Supplementary Materials

Supplementary Analysis 1: Reporting of exact statistics for null effects.

Experiment 1:

Baseline/different: **3 months**: trial type: F(1, 14) = .174, *p* = .683; trial type X habituation: F(1, 14) = .343, *p* = .567; habituation F(1, 14) = 4.03, *p* = .064. **6 months:** trial type: F(1, 14) = 3.48, *p* = .083; trial type X habituation: F(1, 14) = 1.29, *p* = .276; habituation F(1, 14) = .676, *p* = .425. **8 months**: trial type X habituation: F(1, 30) = 3.2, *p* = .082.

Same/different: **3months:** trial type: F(1, 14) = .150, *p* = .704; trial type X habituation: F(1, 14) = .011, *p* = .916; habituation: F(1, 14) = 3.9, *p* = .068 **6 months:** trial type X habituation: F(1, 14) = 2.6, *p* = .129; habituation: F(1, 14) = .281, *p* = .604 **8 months:** trial type X habituation: F(1, 14) = .786, *p* = .382.

Experiment 2:

Baseline/different: **6 months**: trial type: F(1, 55) = .600, *p* = .442; trial type X habituation: F(1, 55) = 1.7, *p* = .200. **7-12 months:** trial type: F(1, 41) = .085, *p* = .772; habituation: F(1, 41) = 2.4, *p* = .128**. 7-12 months relief group:** t(20) = 2.06, *p* = .053. **7-12 month-old triumph group:** t(21) = 1.19, *p* = .246.

Same/different: **6 months:** trial type X habituation: F(1, 55) = 3.9, *p* = .053, habituation: F(1, 55) = 3.5, *p* = .067. **7-12 months:** habituation: F(1, 41) = .952, *p* = .335. **7-12 month-old triumph group:** t(21) = .498, *p* = .624.

Experiment 3:

Baseline/different: **6 months:** trial type X habituation: F(1, 43) = 1.06, *p* = .309; habituation: F(1, 43) = .003, *p* = .956. **7-12 months:** habituation: F(1, 53) = .875, *p* = .354.

Same/different: **6 months:** trial type: F(1, 44) = 2.09, *p* = .155; trial type X habituation: F(1, 44) < .001, *p* = .991; habituation: F(1, 43) = .963, *p* = .332. **7-12 months:** trial type X habituation: F(1, 53) = 3.03, *p* = .087. **7-9 months**: trial type: F(1, 22) = .103, *p* = .751; trial type X habituation: F(1, 22) = .287, *p* = .598; habituation: F(1, 22) = 2.64, *p* = .118. **9-12 months**: trial type X habituation: F(1, 29) = 3.24, *p* = .082; habituation: F(1, 22) = 2.06, *p* = .162.

Supplementary Analysis 2: Explanation of run time analysis and comparison of run time and offline looking time data

Note: Unfortunately, there was a software upgrade partway through data collection for Experiments 2 and 3 which contained a bug that was not detected until later. This bug caused the run-time, rather than looking time to be recorded in the data files. In other words, the short looks by infants away from the stimulus were not excluded from the "looking time" calculation. Testing after the fact confirmed that habituation criteria determined by the software were based on the underlying looking times, not run-time – but run-time, not looking time was not reported in the output file. Although the maximum trial length was 15 s, the time reported in the data file could be slightly longer than 15 s (as trials did not cut off mid-vocalization). Run time values used in the analysis were therefore capped at 15 s (i.e. run times longer than 15 s were analyzed as 15 s). Videorecording of all studies was implemented as a laboratory policy partway through the study, so it was possible to code a frame-by-frame analysis of the test trials for this subset of the data (84 participants for Experiment2, 77 participants for Experiment 3). This

allowed us to verify that looking time and run-time were highly correlated, as expected. Correlations ranged from .924 to .961 across the four trials, *p*'s < .001. Therefore, we report primarily the run-time data, but a separate analysis was run with the looking time data coded offline, yielding similar findings (see Supplementary Table 1). The additional infants excluded from the run-time subsample as noted in the Participants section were excluded because their data were collected prior to the software upgrade and therefore could not be included in the run-time analysis.

Table 1: Statistical tests of the same/different analysis for the offline looking time coding of Experiments 2 and 3. Data are reported separately for 6 month olds. The older age ranges are combined due to limited participant numbers. For each group a 2 (Trial type) X 2 (Habituation Emotion Group) mixed ANOVA was run. As with the primary analyses, follow-up t-tests for each group were run when a significant interaction (Group X TT) effect was found.

