Supplementary

Supplementary Analysis 1: Examining influential data points in Study 1

In this analysis, we examined whether the effects of emotionality on age of acquisition are driven by few highly influential points. To identify highly influential points, we computed Cook's Distance (Cook, 1977), using the `cooks.distance` function in R. We then removed 21 data points that had a Cook's Distance higher than a threshold of 4/n (based on Altman & Krzywinski (2016)), where n is the number of words in the analysis. Emotionality remained a significant predictor of age of acquisition (AoA ~ emotionality; $\beta_{\text{emotionality}} = 0.08$, $t(585) = 1.99$, $p = 0.05$). This suggests that the effects of emotionality on age of acquisition reflect an overall trend across all data points rather than a pattern driven by only a few data points.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, *19*(1), 15-18.

Altman, N., & Krzywinski, M. (2016). Analyzing outliers: influential or nuisance?. *Nature Methods*, *13*(4), 281-283.

Supplementary Analysis 2: Examining the effects of highly valenced words in Study 1

In this analysis, we examined whether the effects of emotionality on age of acquisition were driven by the most extremely positive or negative words. To do so we removed the top and bottom 5% of words in terms of valence (the most extremely positive and negative words) - a total of 62 words. As in Supplementary Analysis 1, emotionality remained a significant predictor of age of acquisition (AoA ~ emotionality; $\beta_{\text{emotionality}} = 0.13$, $t(544) = 2.97$, $p = 0.003$). This suggests that the effects of emotionality on age of acquisition were entirely driven by only a few extremely positive or negative words.

Supplementary Analysis 3: Differences between positive and negative words

First, we examined whether valence predicts age of acquisition and whether it interacts with emotionality in doing so. Valence (measured as a continuous factor from negative to neutral to positive) did not predict age of acquisition (AoA ~ valence: $\beta_{\text{valence}} = 0.06$, $t(606) = 1.42$, $p = 0.16$), even when controlling for frequency for the words where frequency information was available (AoA ~ valence + frequency: $\beta_{\text{valence}} = 0.06$, t(560) = 1.42, p = 0.16). This suggests that positive words are not learned earlier (or later) than negative words, replicating prior work by Braginsky et al. (2016).

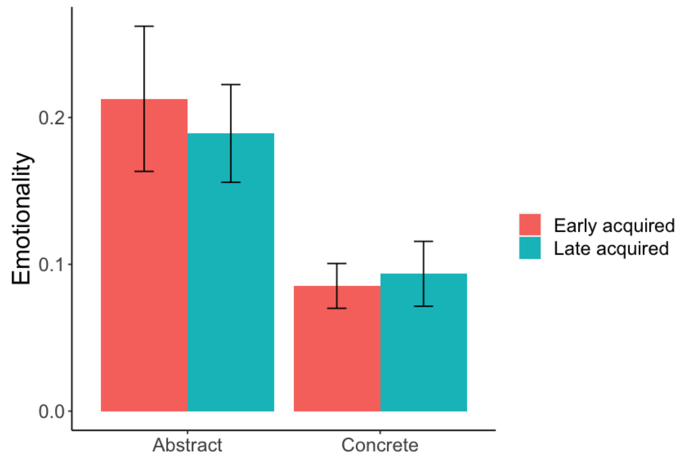
Next, we examined if the effects of emotionality on age of acquisition are modulated by the valence of the word - that is, whether higher emotionality predicts age of acquisition differently for positive vs. negative words. We did not find evidence for an interaction between emotionality and valence (AoA ~ valence × emotionality: $\beta_{\text{valence} \times \text{emotionality}} = -0.04$, $t(604) = -0.42$, $p = 0.67$), suggesting that the effects of emotionality were similar for positive and negative words.

