Supplementary Materials

Experiment 1 Additional Analyses

Age analyses. Although the Age x Test Trial interaction was not significant, separate 4 (Test Trial) x 2 (Condition) mixed-model ANOVAs were conducted for each age group to confirm that both groups failed to form a superordinate category (see Supplementary Figure 1). A significant effect of Test Trial was found for the 14-month-olds, $F(3, 138) = 13.65$, $p < .001$, $\eta_p^2 = .23$. Fourteen-month-olds attended significantly less to the *negative familiar event* compared to the *negative novel face* ($p < .001$, $d = .77$), the *negative novel emotion* ($p < .001$, $d = .62$), and the *positive novel emotion* ($p < .001$, $d = .99$). Infants at this age also looked significantly longer at the *positive novel emotion* compared to the *negative novel face* ($p = .022$, $d = .34$) and marginally longer to the *negative novel emotion* ($p = .067$, $d = .27$). A significant effect of Test Trial was also found for the 18-month-olds, $F(3, 138) = 18.50$, $p < .001$, $\eta_p^2 = .29$. Eighteen-month-olds attended significantly less to the *negative familiar event* compared to the *negative novel face* ($p < .001$, $d = .90$), the *negative novel emotion* ($p < .001$, $d = .97$), and the *positive novel emotion* ($p < .001$, $d = 1.12$). There were no other significant comparisons for either age (see Supplementary Table 1 for cell means).

Did 14-month-old infants track the habituation events in the Disgust-Sad condition? A repeated measures ANOVA confirmed a significant effect of Test Trial for the 14-month-olds in the *Disgust-Sad* condition, $F(3, 69) = 5.71$, $p < .001$, $\eta_p^2 = .20$. Much like the overall sample, these infants attended significantly less to the *negative familiar event* compared to the *negative novel face* ($p = .003$, $d = .68$), the *negative novel emotion* ($p = .013$, $d = .55$), and the *positive novel emotion* ($p < .001$, $d = .87$). This

suggests that the 14-month-olds in the *Disgust-Sad* condition were able to track which event they had seen during habituation, despite being unable to form a superordinate category (see Supplementary Table 1 for cell means).

Experiment 2 Additional Analyses

Age analyses. Although the Age x Trial x Condition interaction was not significant, separate repeated-measures ANOVAs were conducted for each age group to confirm that both groups formed a superordinate category in the *Disgust-Sad* condition (see Supplementary Figure 2). A significant effect of Test Trial emerged for the 14-month-olds, $F(3, 93) = 14.42$, $p < .001$, $\eta_p^2 = .32$, and the 18-month-olds, $F(3, 93) = 5.00$, $p = .003$, $\eta_p^2 = .14$. Similar to the overall sample, paired t-tests revealed that 14-month-olds looked significantly longer at the *positive novel emotion* compared to the *negative familiar event* ($p < .001$, $d = 1.09$), the *negative novel face* ($p < .001$, $d = .85$), and the *negative novel emotion* ($p < .001$, $d = .86$). Infants' looking to the *negative familiar event* did not differ from the *negative novel face* ($p > .25$, $d = .12$) or the *negative novel emotion* ($p > .25$, $d = .03$) and looking to the *negative novel face* did not differ from the *negative novel emotion* ($p > .25$, $d = .16$). Similar results were found for the 18-month-olds. Paired t-tests revealed that 18-month-olds looked significantly longer at the *positive novel emotion* compared to the *negative familiar event* ($p = .002$, $d = .61$), the *negative novel face* ($p = .040$, $d = .38$), and the *negative novel emotion* ($p = .022$, $d = .43$). Infants' looking to the *negative familiar event* did not differ from the *negative novel face* ($p = .13$, $d = .27$) or the *negative novel emotion* ($p = .14$, $d = .27$), and looking to the *negative novel face* did not differ from the *negative novel emotion* ($p > .25$, $d = .03$) (see Supplementary Table 1 for cell means).

