). If an utterance was fewer than 10 utterances away from more than one emotion label, we counted the distance to the closest label only.

**Sentence valenced context:** Second, we assessed the narrow context of the immediate sentence in which the emotion label was embedded. We computed the valenced sentiment of the sentence frame, excluding the label itself, using the *affectr* package (Thornton & Tamir, 2020).

**Data analysis**—All analyses in Study 3 were carried out as mixed-effect linear regressions (Bates et al., 2015). For all models, we started with random intercepts and effects by speaker. However, for the analysis of the valenced context of surrounding utterances, this model did not converge for all except one of the four regressions. Therefore, for this analysis, we only included random intercepts, but not effects by speaker. In the analysis examining valenced context in the labeling sentence, the model with random intercepts and effects by speaker converged. Model formulae are included in the results section below.

## Results and Discussion

**Caregivers precede and follow emotion labels with valenced utterances**—We first characterized the extent to which caregivers provide appropriately valenced context for emotion labels. To do so, we assessed how the valence of caregiver speech changed over the 10 utterances preceding and following their use of an emotion label. A mixed-effect regression predicted the valence of each utterance from its distance to the closest emotion

label. The model included random intercepts by speaker. We constructed two separate models: one for positive emotion labels, one for negative emotion labels.

In the preliminary analysis of 8 emotion labels, we found that caregivers used more positive words as utterances got closer to a positive emotion label (valence of utterance ~ distance of utterance from label + label concreteness + (1 | speaker);  $\beta_{\text{distance}} = -0.1$ ,  $t(65,008.5) = -21.28$ ,  $p = 3.99 \times 10^{-100}$ ), and similarly, that caregivers used more negative words as utterances got closer to a negative emotion label (valence of utterance ~ distance of utterance from label + label concreteness + (1 | speaker);  $\beta_{\text{distance}} = 0.07$ ,  $t(87,835.01) = 17.4$ ,  $p = 9.33 \times 10^{-68}$ ; see the top panel of Figure 3a).

These results were replicated in our preregistered analysis of 94 mental state labels (see the bottom panel of Figure 3a). That is, more positive caregivers' utterances appeared more closely to positive mental state labels (valence of utterance ~ distance of utterance from label + label concreteness + (1 | speaker);  $\beta_{\text{distance}} = -0.01$ ,  $t(112,403.1) = -3.55$ ,  $p = 0.0004$ ), and similarly, more negative caregivers' utterances appeared more closely to negative mental state labels (valence of utterance ~ distance of utterance from label + label concreteness + (1 | speaker);  $\beta = 0.05$ ,  $t(58,092.69) = 9.5$ ,  $p = 2.10 \times 10^{-21}$ ). The results were similar for the smaller preregistered subset of highly emotional mental states (rated over 0.75 out of 1 on emotion), both for positive labels ( $\beta = -0.03$ ,  $t(24,803.19) = -3.61$ ,  $p = 0.0003$ ), and negative labels ( $\beta = 0.06$ ,  $t(17,979.91) = 7.35$ ,  $p = 2.03 \times 10^{-13}$ ).

This increase in valenced context surrounding mental state labels was strongest in the utterances closest to the label (Supplementary Figure 3). For example, ten utterances away from an emotion label (e.g., "happy" or "sad"), parents' utterances were relatively neutral (e.g., "He's coming to get Carl" or "What's he doing?"). But the utterances closer to a positive emotion label became increasingly positive (e.g., "Well the Hokey Cokey that's your favorite isn't it?" [5 utterances away], "Because I wanna get him dancing ya know!" [3 utterances away], "Can I have a big kiss?" [1 utterance away]). Similarly, the utterances closer to a negative emotion label became increasingly negative (e.g., "What did you hurt?" [5 utterances away], "Don't do what you're doing" [3 utterances away], "You're gonna break it" [1 utterance away]). These converging results suggest that caregivers provide valenced context in the utterances preceding and following the production of emotion labels or of mental state labels more broadly.

**Caregivers embed emotion labels within valenced sentences**—Next, we probed whether caregivers provide similarly valenced words in the same sentence as emotion labels. Specifically, we tested whether the valence of the emotion label predicted the sentiment of the sentence in which the label was embedded. We evaluated this prediction in a mixed-effect regression with a fixed effect of the valence of the label and random intercepts by transcript. As expected, the valence of the emotion label predicted the sentiment of the sentence in which the label was embedded both in the 8-label preliminary analysis (valence of sentence ~ valence of label + label concreteness + (1 + valence of label | speaker);  $\beta = 0.23$ ,  $t(193.08) = 5.97$ ,  $p = 1.1 \times 10^{-8}$ ; see Figure 3b top panel) and the 94-state preregistered analysis (valence of sentence ~ valence of label + label concreteness + (1 + valence of label | speaker);  $\beta = 0.21$ ,  $t(163.17) = 9.34$ ,  $p = 7.1 \times 10^{-17}$ ; see Figure 3 bottom panel). Caregivers

introduced negative emotion labels like “mad”, “scared”, and “sad” within sentences that contained other negative words like “mess”, “monster”, and “crying” (e.g., “Hey don’t make a mess because you’re mad”, “I think they’re scared of the monster”, “What are you sad too, are you crying?”). Similarly, caregivers labeled positive emotions, like “happy”, within sentences containing other positive words, like “kiss” and “smiling” (e.g., “Somebody must have given him a kiss there because he’s happy now” and “These look like happy pins - that guy’s laughing, that guy’s smiling”).

Together, these analyses highlight that caregivers surround emotion labels with similarly valenced words, both within the utterance containing the label as well as in the surrounding sentences. The temporal proximity of similarly valenced words may provide opportunities for children to link emotion labels and related words. By providing valenced contexts, caregivers may support the integration of emotion labels into a network of related words and concepts, which in turn may facilitate children’s learning. In Studies 4 and 5 we explore correlational links between caregiver input and child production.

## STUDY 4: Learning contextualized emotion labels

In Study 3, we observed that caregivers position emotion labels in contexts that are matched in valence. This temporal proximity between emotion labels and other similarly valenced words may provide opportunities for children to form semantic connections between them. If true, this could explain why children learn to produce emotion labels and similarly valenced words in tandem. However, not all emotion labels are contextualized to the same degree. Some emotion labels may more consistently appear in well-matched contexts, whereas others may not consistently appear in matched contexts. In Study 4, we examined how variability in the contexts surrounding different labels relates to the age when children produce them. An exploratory analysis examined whether emotion labels that consistently occur in valenced contexts in caregiver speech show an earlier age of acquisition.

### Method

**Dataset**—Study 4 analyzed the same set of caregiver utterances containing emotion labels from CHILDES as in Study 3, along with the utterances preceding and following the emotion label. Age of acquisition was assessed using the same data as in Studies 1 and 2.

### Measures

**Valenced context:** We quantified the valence of the utterances preceding and following each emotion label. Using the same method as Study 3, the valenced context of the surrounding utterances was quantified as the regression coefficient in a model estimating the utterance valence based on its proximity to an emotion label. We constructed a separate model to estimate the valenced context for each of the eight emotion labels used in Studies 2 and 3. We added a random intercept by speaker in the model estimating valenced context to ensure that the estimated context would not be driven by few outlier subjects. This valenced context estimate aggregated over all instances of the emotion label in the available data in CHILDES for caregivers addressing 16–30-month-old children.