	Trial type	Group X TT	Group
Experiment 2	F(1, 49) = 5.286, *p* = .026	F(1, 49) = 5.524, *p* = .023	F(1, 49) = 4.058, *p* = .049
6 months			
Relief N = 29	t(28) = 3.339, *p* = .002		
Triumph N = 22	t(21) = -.037, *p* = .971		
Experiment 2	F(1, 31) = 9.769, *p* = .004	F(1, 31) = 10.770, *p* = .003	F(1, 31) = .053, *p* = .820
7-24 months			
Relief N = 16	t(15) = 3.796, *p* = .002		
Triumph N = 17	t(16) = -.140, *p* = .890		
Experiment 3	F(1, 35) = .440, *p* = .511	F(1,35) = .000, *p* = .992	F(1, 35) = .085, *p* = .772
6 months			
Relief N = 18			
Triumph N = 19			
Experiment 3	F(1,38) = 3.464, *p* = .070	F(1,38)= 4.172, *p* = .048	F(1,38)=4.172, *p* = .048
7-24 months			
Relief N = 21	t(20) = 2.451, *p*= .024		
Triumph N = 19	t(18) = -.158, *p* = .876		

As can be seen in Supplementary Table 1, the general pattern of results is similar to those reported in the main manuscript. However, they differ in that the offline coding analysis found interaction effects in the 6-month-old groups in Experiment 2 and the 7-24 month old age group in Experiment 3 that were not found in the analyses in the main manuscript. Consistent with the 7-12-month-olds in Experiment 2 in the main analysis, the individual t-tests for these groups found discrimination of the same and different test trials only in the relief group. As can be seen above in Supplementary Analysis 1, these interaction effects in the main analysis, while not statistically significant, also approached significance (for the 6 month olds in Experiment 2, *p* = .053, for the 7-12 month olds in Experiment 3, *p* = .087). In sum, for both the main analysis and this analysis with the offline coding, the effect in the triumph group is not as robust across ages and test conditions as in the relief group. Implications of this asymmetry are explored in the General Discussion of the main manuscript.

Supplementary Analysis 3: Reporting of analysis of data from 12-24 month olds in Experiments 2 and 3.

Due to the small sample size of infants across the 12-24 month age range in Experiments 2 and 3, these infants were excluded from the primary analysis and are presented here. For each study, two 2 X 2 repeated measures ANOVAs with trial type (baseline/different or same/different) as within-subjects and habituation emotion group as between-subjects were run, as with the other age groups.

Experiment 2: For the baseline/different analysis, there were no significant effects, Trial type: $F(1, 13) = 2.08$, $p = .173$; Trial type X group: $F(1, 13) = 1.59$, $p = .229$; group: $F(1, 13) = .131$, $p = .724$. For the same/different analysis, there were again no significant effects, Trial type: $F(1, 13) = 3.82$, $p = .072$; Trial type X group: $F(1, 13) < .001$, $p = .986$; group: $F(1, 13) = 1.83$, $p = .199$. However, inspection of the means (see Supplementary Table 2) finds that the infants' pattern of looking is higher to the different trials. Given the small N in this sample, it is likely that there was not sufficient power to show a significant main effect of trial type.

Experiment 3: For the baseline/different analysis, there was a significant main effect of trial type, $F(1, 35) = 23.95$, $p < .001$, and a significant trial type X group interaction, $F(1, 35) = 4.68$, $p = .037$, but no main effect of group, $F(1, 35) = .175$, $p = .679$. For the same/different analysis, there were no significant main effects, Trial type: $F(1, 35) = 1.04$, $p = .315$; group: $F(1, 35) = .006$, $p = .939$. However, there was an interaction effect, $F(1, 35) = 5.4$, $p = .026$. Therefore, each group was examined individually with a paired t-test. Neither group showed significantly different run times between same and different trials, Relief: $t(17) = 2.01$, $p = .060$; Triumph: $t(18) = 1.14$, $p = .270$. These findings show a different pattern of results from those with the 7-12-month-olds in the main analysis, and suggest that further development may be taking place in infants' perception on non-linguistic expressions of emotion. However, given the small sample size across a broad age range, it is difficult to make specific conclusions regarding the exact nature of these changes.

Table 2: Means (standard deviations) of run time (Experiments 2 and 3) for each experiment for the 12-24 month old age range. "Pre" is the average across the first three trials of the habituation phase, "baseline" is the average across the last three trials of the habituation phase. "Same" and "different" are means for the test trials, separately for each habituation group."N_T" is number of participants in the triumph group, "N_R" is the number of participants in the relief group.