Analyses to equate the sample size of experiment 1 with experiment 2. In order to compare the results of Experiment 1 and 2 with the same sample size we also ran a

supplementary analysis in which we used only the first 24 infants tested in each age/condition in Experiment 2, thereby equating the sample size with Experiment 1. The results were the same as when the full *N* was tested (as reported in the main text). For completeness, the analyses with the reduced *N* are provided below (see Supplementary Table 1 for cell means).

Habituation phase. A 2 (Age) x 2 (Condition) x 2 (Trials) mixed-model ANOVA was conducted. A significant main effect of Trials, $F(1, 92) = 405.70$, $p < .001$, $\eta_p^2 = .82$, revealed that infants looked significantly longer to the first three habituation trials ($M = 24.57$s, $SD = 4.83$s) compared to the *negative familiar event* ($M = 10.23$s, $SD = 6.59$s). There were no other significant effects. Thus, infants did not reach the habituation criteria by chance.

Test phase. Infants' looking times during the test trials were analyzed in a 4 (Test Trial) x 2 (Age) x 2 (Condition) mixed-model ANOVA. Significant main effects of Test Trial, $F(3, 276) = 21.96$, $p < .001$, $\eta_p^2 = .19$, and Condition, $F(1, 92) = 7.78$, $p = .006$, $\eta_p^2 = .08$, were qualified by a significant Test Trial x Condition interaction, $F(3, 276) = 6.13$, $p < .001$, $\eta_p^2 = .06$. There were no other significant main effects or interactions (see Supplementary Figure 3).

Follow-up analyses were conducted separately by Condition. For the *Anger-Sad* condition, a significant main effect of Test Trial emerged, $F(3, 138) = 14.87$, $p < .001$, $\eta_p^2 = .24$. Follow-up comparisons revealed that infants looked significantly less at the *negative familiar event* compared to the *negative novel face* ($p < .001$, $d = .58$), the *negative novel emotion* ($p < .001$, $d = .75$), and the *positive novel emotion* ($p < .001$, $d = .80$). There were no other significant effects. Thus, as in Experiment 1, infants did not form a superordinate category of negative valence in the *Anger-Sad* condition.

For the *Disgust-Sad* condition, a significant main effect of Test Trial also emerged, $F(3, 138) = 13.08$, $p < .001$, $\eta_p^2 = .22$. Follow-up comparisons between the test trials revealed that

infants looked significantly more at the *positive novel emotion* compared to the *negative familiar event* ($p < .001$, $d = .81$), the *negative novel face* ($p < .001$, $d = .67$), and the *negative novel emotion* ($p < .001$, $d = .57$). Critically, however, infants' looking to the *negative familiar event* did not differ from the *negative novel face* ($p > .25$, $d = .12$) or the *negative novel emotion* ($p > .25$, $d = .16$), and looking to the *negative novel face* did not differ from the *negative novel emotion* ($p > .25$, $d = .04$). Thus, infants formed a superordinate category of negative valence in the *Disgust-Sad* condition (see Supplementary Table 1 for cell means).

Exploratory vocabulary analyses. A secondary aim of Experiment 2 was to explore whether infants' vocabulary size was related to their looking times during the test trials. Preliminary analyses indicated that18-month-olds had significantly larger receptive vocabularies ($M = 195.31$, $SD = 94.76$), $t(126) = 6.25$, $p < .001$, $d = 1.10$, and expressive vocabularies ($M = 66.78$, $SD = 72.96$), $t(126) = 5.79$, $p < .001$, $d = 1.02$, compared to 14-month-olds (receptive: $M = 102.91$, $SD = 70.79$; expressive: $M = 13.05$, $SD = 13.99$). For this reason, the vocabulary analyses were conducted separately by Age. Given the differences in category formation between the *Anger-Sad* and *Disgust-Sad* conditions, vocabulary analyses were also conducted separately by Condition. Three difference scores were calculated for infants, representing the difference between their looking time to each of the *novel* test trials relative to the *familiar negative emotion* test trial. Difference scores greater than zero indicate longer looking time to the novel test trial. Correlations were then calculated to examine whether receptive and/or expressive vocabulary scores were related to these difference scores.