**Age of acquisition:** We used the method outlined in Study 1 to estimate the age of acquisition for each emotion label based on child vocabularies in Wordbank (Braginsky et al., 2018; Frank et al., 2017). The emotion label “hate” was excluded from this analysis, because fewer than 50% of children produced it within the 16–30-month age range.

**Data Analysis—**All analyses in Study 4 were linear regressions, predicting AoA. Model formulae are included in the results section.

## Results and Discussion

We tested whether emotion labels that consistently occur in valenced contexts have an earlier age of acquisition. In a linear regression, we found that emotion labels that were consistently surrounded by valenced context in caregiver input in CHILDES had an earlier age of acquisition in Wordbank (AoA ~ valenced context;  $\beta_{valenced\ context} = -0.80$ ,  $t(5) = -3.01$ ,  $p = 0.03$ ; see Figure 4). This effect remained marginally significant when controlling for the label frequency and concreteness (AoA ~ valenced context + frequency + label concreteness;  $\beta_{valenced\ context} = -0.98$ ,  $t(3) = -3.25$ ,  $p = 0.048$ ). This result is consistent with a possible bidirectional link, such that related words surrounding an emotion label may support children in constructing the label’s complex meaning, and that early production of an emotion label may provide opportunities for introducing related valenced words in close proximity. However, because of the small number of emotion labels tested, it is difficult to disentangle this possibility from other confounding factors, such as the valence of the label itself (see Supplementary Analysis 9).

## STUDY 5: Longitudinal effect of caregivers’ valenced context on children’s emotion label production

Study 5 probed links between caregivers’ input and their child’s development of emotion labels. In this longitudinal analysis, we tested whether caregivers who surround emotion labels with more valenced contexts have children who later produce emotion labels with higher accuracy in appropriately valenced contexts. We examined this prediction in an exploratory analysis using the same 8 emotion labels that were used in previous studies, as well as a preregistered analysis of 87 mental state labels. By measuring speech within the same dyad, we can assess how caregivers’ input shapes children’s production of specific emotion labels over time.

### Method

**Dataset—**For this longitudinal analysis, we used a subset of corpora from CHILDES that included only parent-child dyads with multiple recorded interactions spanning at least two months of the child’s development. For each dyad, we defined two age ranges: one for evaluating initial caregiver input and another for evaluating later child production. Because recordings for each dyad spanned different ages and amounts of time (see Figure 5a), we split the total number of recorded interactions in two, based on the median age range across all transcripts for that dyad (Time 1 and Time 2).

We quantified caregiver input in the first part of the dyad's data (T1) as the valenced context surrounding emotion labels. We quantified the child's production in the second part of the dyad's data (T2) as an estimate of the child's appropriate production of emotion labels, defined specifically as the valenced context in which the child produced emotion labels. For example, if the dataset included 10 transcribed interactions for a given dyad at different ages, the first 5 interactions would be included in the parent input data, and the second 5 interactions would be included in the child production data. Interactions recorded at the median age for the dyad were included in the parent input estimate. In some cases, this process resulted in a slightly uneven split between the two parts of the dyad's data, as some dyads had multiple recorded interactions at the median age, and some recorded interactions did not include emotion labels. We excluded dyads if the caregiver did not produce emotion labels in the first part, or if the child did not produce emotion labels in the second part of the dyad's data.

For the preliminary analysis of 8 emotion labels, this resulted in a final sample of 35 children (19 female, 16 male, unknown ethnicity), with 2,199 parent and 1,212 child instances of these labels during T1 and T2, respectively. Because some of the 87 labels included in the preregistered analysis were not produced by young children (and did not appear on the MCDI), only 26 dyads had both parent and child productions of even one of these words. Thus, our analysis included 2,483 parent and 583 child instances of these labels during T1 and T2, respectively.

### Measures

**Valenced context:** As in Study 3, we characterized the valenced context in which caregivers and children used emotion labels in two ways: first, as the broad context of the 10 utterances preceding and following an emotion label, and second, as the narrow context of the sentence within which the label was embedded. For each dyad, we computed these two indices both in caregiver input and in child production.

To quantify the broad context of the preceding and following utterances, as in Study 3, we measured caregiver valenced context in the utterances preceding and following an emotion label. This measure was quantified as the regression coefficient in a model estimating the utterance valence based on its proximity to an emotion label. In Study 3, valenced context was similar for positive and negative labels, therefore we computed a single index for each speaker by averaging the valence coefficients across all positive and negative emotion labels.

As in Study 3, we quantified the narrow context of the sentence within which an emotion label was embedded. The resulting coefficient served as a valenced context index of how well the valence of the sentence matched the valence of the label within it.

**Data analysis**—In Study 5, we used linear regression to predict the valenced context surrounding children's mental state labels during T2 based on the valenced context of caregivers' productions during T1. In Supplementary Analysis 8, we additionally report concurrent links between caregiver input and child production within T1 and within T2. Model formulae are included in the results section below. The preregistration can be found



here: [https://aspredicted.org/G38\\_LRH](https://aspredicted.org/G38_LRH), and the list of mental state labels can be found on the project [osf page](#).

We controlled for the child's median age, as well as two measures of the child's language development during T2: the mean length of the child's utterances (MLU; Miller & Chapman, 1981), and the child's lexical diversity, measured as the average type-token ratio (TTR; Templin, 1957). We did so to isolate broader effects of language development, because children who produce many different words and longer utterances during T1 may be more likely to produce emotion labels and mental state labels in appropriate contexts during T2.

In addition, we controlled for the mean emotionality (i.e., absolute valence of emotion words, whether positive or negative) of caregiver utterances during T1 in order to distinguish between the effects of caregivers' overall use of valenced language from caregivers' use of appropriately valenced words in context (i.e., speech surrounding emotion labels). Similarly, we controlled for the number of instances of mental state labels in caregiver speech during T1. Additionally, we controlled for the average concreteness of the emotion and mental state labels used by the caregiver and the child in T1 and T2 respectively. Thus, all regression results reported below include the following control variables: median child age, the two measures of child language development (MLU and TTR at T2), caregivers' overall use of valenced language during T1, and the number of caregiver mental state label instances during T1, and concreteness. For simplicity, when reporting model formulae, these variables are referred to collectively as control variables.

## Results and Discussion

**Valence in caregiver input predicts child emotion label production**—First, we examined if the valenced context caregivers provided in the utterances surrounding emotion labels and mental state labels would predict the contexts in which their child later produced emotion labels and mental state labels. That is, do patterns of child production follow patterns of caregiver input? After controlling for the dimensions of speech outlined above, caregiver input during T1 predicted the child's production of emotion and mental state labels during T2, both in the preliminary analysis of 8 emotion labels (child surrounding context (T2) ~ caregiver surrounding context (T1) + control variables ~  $\beta_{\text{caregiver context (T1)}} = 0.35$ ,  $t(31) = 2.18$ ,  $p = 0.037$ ; see Figure 5b top panel) and in the preregistered analysis of 87 mental state labels (child surrounding context (T2) ~ caregiver surrounding context (T1) + control variables ~  $\beta_{\text{caregiver context (T1)}} = 0.4$ ,  $t(24) = 2.36$ ,  $p = 0.027$ ; see Figure 5b bottom panel).