Experiment	Pre (s)	Baseline (s)	habituation t-statistic p-value	Relief group same (s)	Relief group different (s)	Triumph group same (s)	Triumph group different (s)
Experiment 2 $N_R = 6$, $N_T = 9$	13.1 (2.4)	9.2 (1.9)	$t(14) = 6.7$ $p < .001$	9.4 (3.0)	11.5 (3.8)	7.7 (3.0)	9.8 (3.0)

Experiment 3 $N_R = 18, N_T = 19$	12.5 (3.0)	8.4 (2.0)	$t(36) = 10.6$ $p < .001$	9.7 (4.1)	11.7 (3.0)	11.0 (3.6)	10.2 (3.6)

Supplementary Analysis 4: Exclusion of infants who failed to habituate

Although there some differences in individual tests of significance, the overall pattern of findings is similar (see Supplementary Tables 3 and 4). There are two cases in which the habituation emotion group main effects differ, but this effect is not directly relevant to our question of infants' discrimination of the trial types. Another difference is the lack of interaction at 8 months in Experiment 2 in the baseline/different test in this supplementary analysis which was found in the main analysis. However, in the main analysis, despite the interaction, separate t-tests did not find effects of trial type for either group, so the conclusion is the same in either case. Another difference is the interaction effect in Experiment 1 for the 8 month olds in the baseline/different analysis that was not found in the main analysis. In this case, separate t-tests did find that the significant effect was restricted to the relief group. This is consistent with our overall finding that effects in the triumph group overall are less robust, and also with the lack of such an effect for the 7-12 month olds in Experiment 2.

The one instance in which the results diverge more substantially was for the 6-month-olds in Experiment 2. As can be seen in Supplementary Table 3, all three effects in the baseline/different analysis were different from those reported in the main manuscript, and in addition an interaction effect was found in the same/different analysis which was not found in the analysis main manuscript. However, the overall interpretation is not radically different from the data reported in the main manuscript. In the baseline/different analysis here, unlike in the main analysis, we do find evidence that 6-month-olds dishabituate to the different trials compared with the baseline trials. However, this effect is restricted the relief group, again supporting the relief/triumph asymmetry found in several of our analyses. The analysis excluding infants who failed to habituate therefore suggests a slightly more robust discrimination on the part of the 6-month-olds, in that one additional test/sub-group combination supports this discrimination. Conversely, in the same/different analysis, the interaction effect found here in Experiment 2 approached, but did not reach, significance in the main analysis. Therefore, the main conclusions reached in Experiment 2, that 6-month-olds discriminate relief and triumph, but that the effect is more robust for the relief group, are also reached for this subsample, despite the different pattern of significance.

Table 3: Statistical tests for the baseline/different analyses excluding infants who failed to habituate. Tests for which the significance differed from that reported in the manuscript are in bold.

	Trial type	Group X TT	Group
Experiment 1			
3 months Relief N = 8 Triumph N = 7	F(1, 13) = .528, p = .480	F(1, 13) = .796, p = .388	**F(1, 13) = 5.75, p = .032**
6 months	F(1, 14) = 3.48, p = .083	F(1, 14) = 1.29, p = .276	F(1, 14) = .676, p = .425

Relief N = 8			
Triumph N = 8			
8 months	F(1, 28) = 25.4, $p < .001$	**F(1, 28) = 5.56, $p = .026$**	F(1, 28) = 7.32, $p = .011$
Relief N = 16	**t(15) = 6.35, $p < .001$**		
Triumph N = 14	**t(13) = 1.60, $p = .133$**		
Experiment 2			
6 months	**F(1, 37) = 4.86, $p = .034$**	**F(1, 37) = 4.13, $p = .049$**	**F(1, 37) = 3.54, $p = .068$**
Relief N = 19	**t(18) = 3.49, $p = .003$**		
Triumph N = 20	**t(19) = .110, $p = .914$**		
7-12 months	F(1, 28) = .662, $p = .423$	**F(1, 28) = 1.46, $p = .237$**	F(1, 28) = 2.52, $p = .124$
Relief N = 15			
Triumph N = 15			
Experiment 3			
6 months	F(1, 36) = 12.7, $p = .001$	F(1,36) = 1.20, $p = .281$	F(1, 36) = .452, $p = .505$
Relief N = 19			
Triumph N = 19			
7-12 months	F(1,44) = 72.0, $p < .001$	F(1,44)= 9.21, $p = .004$	**F(1,44)=4.172, $p = .048$**
Relief N = 23	t(22) = 12.79, $p < .001$		
Triumph N = 23	t(22) = 3.05, $p = .006$		

Table 4: Statistical tests for the same/different analyses excluding infants who failed to habituate.