Finally, we probed whether the valenced context of the surrounding utterances differs for positive vs. negative emotion and mental state labels in Study 3. We found conflicting results, such that valenced context was more strongly matched to the label for positive (vs. negative) labels in the preliminary set of 8 labels (match in valence of utterance ~ distance of utterance from label × valence of label+ label concreteness + (1 | speaker); $\beta_{\text{distance} \times \text{valence}} = -0.02$, $t(152969.03) = -6.03$, $p = 1.65\times10^{-9}$), however was less strongly matched (compared to that of negative labels) in the preregistered set of 94 mental states ($\beta_{\text{distance} \times \text{valence}} = 0.02$, $t(170742.13) = 6.25$, $p = 4.14 \times 10^{-10}$).

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. (2016, August). From uh-oh to tomorrow: Predicting age of acquisition for early words across languages. In *CogSci*.

Supplementary Analysis 4: Interaction between emotionality and abstractness

This supplementary analysis explored the interaction between emotionality and abstractness in predicting the age of acquisition of a word. We obtained abstractness ratings from Köper & im Walde (2017) for the words in the MCDI (available for 560 of the words). There was a significant correlation between abstractness and valence ($r = 0.4$, $p < 0.001$). Next, wWe split the MCDI words into three quantiles. The lowest quantile was labeled as “Concrete” and the highest quantile as “Abstract”. Similarly, we split words by their age of acquisition (see Study 1 methods for details) into three quantiles. Words in the lowest quantile were marked as “Early acquired” and those in the highest quantile as “Late acquired”. Qualitatively, the earliest acquired abstract words are higher in emotionality (that is. they are either highly positive or highly negative), whereas the later acquired abstract words were relatively more neutral. Conversely, the opposite trend was true for concrete words - the earlier acquired words were relatively more neutral, whereas the later acquired words were relatively higher in emotionality. Additionally, we tested this interaction in a regression predicting age of acquisition, however, found no significant interaction (AoA ~ emotionality * concreteness; $\beta_{emotionality * concreteness} \sim 0.02$, $t(556) = 0.73$, $p = 0.46$).



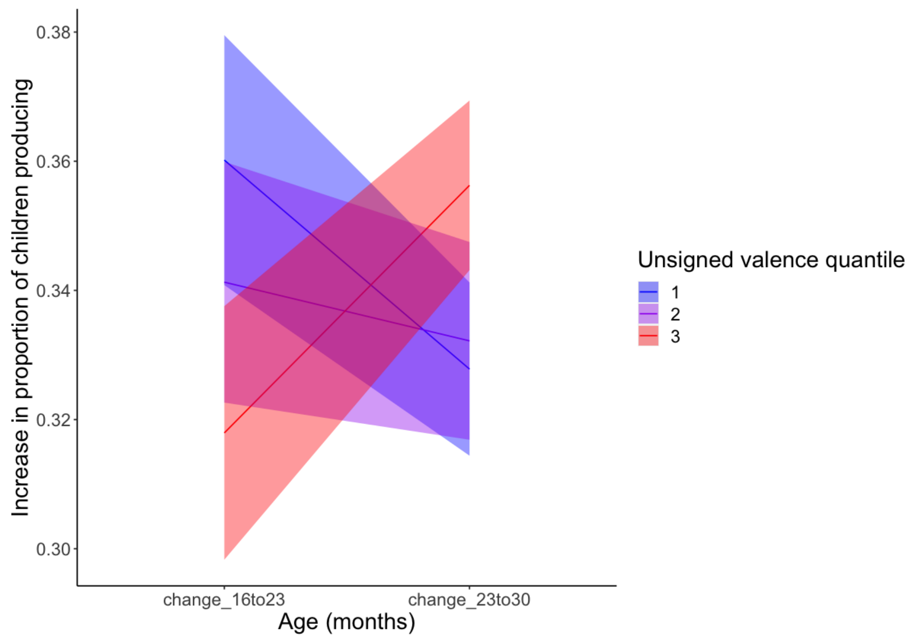
Supplementary Figure 1. Emotionality is plotted on the y-axis, with low values indicating relatively more neutral words and high values - highly positive or negative words. Abstract and concrete words represent the top and bottom third of words in terms of abstractness respectively. Similarly, early and late acquired words were selected as the bottom and top third in terms of age of acquisition.