This directionality of the effect was specific to parent input later predicting child output (Supplementary Analysis 8). Child production during T1 did not predict later caregiver valenced context during T2. Similarly, we did not observe concurrent links within T1 or within T2 between caregiver and child valenced context (Supplementary Analysis 8). Together, these results are consistent with the idea that the broad, multi-utterance valenced context in caregivers' input may support children's ability to integrate emotion labels into a similarly complex and dynamic context. Further, we provide some evidence that caregiver valenced context may support child production over time, possibly by allowing children to

accumulate aggregate co-occurrence statistics of mental state labels and related valenced words.

Second, we tested whether the valence of the sentences in which caregivers embed emotion labels predicts children's appropriate production of emotion labels. In the preliminary analysis of 8 emotion labels, after controlling for the dimensions of speech outlined above, the valenced context of sentences surrounding an emotion label in caregiver input during T1 was marginally predictive of the valenced context of sentences surrounding the child's production of emotion labels during T2 (child sentence context (T2) ~ caregiver sentence context (T1) + control variables  $\sim \beta_{caregiver\ context\ (T1)} = 0.37, t(26) = 2, p \sim 0.056$ ; see Figure 5c top panel). However, we did not find evidence for this in the preregistered analysis of 87 mental state labels (child sentence context (T2) ~ caregiver sentence context (T1) + control variables  $\sim \beta_{caregiver\ context\ (T1)} = -0.27, t(17) = -1.01, p \sim 0.33$ ; see Figure 5c bottom panel). These findings suggest only weak evidence for effects of caregiver valenced context on child valenced context within the labeling sentence, potentially resulting from the fact that child utterances at this age are still quite short.

## General Discussion

Children's ability to communicate about their emotions is a key determinant of their socio-emotional development (Hoemann et al., 2019). In this investigation, we describe the emergence of a network of emotion labels and mental state labels in toddlerhood, as well as the contextual cues in caregiver input that may give rise to the development of children's emotion vocabulary. Overall, we showed that children's production of emotion labels is embedded in a developing network of related positive and negative words. We propose that caregivers may support children's ability to draw these connections by surrounding emotion labels with other words matched in valence. Indeed, we found correlational evidence that individual differences across caregivers predicted children's later production of emotion labels in appropriate contexts.

In Studies 1 and 2, we situated children's production of emotion labels within the developmental trajectory of toddler's expanding network of valenced words. Study 1 revealed that toddlers start off by learning predominantly concrete neutral words, and over time, expand the emotional range of their semantic network by incorporating highly positive and highly negative abstract words. This is the case even though valenced words are just as frequent as concrete neutral words. What might explain this learning trajectory, if not frequency? The words included in the MCDI, which we used for the analyses in Studies 1 and 2, were selected to represent children's first words. Since infants learn abstract words later (Bergelson & Swingley, 2013), this set is skewed toward words with concrete referents. Because highly emotional words, like "good" or "bad", are, on average, less concrete than relatively neutral words, like "spoon" or "shake" (Vigliocco et al., 2014), the valenced words in a toddler's vocabulary may be some of the most abstract words they produce. Relatedly, valenced words may be some of the *first* abstract words that young children learn. Supplementary Figure 2 qualitatively shows that the earliest acquired abstract words are higher in emotionality, whereas the later acquired abstract words were relatively more neutral. This is consistent with the findings from Ponari et al. (2018, 2020), which suggest

that older children learn highly positive or negative abstract words earlier than neutral abstract words. Future research across a wider age range from infancy to childhood to adolescence will be needed to understand how abstractness and emotionality may shape children's vocabularies in different ways and at different ages, and how these changes interact with other developing cognitive systems and abilities.

Toddlers' early vocabularies only sparsely include highly positive and negative words and concepts, which shapes the early semantic connections that they can access. Simulation-based and empirical work both suggest that words that are connected to many other words in the child's vocabulary are learned better (Borovsky et al., 2016; Hills et al., 2009). Similarly, in Study 2, we found that children were more likely to produce a given emotion label when they knew other words with similar valence or emotionality. That is, our results suggest that children do not learn emotion labels in isolation, but rather emotion labels are integrated into a network of valenced concepts. There are several possible explanations for this relation. For example, before they learn emotion labels, children may need to learn related valenced words that can then scaffold their learning of emotion labels. As children learn more valenced words, it may be easier for novel emotion labels to become integrated into a pre-existing rich semantic network that contains a full continuum of positive and negative concepts. It is also possible that learned emotion labels serve as 'hubs' that link new, similarly valenced supporting words. The most likely explanation is that there is a bi-directional link, such that knowing more valenced words supports learning new emotion labels and vice-versa. That is, emotion labels, mental state labels, and valenced words may be part of an inseparable interconnected network that is learned as part of a shared semantic space. However, additional research is needed to understand and disentangle the mechanisms that support children's learning of highly positive and negative words, whether those label specific emotions, mental states or neither. The current investigation suggests that children's learning of emotion labels may be best explored in the context of children's full vocabulary network (Wojcik, 2018), rather than in isolation.

What kind of input supports children in forming these semantic connections? In Studies 3, 4 and 5, we found that caregiver input includes consistent links between emotion labels and similarly-valenced words, which may facilitate children's learning over time, although this would need to be investigated in a study that can appropriately pinpoint causal links. Nonetheless, our data suggest that caregivers' contextual scaffolding may support children's learning at two timescales. In the moment, it may activate the appropriate valenced semantic space before and after an emotion label is introduced. Such extended activation can facilitate the consolidation of emotion labels by creating an 'optimal moment' for learning, similar to the contextual effects on memory observed for other words (Borovsky et al., 2010; Coutanche & Thompson-Schill, 2014). Additionally, over the course of many instances of an emotion label in caregiver speech, the surrounding valenced context may increase the co-occurrence probabilities between the emotion label and related valenced words, which may help children to infer similarity between the emotion label and other similarly valenced words. There is evidence that toddlers perceive words with similar occurrence statistics as more similar to each other in meaning (Lany & Saffran, 2010; Wojcik & Saffran, 2015), and that adults perceive novel emotion labels as more similar when they follow each other closely in time (Thornton et al., 2020). Evidence from adult neuroscience suggests

that when events consistently follow each other closely in time, they form overlapping neural representations (Raymond et al., 1992; Thornton et al., 2019). This allows the brain to link representations together that are more similar, as well as to co-activate related concepts in the future. Over time, these statistical associations may allow children to construct dimensions, such as valence, that are important for understanding emotion labels with specificity. Importantly, semantic associations are one of many kinds of statistical regularities that accompany emotional experiences.