	Trial type	Group X TT	Group
Experiment 1			
3 months	F(1, 13) = .748, $p = .403$	F(1, 13) = .329, $p = .576$	F(1, 13) = 3.57, $p = .081$
Relief N = 8			
Triumph N = 7			
6 months	F(1, 14) = 11.47, $p = .004$	F(1, 14) = 2.61, $p = .129$	F(1, 14) = .281, $p = .604$
Relief N = 8			
Triumph N = 8			
8 months	F(1, 28) = 9.87, $p = .004$	F(1, 28) = 1.53, $p = .226$	F(1, 28) = 10.6, $p = .003$
Relief N = 16			
Triumph N = 14			
Experiment 2			
6 months	F(1, 37) = 9.31, $p = .004$	**F(1, 37) = 5.28, $p = .027$**	F(1, 37) = 2.83, $p = .101$
Relief N = 19	**t(18) = 4.51, $p < .001$**		
Triumph N = 20	**t(19) = .474, $p = .641$**		
7-12 months	F(1, 28) = 5.62, $p = .025$	F(1, 28) = 4.81, $p = .037$	F(1, 28) = .035, $p = .853$
Relief N = 15	t(14) = 2.78, $p = .015$		
Triumph N = 15	t(14) = .155, $p = .879$		
Experiment 3			
6 months	F(1, 36) = .674, $p = .417$	F(1,36) = .060, $p = .808$	F(1, 36) = .908, $p = .347$

Relief N = 19 Triumph N = 19			
7-12 months Relief N = 23 Triumph N = 23	F(1,44) = 4.68, p = .036	F(1,44)= 1.05, p = .312	F(1,44)=3.40, p = .072
7-9 months Relief N = 8 Triumph N = 10	F(1,16) = .005, p = .946	F(1,16)= .004, p = .950	F(1,16)=1.97, p = .180
10-12 months Relief N = 15 Triumph N = 13	F(1,26) = 9.89, p = .004	F(1,26)= 2.06, p = .163	F(1,26)=1.25, p = .273

Supplementary Analysis 5: Examination of the influence of language background

Since our sample in Experiments 2 and 3 contained a large number of infants with multilingual experience, we examined whether this experience might influence the infants' ability to discriminate the same and different trials. In particular, we hypothesized that infants with experience with more than one language might be better able to detect the within-emotion similarities from the habituation to the test sample in Experiment 3, despite the variance created by the two linguistic/cultural contexts. We divided our sample into two groups. Language group was assigned as "monolingual" if the infant heard 20% or less of languages other than English and "multilingual" if the infant heard more than 20% of one or more languages other than English by parental report. To reduce the number of factors in the analysis, a difference score was created by subtracting the same test trials from the different test trials (i.e. different – same). Using this difference score as the dependent variable, a univariate ANOVA was run at each age group for Experiments 2 and 3 with habituation emotion group and language background as fixed factors.

Table 5: Statistical tests examining whether language background influenced infants' responding. "N_T" is number of participants in the triumph group, "N_R" is the number of participants in the relief group.

	Language	Language X Group	Group
Experiment 2			
6 months Mono N_R= 25 N_T=18 Multi N_R= 8 N_T=6	F(1, 53) = .855, p = .359	F(1, 53) = .240, p = .626	F(1, 53) = 2.10, p = .153
7-12 months Mono N_R= 10 N_T=8 Multi N_R= 11 N_T=14	F(1, 39) = 1.41, p = .242	F(1, 39) = .123, p = .728	F(1, 39) = 8.90, p = .005
Experiment 3			
6 months Mono N_R= 15	F(1, 42) = .254, p = .617	F(1, 42) = 1.02, p = .317	F(1, 42) = .425, p = .518

	N_T=21		
Multi	N_R= 7		
	N_T=3		
7-12 months	F(1, 51) = 4.64, *p* = .036	F(1, 51) = .133, *p* = .717	F(1, 51) = 2.76, *p* = .103
Mono	N_R= 16		
	N_T=18		
Multi	N_R= 10		
	N_T=11		

As can be seen from Supplementary Table 5, the only instance in which language background showed a significant effect on the relative preference for same and different trials was in Experiment 3, 7-12 month old age group. The monolingual infants showed a stronger preference for the different over same trials (Mean = 2.0 s, SD = 3.2 s) than the multilingual infants (Mean = .27 s, SD = 2.3 s). Notably, this suggests that the multilingual experience of our sample cannot be driving the effects we found in Experiment 3. It is not clear why the monolingual infants might be at an advantage. One possibility is that bilingual infants may be slightly slower to develop speech processing capabilities, and may have more difficulty at this age in attending to multiple dimensions of variance in acoustic input. However, it is important to state this this finding must be viewed as exploratory – there may be other differences between our two groups (such as socio-economic differences) that may be driving the difference seen here.