In the receptive vocabulary analyses, one 14-month-old in the *Anger-Sad* condition was excluded due to an extremely high vocabulary score (+3 *SD* above the

mean). There were no significant correlations for either Age or Condition (see Supplementary Table 2).

In the expressive vocabulary analyses, two 14- and two 18-month-olds (one infant in each condition) were excluded due to extremely high vocabulary scores (+3 *SD* above the mean). For the 14-month-olds in the *Anger-Sad* condition, expressive vocabulary was significantly and negatively related to looking time to the *negative novel face*, $r = -.38$, $p = .033$, and the *positive novel emotion*, $r = -.36$, $p = .045$. Expressive vocabulary was also marginally negatively related to looking time to the *negative novel emotion*, $r = -.35$, $p = .055$. There were no significant correlations for the 18-month-olds (see Supplementary Table 2).

Thus, 14-month-olds with larger expressive vocabularies spent less time looking at the *negative* test events in the *Anger-Sad* condition. This finding seems to be consistent with findings that labels are more likely to facilitate categorization among infants with larger vocabularies (e.g., Waxman & Markow, 1995). However, it is important to note that 14-month-olds also showed shorter looking times to the *positive novel emotion*. Given the relatively small sample sizes used to conduct these correlations ($n = 32$), this finding requires replication.

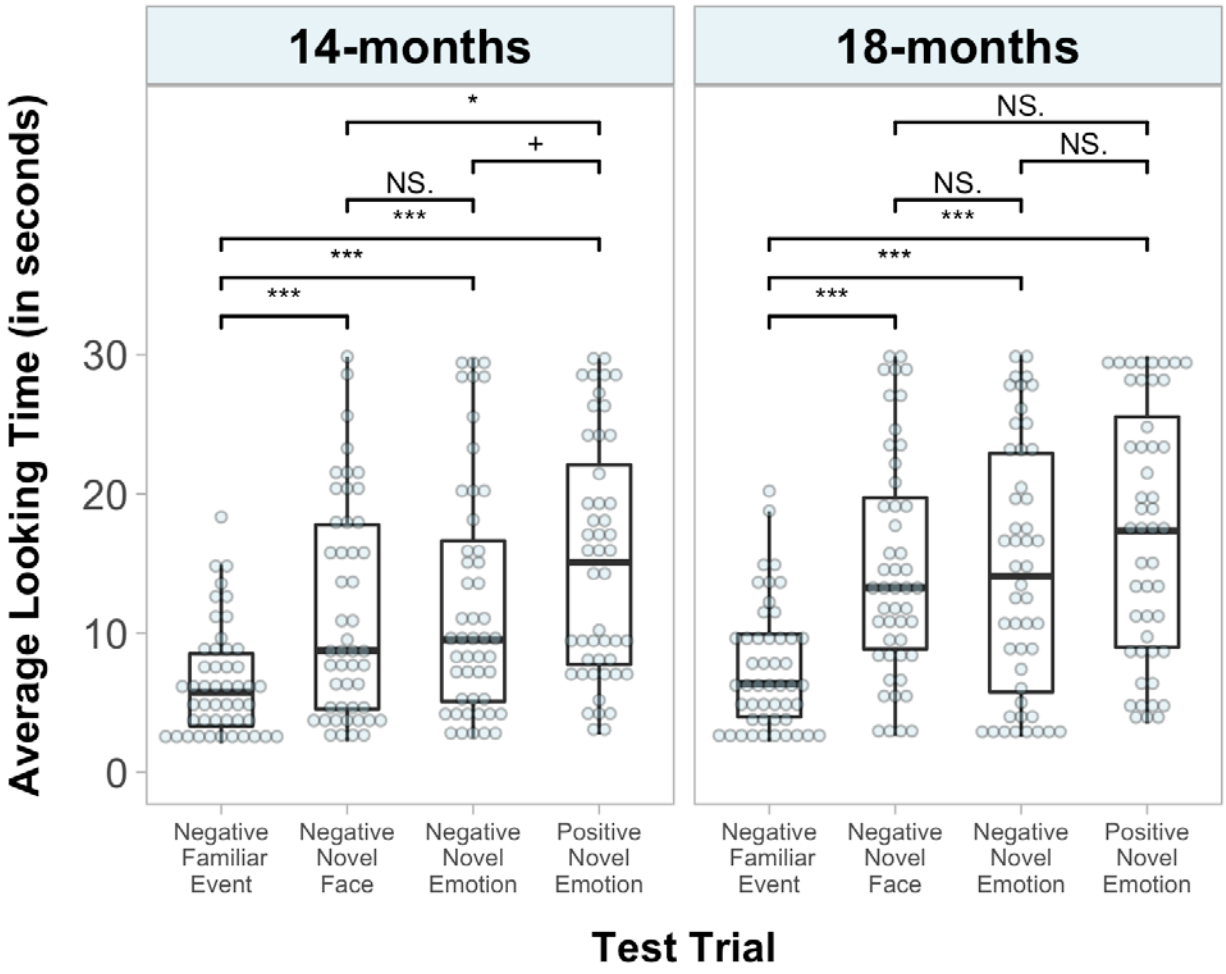
Supplementary Table 1. Mean looking times (and standard deviations) in seconds during the test trials. Conditions are Anger-Sad (AS) and Disgust-Sad (DS).