In Study 4, we found that children learn emotion labels earlier if those labels are consistently surrounded by informative valenced contexts. Study 4 first demonstrated a relation between an aggregate summary of the valenced context of each emotion label and children's average production of the label. Study 5 then revealed links between each child's input across time and their later production. This longitudinal analysis showed that children whose caregivers surrounded emotion labels with matching valenced contexts in their input went on to also produce emotion labels that matched the sentiment of the surrounding context. Thus, we could trace links between emotion-related context in both caregiver and child speech. These links were robust even when controlling for the child's age, markers of the child's overall language development, specifically, the length of the child's utterances and the complexity of children's overall language use, caregivers' use of valenced language and frequency of labeling emotions and mental states. That said, it is still uncertain whether there is a direct causal link between caregiver input and child production, particularly in the extent to which sentence-level valenced context in caregivers' speech predicts that in young children's speech. We posit that caregivers who use more valenced language surrounding emotion labels provide opportunities for the child to create stronger semantic connections between the label and related words, and therefore enable more successful use of the label in appropriate contexts later on. At the same time, caregivers provide these valenced contexts in a highly interactive setting. It is possible that children who have a better understanding of emotion labels may elicit more contextual elaboration from their caregivers. In fact, there is likely to be bidirectional reinforcement between caregiver and child valenced speech, such that caregivers and children co-create learning moments that activate and expand a broader network of valenced concepts surrounding emotion labels.

### **Limitations and future directions**

The current study relies strongly on descriptive methods based on large pre-existing databases. This approach comes with key strengths: it offers many datapoints per child and in aggregate, and provides insights into naturally occurring behaviors in caregiver-child interactions. However, it comes with key limitations too.

One major limitation in this investigation is that we were limited by the words included in the MCDI, which includes very few emotion labels and was not designed to include control words that span the full spectrum of positive and negative as well as neutral concepts – particularly words that are matched for frequency and concreteness. Additionally, this questionnaire was specifically designed for infants and toddlers. Future research across a wider age range – from infancy to childhood to adolescence – will be needed to understand how abstractness and emotionality may shape children's vocabulary growth in similar

or different ways both within and across ages, and how any potential changes interact with other developing cognitive systems and abilities. Moreover, parent-report measures of children's productive vocabulary pale in comparison to robust, child-driven measures of production and comprehension.

Importantly, our descriptive approach limits our ability to make causal inferences about potential links between caregiver input and child learning. Our measures of caregiver input (quantified using data from the CHILDES database) are disconnected from standardized measures of production (quantified using Wordbank), which makes it impossible to link an individual child's input and their parent-reported ability to produce a given word. Further, all standardized measures of children's production of emotion labels (in Studies 1, 2 and 4) relied on cross-sectional data, due to the deidentified nature of data in WordBank. A combination of densely sampled caregiver input across time and robust measures of production and comprehension would enable firmer conclusions about causal links between valenced context and emotion word learning, especially alongside carefully controlled experimental work. Additionally, going forward, it will be important to examine relations between caregiver input and child production on a shorter timescale, as we do not know how children learn emotion labels or abstract words during real-time interaction with caregivers. It is likely that children play an important role in ramping up emotion during everyday conversations, and therefore play an important (and under-studied) role in shaping their own emotion-related input. Additional research is needed to elucidate how the dynamics of caregiver-child interactions that surround emotion labels give rise to children's context-appropriate production of new emotion words.

In this investigation we characterize the context of emotion labels in terms of the valenced words that surround them. However, in reality, this context is much broader as it is intertwined with the child's personal emotional experiences, as well as their perception of emotion displays from caregivers. For example, Wu et al. (2022) showed that children relied on observed facial configurations and body postures in labeling emotions. This makes emotion labels very different from other types of abstract words in that they accompany internal sensations and external percepts. It is important to expand what exactly constitutes valenced context by including measures of affect-related internal physical sensations, social touch, vocal prosody, and facial and body movements. Further, the dynamics of emotion in speech and caregivers' use of valenced words is likely to vary substantially across cultures, communities and languages. Even though the Wordbank and CHILDES databases allow us to access a much larger English-speaking participant pool than would be possible with individual lab studies, these datasets oversample White, high socio-economic status participants in a limited number of Western countries. Exploring the role of a variety of emotion-related cues across different communities would position the field to uncover a richer complexity of emotion contexts and their impact on learning.

## Conclusion

This investigation introduces new ways to harness the complex statistics of caregivers' language input and children's word learning. Here we used an approach that specifically analyzes abstract, valenced words, but the applicability of this approach extends beyond

just this domain. Much of the research on how children learn novel labels has focused on how infants learn statistical co-occurrences between labels and the physical objects they refer to (Frank et al., 2009; Trueswell et al., 2013; Yu & Smith, 2007). But, how do young children learn novel labels whose referent or meaning is less concrete? Understanding context, broadly defined, offers a crucial path forward (Wojcik, 2018). Our findings have implications for understanding children's word learning beyond emotion labels and related valenced words, and future work can use our approaches to examine the roles of different contextual caregiver cues in shaping children's learning of different types of words. That is, this investigation provides techniques for characterizing the 'quality' of infant-directed speech in a way that is grounded, inclusive, and nuanced. We showed that patterns of words and contexts converged to support the emergence of emotion labels and similarly-valenced words in children's productions, and our approach will enable other researchers to quantify how caregivers dynamically use words that support children's learning of words with complex, abstract meanings.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

The data necessary to reproduce the analyses presented here are publicly accessible as part of the open-source databases Wordbank (<http://wordbank.stanford.edu/data>) and CHILDES (<https://childes.talkbank.org/>). The analytic code necessary to reproduce the analyses presented in this paper is publicly accessible. Code is available at the following URL: [https://osf.io/n689q/?view\\_only=d5484327bd2f4e06ab3d195637459e53](https://osf.io/n689q/?view_only=d5484327bd2f4e06ab3d195637459e53).

The analyses in Studies 1, 2, and 4 as well as the preliminary analyses in Studies 3 and 5 were exploratory. Studies 3 and 5 included preregistered analyses as well. The preregistration for Study 3 can be found at the following URL: [https://aspredicted.org/Y7J\\_MT4](https://aspredicted.org/Y7J_MT4). The preregistration for Study 5 can be found at the following URL: [https://aspredicted.org/G38\\_LRH](https://aspredicted.org/G38_LRH).

We thank the families and researchers that contributed to the Wordbank and CHILDES databases, and the reviewers for their valuable input. This work was supported by NIH R01HD095912 (C.L.W.), NIMH R01MH114904 (D.I.T.), and ACM SIGHPC Computational & Data Science Fellowship (M.L.N.).

## Abbreviations

<b>MCDI</b>	MacArthur-Bates Communicative Development Inventory
<b>CHILDES</b>	Child Language Data Exchange System