Köper, M., & im Walde, S. S. (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).

Supplementary Analysis 5: Growth in child production along the valence continuum

We explored how the valence of a word relates to the growth in child production of that word across this age range. We measured the increase (difference) in the proportion of children producing each word across the first half of the age range (from 16 to 23 months), and in the second half of the age range (from 24 to 30 months). In a linear regression predicting the increase in proportion of children producing a given word, there was a significant positive interaction ($\beta \sim 0.24$, $p < 0.001$) between the continuous emotionality of the word and the time period (first vs. second part of the age range). This suggests that for the least valenced words, child production increased the most during the first half of the age range, and less so during the second half of the age range. However, for the most valenced words, child production did not increase as much during the first half of the age range, and instead increased the most during the second half of the age range (see Supplementary Figure 2).

Supplementary Figure 2: This plot shows the increase in proportion of children producing a given word over the first half of the age range (16-23mo) and the second half of the age range (23-30mo) - important note, they don't show the proportion of children producing word, just the change in proportion which is always positive because more children produce a word as we go up in age. The three lines illustrate the 3 quantiles of emotionality (1 being the lowest, or most neutral one, 3 being the highest or most strongly valenced - I'll fix the labels) are there for illustration purposes (even though the analysis was done using continuous absolute valence). The idea is that the most neutral words (blue -1) increase a lot in production between 16 and 23 months and then slightly taper off with smaller increases in production between 24 and 30 months. The most valenced words (red -3) increase slightly in production between 16 and 23 months but quite a bit more so between 24 and 30 months.

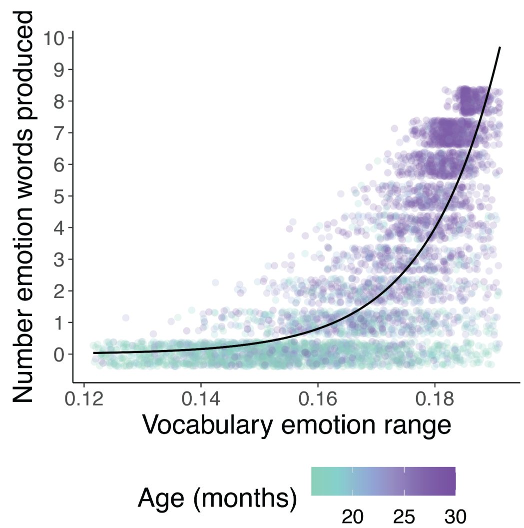


Supplementary Analysis 6

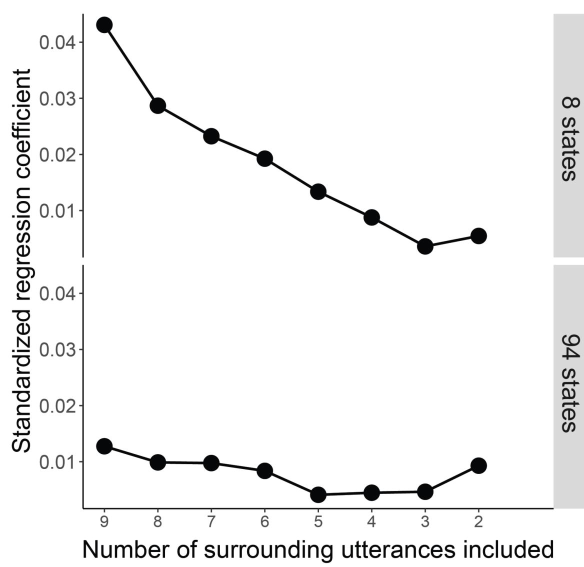
As in Study 1, we computed the emotional range of each child's vocabulary as the standard deviation of the valence of all the words they produce. We excluded any emotion labels they produce when computing this measure.