Exp.	**Condition**	**Age**	***n***	**Negative Familiar Event**	**Negative Novel Face**	**Negative Novel Emotion**	**Positive Novel Emotion**
1	AS, DS	14	48	6.54 (3.90)	11.42 (7.74)	12.26 (8.57)	15.16 (8.51)
1	AS, DS	18	48	7.46 (4.51)	14.71 (8.04)	14.55 (9.09)	17.25 (9.03)
1	DS	14	24	6.10 (2.80)	11.38 (8.12)	9.94 (7.16)	13.58 (8.11)
2	DS	14	32	10.68 (5.85)	11.77 (6.92)	10.45 (7.03)	19.18 (7.50)
2	DS	18	32	10.76 (6.51)	13.17 (7.69)	12.94 (6.96)	17.02 (8.53)
2	AS	14, 18	48	10.00 (6.75)	16.69 (8.98)	17.97 (8.79)	18.17 (8.37)
2	DS	14, 18	48	10.46 (6.48)	11.45 (6.31)	11.76 (7.11)	17.81 (7.92)

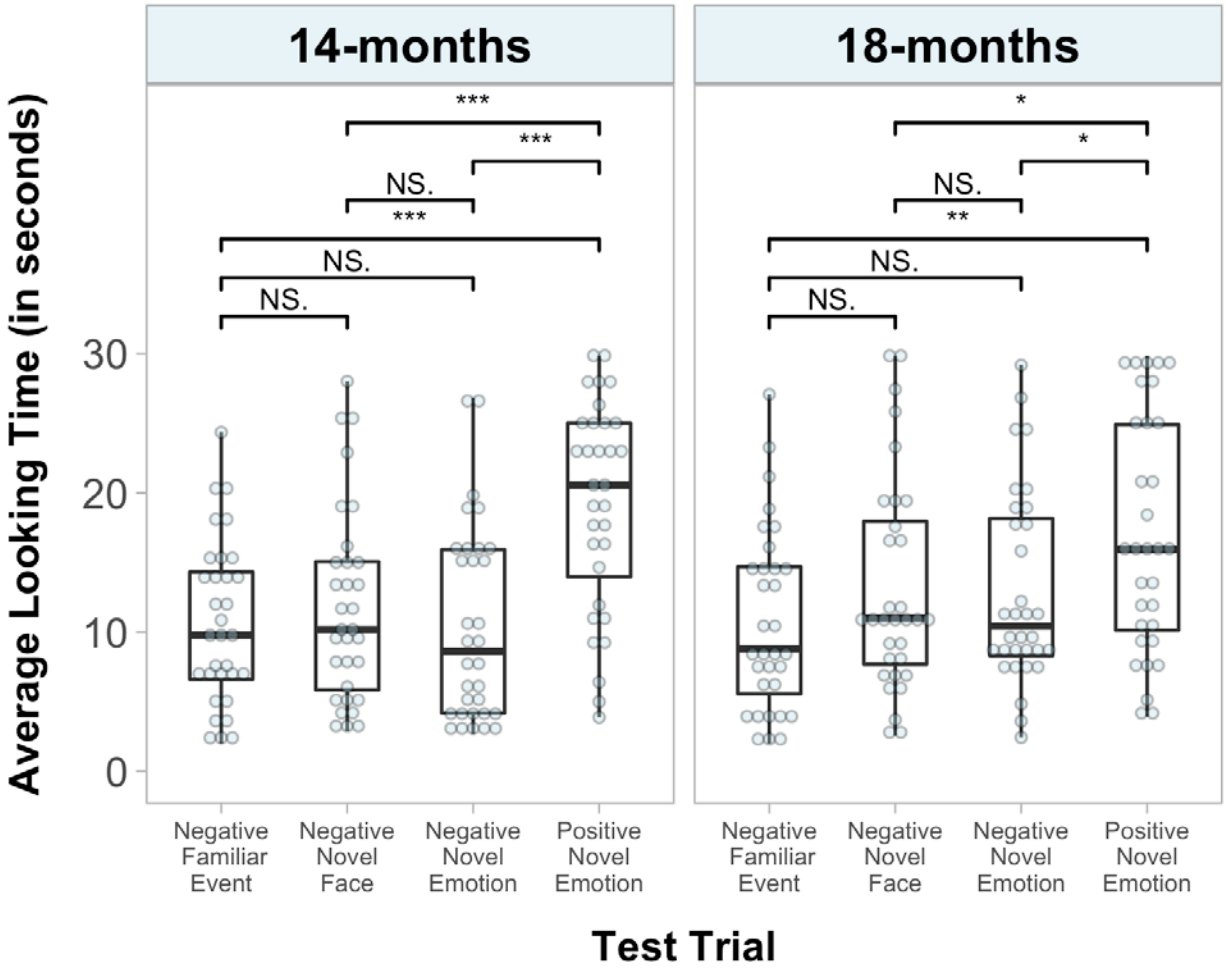
Supplementary Table 2. Correlations between receptive and expressive vocabulary and emotion difference scores in Experiment 2, *$p < .05$, ^$p < .10$.

	14-month-olds		18-month-olds	
	Receptive	Expressive	Receptive	Expressive
Anger-Sad				
Nov Neg Face	-.01	-.38*	-.20	-.06
Nov Neg Emo	.13	-.35^	-.27	-.17
Nov Pos Emo	.19	-.36*	-.12	-.08
Disgust-Sad				
Nov Neg Face	-.17	-.19	.22	.22
Nov Neg Emo	-.07	-.04	.22	.10
Nov Pos Emo	-.21	-.26	.06	-.08

Supplementary Figure 1. Experiment 1: Infants' mean total looking time (and *SE*) during the test trials, *$p < .05$, ***$p < .001$, +$p<.10$, for pairwise comparisons between events



Supplementary Figure 2. Experiment 2: Infants' mean total looking time (and *SE*) during the test trials in the *Disgust-Sad* condition, **p* < .05, ***p* < .01, ****p* < .001, for pairwise comparisons between events



Supplementary Figure 3. Experiment 2: Infants' (first 24 tested per Age and Condition) mean total looking time (and *SE*) during the test trials, ***$p < .001$, for pairwise comparisons between events