## References

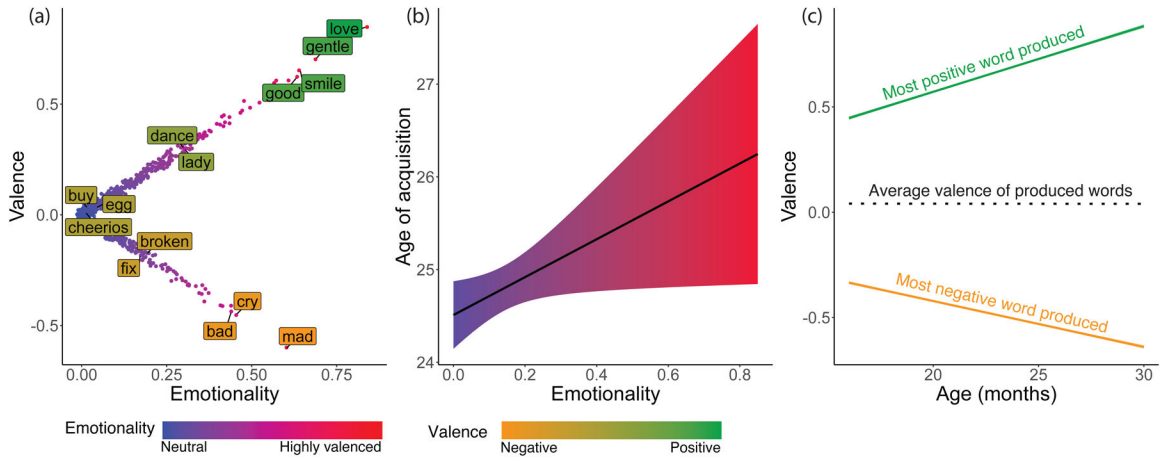
- Barrett LF, Mesquita B, & Gendron M (2011). Context in Emotion Perception. *Current Directions in Psychological Science*, 20(5), 286–290. 10.1177/0963721411422522
- Bates D, Maechler M, Bolker BM, & Walker S (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48. 10.48550/arXiv.1406.5823
- Behrendt S (2014). lm.beta: Add Standardized Regression Coefficients to lm-Objects. R package version 1.5–1 <https://CRAN.R-project.org/package=lm.beta>
- Bergelson E, & Swingle D (2013). The acquisition of abstract words by young infants. *Cognition*, 127(3), 391–397. 10.1016/j.cognition.2013.02.011 [PubMed: 23542412]
- Borghi AM (2022). Concepts for which we need others more: The case of abstract concepts. *Current directions in psychological science*, 09637214221079625. 10.1177/09637214221079625

- Borovsky A, Ellis EM, Evans JL, & Elman JL (2016). Lexical leverage: Category knowledge boosts real-time novel word recognition in 2-year-olds. *Developmental Science*, 19(6), 918–932. 10.1111/desc.12343 [PubMed: 26452444]
- Borovsky A, Kutas M, & Elman J (2010). Learning to use words: event-related potentials index single-shot contextual word learning. *Cognition*, 116(2), 289–296. 10.1016/j.cognition.2010.05.004 [PubMed: 20621846]
- Braginsky M, Yurovsky D, Marchman VA, & Frank M (2016, August). From uh-oh to tomorrow: Predicting age of acquisition for early words across languages. In *CogSci*.
- Braginsky M, Yurovsky D, Frank M, & Kellier D (2018). wordbankr: Tools for connecting to wordbank, an open repository for developmental vocabulary data.
- Cabanac M (2002). What is emotion?. *Behavioural processes*, 60(2), 69–83. 10.1016/S0376-6357(02)00078-5 [PubMed: 12426062]
- Candeloro L, Savini L, & Conte A (2016). A New Weighted Degree Centrality Measure: The Application in an Animal Disease Epidemic. *PLoS One*, 11(11), e0165781. 10.1371/journal.pone.0165781 [PubMed: 27802327]
- Cole PM, Armstrong LM, & Pemberton CK (2010). The role of language in the development of emotion regulation. In Calkins SD & Bell MA (Eds.), *Child development at the intersection of emotion and cognition* (pp. 59–77). American Psychological Association. 10.1037/12059-004
- Coutanche MN, & Thompson-Schill SL (2014). Fast mapping rapidly integrates information into existing memory networks. *Journal of Experimental Psychology: General*, 143(6), 2296–2303. 10.1037/xge0000020 [PubMed: 25222265]
- Dolcos F, LaBar KS, & Cabeza R (2004). Interaction between the amygdala and the medial temporal lobe memory system predicts better memory for emotional events. *Neuron*, 42(5), 855–863. 10.1016/S0896-6273(04)00289-2 [PubMed: 15182723]
- Eisenberg N, Sadovsky A, & Spinrad TL (2005). Associations of emotion-related regulation with language skills, emotion knowledge, and academic outcomes. *New Directions for Child and Adolescent Development*, 2005(109), 109–118. 10.1002/cd.143
- Fabes RA, Eisenberg N, Hanish LD, & Spinrad TL (2001). Preschoolers' Spontaneous Emotion Vocabulary: Relations to Likability. *Early Education and Development*, 12(1), 11–27. 10.1207/s15566935eed1201\_2
- Fenson L (2007). *MacArthur-Bates communicative development inventories*. Paul H. Brookes Publishing Company Baltimore, MD.
- Fernald A, Frank M, Goodman N, & Tenenbaum J (2009). Continuity of discourse provides information for word learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 31. <https://escholarship.org/content/qt7ks028bq/qt7ks028bq.pdf>
- Frank MC, Braginsky M, Yurovsky D, & Marchman VA (2017). Wordbank: an open repository for developmental vocabulary data. *Journal of Child Language*, 44(3), 677–694. 10.1017/S0305000916000209 [PubMed: 27189114]
- Frank MC, Goodman ND, & Tenenbaum JB (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, 20(5), 578–585. 10.1111/j.1467-9280.2009.02335.x [PubMed: 19389131]
- Hamann SB, Ely TD, Grafton ST, & Kilts CD (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. *Nature Neuroscience*, 2(3), 289–293. 10.1038/6404 [PubMed: 10195224]
- Hills T (2013). The company that words keep: comparing the statistical structure of child- versus adult-directed language. *Journal of Child Language*, 40(3), 586–604. 10.1017/S0305000912000165 [PubMed: 22584041]
- Hills TT, Maouene M, Maouene J, Sheya A, & Smith L (2009). Longitudinal analysis of early semantic networks: preferential attachment or preferential acquisition? *Psychological Science*, 20(6), 729–739. 10.1111/j.1467-9280.2009.02365.x [PubMed: 19470123]
- Hoemann K, Wu R, LoBue V, Oakes LM, Xu F, & Barrett LF (2020). Developing an Understanding of Emotion Categories: Lessons from Objects. *Trends in Cognitive Sciences*, 24(1), 39–51. 10.1016/j.tics.2019.10.010 [PubMed: 31787499]