Emotional range predicts the production of emotion labels

In a confirmatory analysis, we tested whether children produce more emotion labels when their vocabularies had a wider emotional range. We ran a Poisson regression predicting the number of emotion labels the child produces from the emotional range of the rest of their productive vocabulary. Children who had vocabularies with a wider emotional range produced more emotion labels (number of mental state labels ~ emotion range; $\beta = 0.18$, $z = 96.63$, $p < 0.0001$). This effect remained significant when controlling for total vocabulary size and age ($b = 0.49$, $\beta = 0.17$, $z = 40.11$, $p < 0.0001$). Therefore, children whose vocabularies represented a wider range of positive and negative words produced more emotion labels. This finding offers preliminary evidence for the way in which children may bootstrap knowledge of emotion labels. As the emotional range of children's vocabularies increases, they can form stronger semantic connections between emotion labels and related valenced words. This in turn can facilitate children's production of both emotion labels and valenced words.



Supplementary Figure 3. Emotion range predicts number of emotion labels produced. The black line represents the predictions of a Poisson regression, and the gradient represents age in months, such that teal is 16 months and purple is 30 months.



Supplementary Figure 4. The increase in valence match in surrounding utterances with greater proximity to the label is strongest when including utterances closest to the label. In this analysis, we computed the regression coefficient of the Study 3 analysis of the valenced context of surrounding utterances, by shifting the analysis window start away from the labeled utterance. For example, for the analysis including 9 data points (plotted on the x-axis), we included utterances that were between 1 and 10 utterances away from the label. For the analysis including 4 data points, we only included utterances that were between 6 and 10 utterances away from the labeled utterance. On the y-axis, we plotted the standardized regression coefficient (averaged over positive and negative emotion labels). The drop in the regression coefficient indicated that the result was primarily driven by points closest to the label.

Supplementary Analysis 7: Valenced context by age

We probed whether valenced context, as measured in Studies 3 and 4, varies depending on the age of the child. We found mixed results for the different analyses, which we summarize in the table below.

Label set	Type of context	Model formula	Interaction results
8 labels	Surrounding utterances (positive labels)	valence ~ distance from label × age + (1|speaker)	$\beta_{\text{distance×age}} = 0.01$, $t(65014.72) = 0.5$, $p = 0.62$
	Surrounding utterances (negative labels)	valence ~ distance from label × age + (1|speaker)	$\beta_{\text{distance×age}} = 0.05$, $t(87868.33) = 1.92$, $p = 0.05$
	Sentence (all labels)	valence of sentence ~ valence of label × age + (1|speaker)	$\beta_{\text{label valence×age}} = -0.22$, $t(7451.43) = -2.41$, $p = 0.01$

94 labels	Surrounding utterances (positive labels)	valence ~ distance from label × age + (1|speaker)	$\beta_{\text{distance×age}}$ = 0.57, $t(907.71) = 2.93$, p = 0.003
	Surrounding utterances (negative labels)	valence ~ distance from label × age + (1|speaker)	$\beta_{\text{distance×age}}$ = 0.06, $t(747.3) = 0.36$, $p = 0.71$
	Sentence (all labels)	valence of sentence ~ valence pf label × age + (1|speaker)	$\beta_{\text{label valence×age}}$= 0.03, $t(8385.06) = 0.53$, p = 0.59

Supplementary Analysis 8: Concurrent links between caregiver input and child production