- Hoemann K, Xu F, & Barrett LF (2019). Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis. *Developmental Psychology*, 55(9), 1830–1849. <https://doi.org/10.1037%2Fdev0000686> [PubMed: 31464489]
- Horowitz AC, & Frank MC (2015). Young children’s developing sensitivity to discourse continuity as a cue for inferring reference. *Journal of Experimental Child Psychology*, 129, 84–97. 10.1016/j.jecp.2014.08.003 [PubMed: 25279437]
- Köper M, & im Walde SS (2017, April). Improving verb metaphor detection by propagating abstractness to words, phrases and individual senses. In *Proceedings of the 1st Workshop on Sense, Concept and Entity Representations and their Applications* (pp. 24–30). 10.18653/v1/W17-1903
- Lany J, & Saffran JR (2010). From statistics to meaning: infants’ acquisition of lexical categories. *Psychological Science*, 21(2), 284–291. 10.1177/0956797609358570 [PubMed: 20424058]
- Lew-Williams C, Pelucchi B, & Saffran JR (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science*, 14(6), 1323–1329. 10.1111/j.1467-7687.2011.01079.x [PubMed: 22010892]
- MacWhinney B (2000). *The CHILDES Project: Tools for Analyzing Talk* (third edition): Volume I: Transcription format and programs, Volume II: The database. *Computational Linguistics*, 26(4), 657–657. 10.4324/9781315805641
- MacWhinney B (2014). *The CHILDES project: Tools for analyzing talk, Volume II: The database*. Psychology Press. 10.4324/9781315805672
- Miller JF, & Chapman RS (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research*, 24(2), 154–161. 10.1044/jshr.2402.154 [PubMed: 7265928]
- Nencheva ML, Thornton MA, Lew-Williams C, Tamir DI (2021, April). Development of emotion dynamics in early childhood. Talk at the 2021 Society for Research in Child Development Biennial Meeting (Virtual)
- Nook EC, Sasse SF, Lambert HK, McLaughlin KA, & Somerville LH (2017). Increasing verbal knowledge mediates development of multidimensional emotion representations. *Nature Human Behaviour*, 1, 881–889. 10.1038/s41562-017-0238-7
- Peters RE, Kueser JB, & Borovsky A (2021). Perceptual connectivity influences toddlers’ attention to known objects and subsequent label processing. *Brain Sciences*, 11(2), 163. 10.3390/brainsci11020163 [PubMed: 33513707]
- Ponari M, Norbury CF, & Vigliocco G (2018). Acquisition of abstract concepts is influenced by emotional valence. *Developmental Science*, 21(2). 10.1111/desc.12549
- Ponari M, Norbury CF, & Vigliocco G (2020). The role of emotional valence in learning novel abstract concepts. *Developmental Psychology*. 10.1037/dev0001091
- Raymond JE, Shapiro KL, & Arnell KM (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *Journal of Experimental Psychology. Human Perception and Performance*, 18(3), 849–860. [PubMed: 1500880]
- Roben CKP, Cole PM, & Armstrong LM (2013). Longitudinal relations among language skills, anger expression, and regulatory strategies in early childhood. *Child Development*, 84(3), 891–905. 10.1111/cdev.12027 [PubMed: 23278601]
- R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Russell JA (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. 10.1037/h0077714
- Saffran JR, Aslin RN, & Newport EL (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928. 10.1126/science.274.5294.1926 [PubMed: 8943209]
- Saffran JR, & Kirkham NZ (2018). Infant Statistical Learning. *Annual Review of Psychology*, 69, 181–203. <https://doi.org/10.1146%2Fannurev-psych-122216-011805>
- Sanchez A, Meylan SC, Braginsky M, MacDonald KE, Yurovsky D, & Frank MC (2019). *childes-db: A flexible and reproducible interface to the child language data exchange system*. *Behavior Research Methods*, 51(4), 1928–1941. 10.3758/s13428-018-1176-7 [PubMed: 30623390]

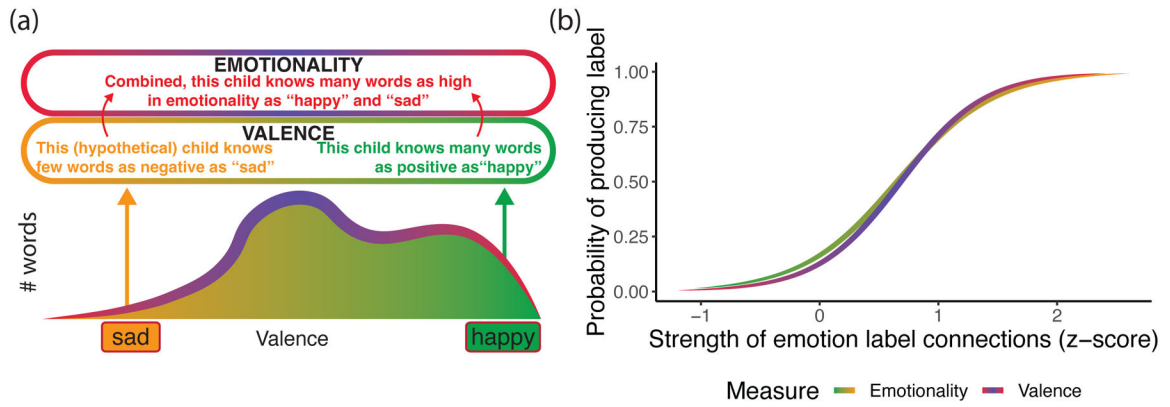


- Schwab JF, & Lew-Williams C (2017). Discourse continuity promotes children's learning of new objects labels. *CogSci*. <https://cogsci.mindmodeling.org/2017/papers/0586/paper0586.pdf>
- Schwab JF, & Lew-Williams C (2020). Discontinuity of Reference Hinders Children's Learning of New Words. *Child Development*, 91(1), e29–e41. 10.1111/cdev.13189 [PubMed: 30516268]
- Shablack H, Stein AG, & Lindquist KA (2020). Comment: A role of Language in Infant Emotion Concept Acquisition. *Emotion Review*, 12(4), 251–253. 10.1177/1754073919897297
- Sharot T, & Phelps EA (2004). How arousal modulates memory: disentangling the effects of attention and retention. *Cognitive, Affective & Behavioral Neuroscience*, 4(3), 294–306. 10.3758/CABN.4.3.294
- Tamir DI, Thornton MA, Contreras JM, & Mitchell JP (2016). Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. *Proceedings of the National Academy of Sciences of the United States of America*, 113(1), 194–199. 10.1073/pnas.1511905112 [PubMed: 26621704]
- Templin MC (1957). *Certain language skills in children: Their development and interrelationships* (Vol. 10). Minneapolis: University of Minnesota Press.
- Thornton MA, Rmus M, Tamir D, & Thornton MA (2020). Transition dynamics shape mental state concepts. 10.31234/osf.io/kbcjs
- Thornton MA, & Tamir DI (2017). Mental models accurately predict emotion transitions. *Proceedings of the National Academy of Sciences of the United States of America*, 114(23), 5982–5987. 10.1073/pnas.1616056114 [PubMed: 28533373]
- Thornton MA, & Tamir DI (2020). People represent mental states in terms of rationality, social impact, and valence: Validating the 3d Mind Model. *Cortex*, 125, 44–59. 10.1016/j.cortex.2019.12.012 [PubMed: 31962230]
- Thornton MA, Weaverdyck ME, & Tamir DI (2019). The Social Brain Automatically Predicts Others' Future Mental States. *The Journal of Neuroscience*, 39(1), 140–148. 10.1523/JNEUROSCI.1431-18.2018 [PubMed: 30389840]
- Trentacosta CJ, & Izard CE (2007). Kindergarten children's emotion competence as a predictor of their academic competence in first grade. *Emotion*, 7(1), 77–88. 10.1037/1528-3542.7.1.77 [PubMed: 17352565]
- Trueswell JC, Medina TN, Hafri A, & Gleitman LR (2013). Propose but verify: fast mapping meets cross-situational word learning. *Cognitive Psychology*, 66(1), 126–156. 10.1016/j.cogpsych.2012.10.001 [PubMed: 23142693]
- Vigliocco G, Kousta S-T, Della Rosa PA, Vinson DP, Tettamanti M, Devlin JT, & Cappa SF (2014). The neural representation of abstract words: the role of emotion. *Cerebral Cortex*, 24(7), 1767–1777. 10.1093/cercor/bht025 [PubMed: 23408565]
- Warriner AB, Kuperman V, & Brysbaert M (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4), 1191–1207. 10.3758/s13428-012-0314-x [PubMed: 23404613]
- Widen SC (2013). Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review*, 5(1), 72–77. 10.1177/1754073912451492
- Wojcik EH (2018). The development of lexical--semantic networks in infants and toddlers. *Child Development Perspectives*, 12(1), 34–38. 10.1111/cdep.12252
- Wojcik EH, & Saffran JR (2015). Toddlers encode similarities among novel words from meaningful sentences. *Cognition*, 138, 10–20. 10.1016/j.cognition.2015.01.015 [PubMed: 25704579]
- Woodard K, Zettersten M, & Pollak SD (2021). The representation of emotion knowledge across development. *Child development*. 10.1111/cdev.13716
- Wu Y, Matteson HM, Baker C, & Frank MC (2022, May 11). Angry, sad, or scared? Within-valence mapping of emotion words to facial and body cues in 2- to 4-year old children. 10.31234/osf.io/ka3ed
- Yu C, & Smith LB (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18(5), 414–420. 10.1111/j.1467-9280.2007.01915.x [PubMed: 17576281]



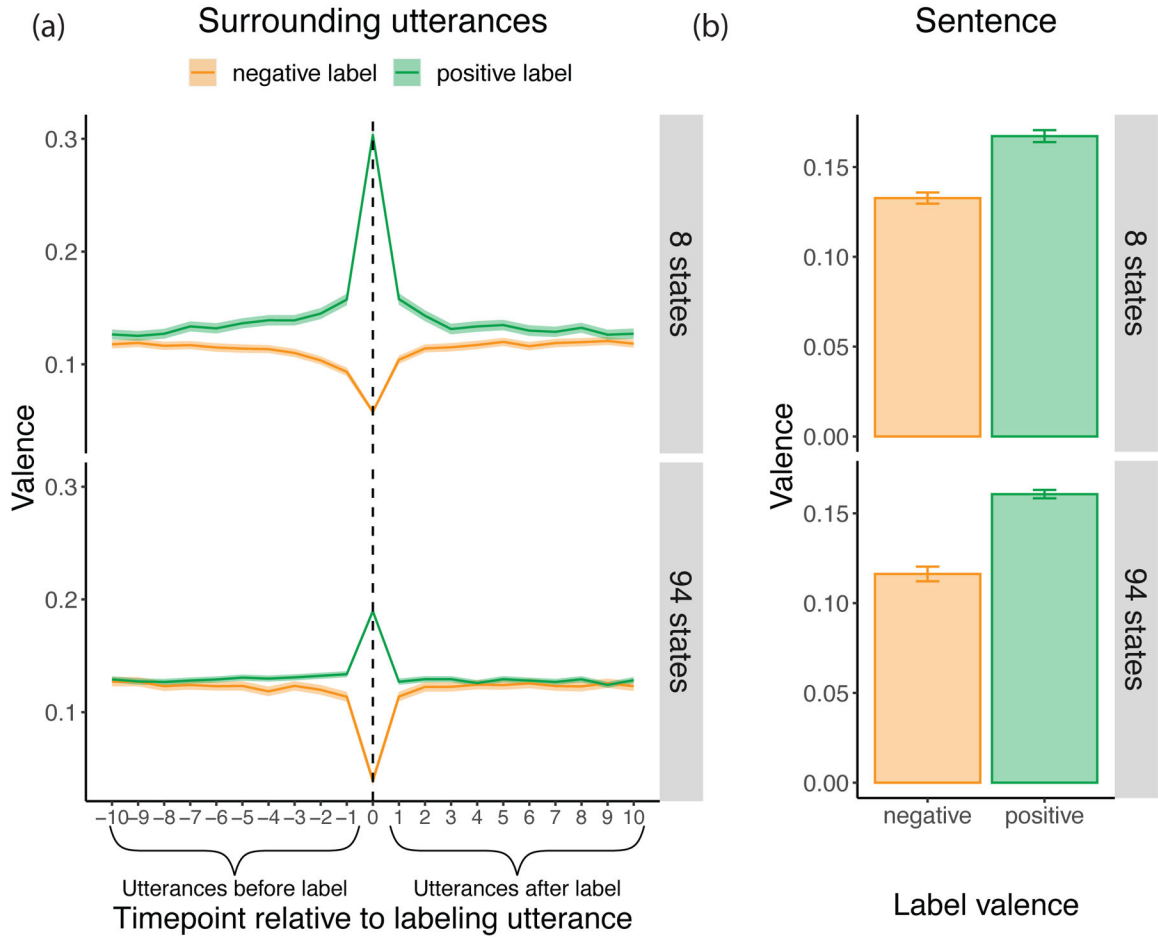
**Figure 1.**

Panel (a) shows the distribution of valence (y-axis) and emotionality (x-axis) across all words in the MCDI. Since emotionality is computed as the absolute value of valence, highly positive and negative words are both high in emotionality. Points were randomly jittered to show the spread of values. A few select words are highlighted to illustrate the kinds of words that are in different parts of the spectrum. The color of the points represents their emotionality (highly positive / negative words in red and neutral words in blue), and the color of the words represents their valence (negative words in yellow and positive words in green). Panel (b) shows the relation between the emotionality of a word (on the x-axis), and the age of acquisition of the word as the age at which 50% of children produce the word based on wordbankr (on the y-axis). The shaded area shows a 95% confidence interval of the prediction. Panel (c) shows the link between age and the valence of the most positive (green) and most negative (yellow) word that children at the corresponding age in this sample produced. The dashed line represents the average valence of all words produced by children at a given age.



**Figure 2:**

Panel (a) shows an example of how two labels can be similar or different in how well-connected they are to other words in the child’s vocabulary. The mock density plot represents how many words the child knows in different parts of the valence continuum. Yellow represents negative words and green - positive ones. Red represents words that are highly positive or negative, and blue - relatively neutral words. For example, this child knows quite a few positive and neutral words, but relatively few negative labels. This means that the label “happy” is strongly connected both in terms of valence (as the child knows many other positive words) as well as emotionality (as there are many highly positive or highly negative words in the child’s vocabulary). The label “sad” is similarly strongly connected in terms of emotionality, but less so in terms of valence (as the child doesn’t know many negative words). Panel (b) shows the probability of a child producing an emotion label (y-axis) as a function of how strongly the emotion label is connected to other words the child knows (x-axis) based on the predictions of a logistic regression fitted to the data. The green-yellow gradient line represents connection strength in terms of valence, and the red-blue gradient line represents connection strength in terms of emotionality.