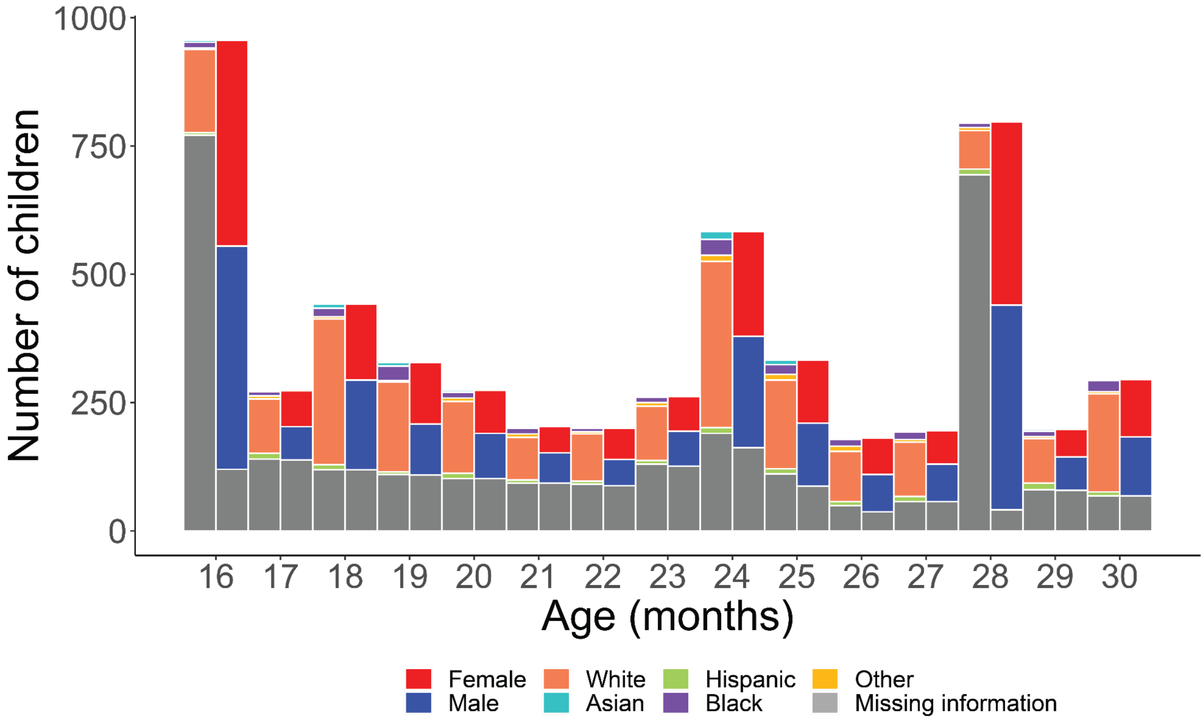
In this supplementary analysis, we explored concurrent links between child production and caregiver input in Study 5. Additionally, we probed whether child production during the first part of the dyad's data (TP1) predicted later caregiver input in the second part of the dyad's data (TP2). We did not find evidence for significant concurrent or reverse links between caregiver input and child production. Regression results are reported in the table below.

Label set	Model formula	Results
8 labels	child surrounding valenced context (T2) ~ caregiver surrounding valenced context (T2) + median child age	$\beta_{\text{caregiver context(T2)}}$= -0.08, $t(39) = -0.48$, $p = 0.63$
87 labels	child surrounding valenced context (T1) ~ caregiver surrounding valenced context (T2) + median child age	$\beta_{\text{caregiver context(T2)}}$= 0.15, $t(32) = 0.83$, $p = 0.41$