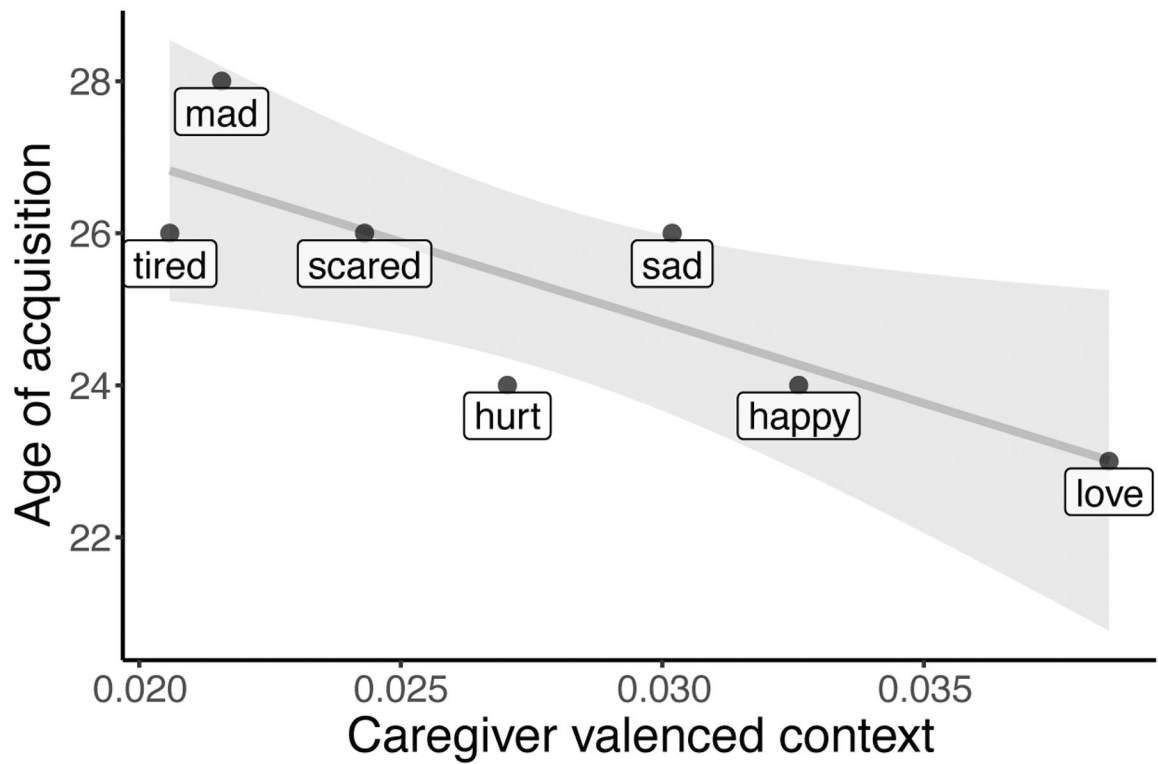
**Figure 3:** Panel (a) shows the sentiment of the utterances preceding and following an emotion label for the preliminary analysis with 8 emotion labels (top) and the preregistered analysis with 94 mental state labels (bottom). The lines represent the average sentiment at each time point relative to the labeling utterance for positive (green) and negative (yellow) labels respectively, with the shaded area representing a 95% confidence interval. Panel (b) shows the sentiment of the sentence (excluding the emotion label) in labeled utterances for the 8-emotion-label analysis (top) and the 94-mental-state-label analysis (bottom). Panel (b) shows the average sentence valence for positive and negative labels (in green and yellow respectively), with 95% confidence interval error bars. Note that the average valence of all utterances is positive (above 0). This reflects the generally positive sentiment of child-directed speech. However, there is still an overall difference in valence such that the utterances preceding negative emotion labels become less positive.

Author Manuscript

Author Manuscript

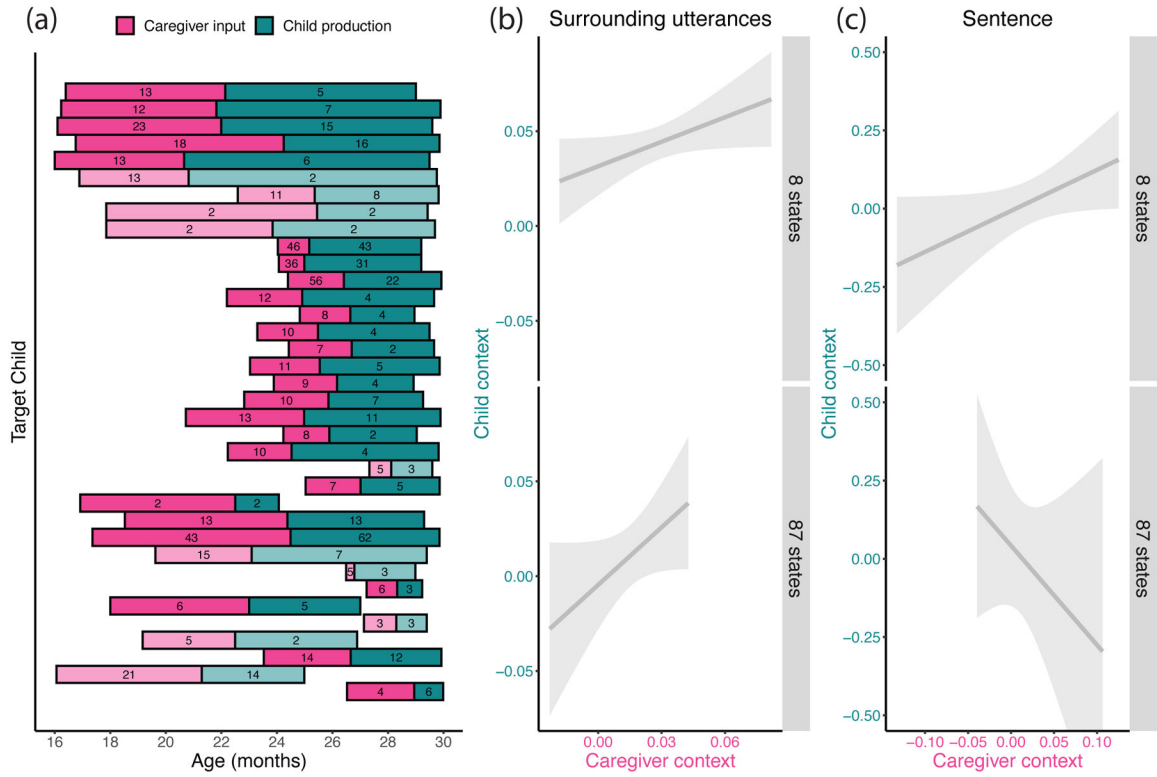
Author Manuscript

Author Manuscript



**Figure 4:**

The x-axis shows the average valenced contexts of all instances for each emotion label in caregiver speech for children between 16 and 30 months. The y-axis shows the age of acquisition of each emotion label (measured as the age at which half of children produce the label in a normative sample from Wordbank).



**Figure 5:**

Panel (a) is a descriptive plot of the age range for each dyad (x axis). Each row represents one dyad, with the darkened part of the age range marking the earlier section used to assess caregiver input. The lighter section shows the part of the age range used to assess child production. The numbers in each rectangle show the number of transcripts included in that portion of the dyad’s age range. Panels (b) and (c) show how caregiver valenced context in the earlier portion of the age range (x-axis) relates to the child’s valenced context in the later part of the age range (y-axis). Panel (b) shows the valenced context of the utterances surrounding an emotion or mental state label, whereas panel (c) shows the valenced context of the sentence in which the emotion or mental state label was embedded. The top panels in (b) and (c) show the results from the preliminary analysis of 8 emotion labels, and the bottom panels show the preregistered analysis of 87 mental state labels.