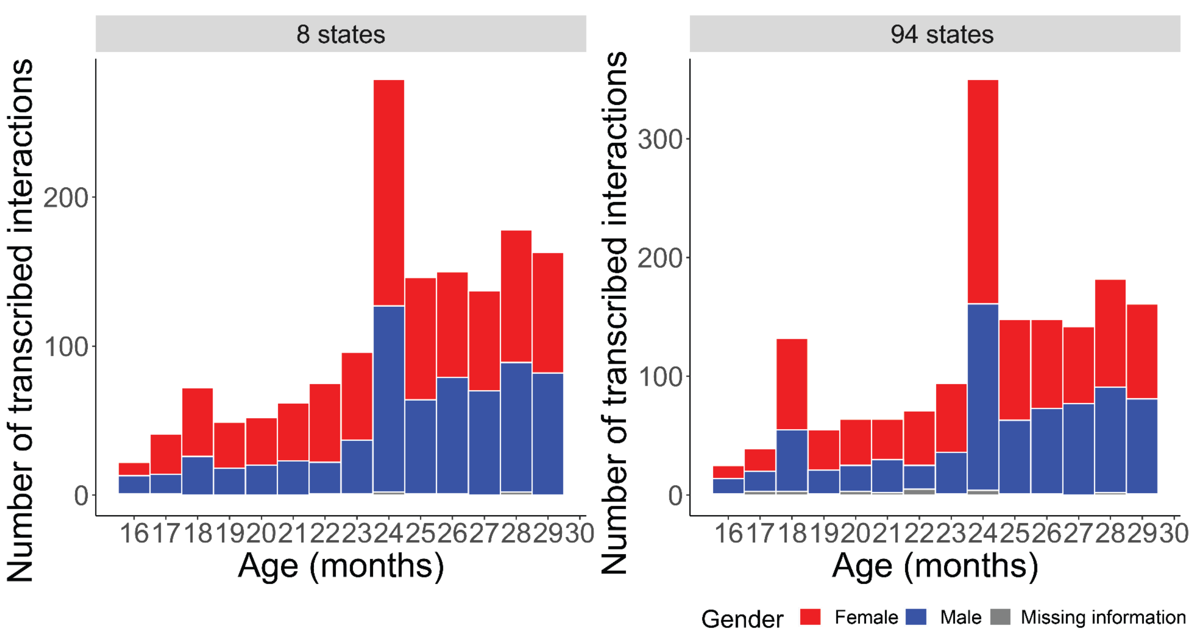
8 labels	child surrounding valenced context (T1) ~ caregiver surrounding valenced context (T1) + median child age	$\beta_{\text{caregiver context(T1)}}$= 0.29, t(31) =1.69, p = 0.10
87 labels	child surrounding valenced context (T1) ~ caregiver surrounding valenced context (T1) + median child age	$\beta_{\text{caregiver context(T1)}}$= 0.27, t(23) = 1.27, p = 0.21
8 labels	caregiver surrounding valenced context (T2) ~ child surrounding valenced context (T1) + median child age	$\beta_{\text{child context(T1)}}$= 0.01, t(31) =0.06, p = 0.95
87 labels	caregiver surrounding valenced context (T2) ~ child surrounding valenced context (T1) + median child age	$\beta_{\text{child context(T1)}}$= 0.06, t(23) = 0.29, p = 0.77

Supplementary Analysis 9:

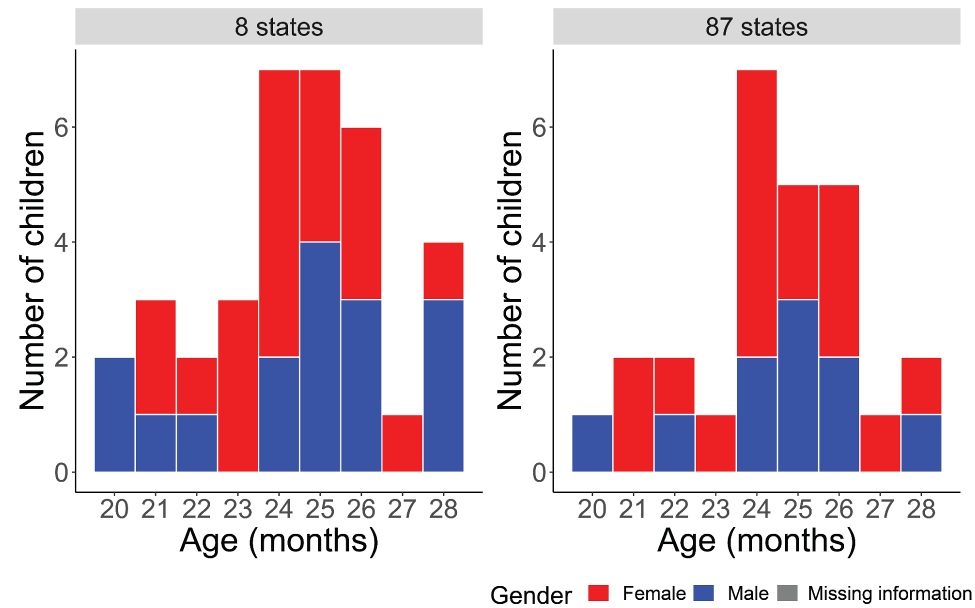
In Study 4, we found a relation between the valence of the label and its valenced context. Negative emotion labels (such as "mad") were less likely than positive emotion labels (such as "happy") to be surrounded by utterances with matched valenced (AoA ~ label valence; β = 0.81, t(6) = 3.32, p = 0.02). This may be because the tone of child-directed speech tends to be positive on average.



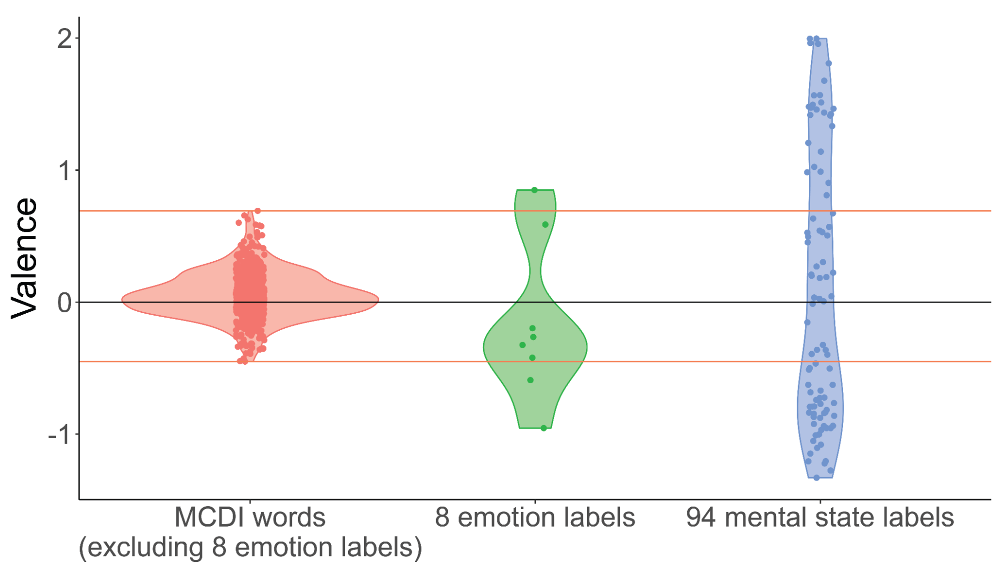
Supplementary Figure 5. Histogram of Wordbank participants. This histogram reflects participants included in Studies 1 and 2, as well as in estimating production data in Study 4. Each 1-month age bin is shows as two side-by-side rectangles, split by ethnicity (on the left) and gender (on the right), where this information was available.



Supplementary Figure 6. Histogram of CHILDES transcribed interactions by child age. This histogram shows transcripts included in Study 3 and the input portion of Study 4. The left panel includes transcripts that were used in th preliminary analysis of 8 labels and the right panel is based on data included in the preregistered analysis of 94 labels.



Supplementary Figure 7. Histogram of CHILDES dyads included in the Study 5 analyses, by the median child age in each dyad's data. The left panel includes dyads that were used in th preliminary analysis of 8 labels and the right panel is based on data included in the preregistered analysis of 87 labels.



Supplementary Figure 8. Distribution of valence in MCDI words (left, in orange), the preliminary set of 8 emotion labels (middle, in green), and the preregistered set of 94 emotion labels in Study 3 (right, in blue). The black horizontal line represents neutral valence (zero). The top and bottom orange lines represent the valence of the most positive and most negative MCDI word respectively (not counting the 8 emotion labels).