**Raw data (non log-transformed) statistical analyses**

*Combined analysis*

A mixed-model ANOVA was first conducted on looking time duration for both the pitch and formant conditions, with trial type as a within-subjects factor (inconsistent, consistent) and presentation order (inconsistent first, consistent first) and condition (pitch, formant) as between-subjects factors. These analyses revealed a significant main effect of trial type such that, as predicted, infants looked longer at the inconsistent sound/size pairing trials (*M* = 14.04, *SD* = 10.21) than the consistent sound/size pairing trials (*M* = 10.45, *SD* = 8.86; *F*(1,28) = 8.84, *p* = .006, *η*p2 = .24). There was no difference found between the pitch and formant conditions *F*(1,28) = .03, *p* = .875, *η*p2 = .001), and likewise no main effect of presentation order (*F*(1,28) = .38, *p* = .545, *η*2 = .013). However, there was also a trial type x presentation order interaction (*F*(1,28) = 10.18, *p* = .003, *η*2 = .27), meaning that the effect of trial type differed depending on which trial order infants saw.

Because of the significant interaction, it was necessary to examine the effects of trial type broken up by each of the presentation orders. Infants who saw the inconsistent trial block first looked substantially longer at the inconsistent size/sound pairings than at the consistent size/sound pairings (inconsistent trials *M* = 16.95, *SD* = 9.01, consistent trials *M* = 9.51, *SD* = 6.69; *t*(15) = 3.33, *p* = 0.005, *r* = .65). In contrast, infants who saw the consistent trial block first did not show a looking time difference across the inconsistent versus consistent pairings (inconsistent trials *M* = 11.13, *SD* = 10.77, consistent trials *M* = 11.40, *SD* = 10.75; *t*(15) = -.29, *p* = 0.775, *r* = .08). Thus, the interaction reflects that infants who saw the inconsistent events first drove the difference between inconsistent and consistent event types.

*Analysis broken up by pitch and formant*

To examine the independent effects of pitch versus formant, the two conditions were also analyzed separately using a 2 x 2 mixed model ANOVA with trial type as a within-subjects factor and order as a between-subjects factor. Again, a trial type X presentation order interaction was evident in the pitch condition, *F*(1, 14) = 9.91, *p* = .01, *η*p2 = .42, and in the formant condition, although the interaction effect was weaker for formant, F (1, 14) = 1.90, *p* = .19, *η*p2 = .12. Given these interactions with presentation order, it was necessary to examine the effects of trial type within each condition broken up by each of the different trial order.

Paired-samples t-tests indicated that infants who saw the inconsistent trial block first looked longer at the inconsistent size/sound pairings than the consistent size/sound pairings in the pitch condition (inconsistent trials *M* = 18.83, *SD* = 8.41, consistent trials *M* = 9.16, *SD* = 3.24; *t*(7) = 3.11, *p* = 0.02, *r* = .76) and marginally longer in the formant condition (inconsistent trials *M* = 15.08, *SD* = 9.77, consistent trials *M* = 9.85, *SD* = 9.23; *t*(7) = 1.63, *p* = 0.15, *r* = .52). No looking time differences were found for infants who saw the consistent trial block first, (*pitch*: inconsistent trials *M* = 9.49, *SD* = 8.68, consistent trials *M* = 10.50, *SD* = 7.97; *t*(7) = -.74, *p* = 0.48, *r* = 0.27; *formant*: inconsistent trials *M* = 12.77, *SD* = 12.93, consistent trials *M* = 12.30, *SD* = 13.50; *t*(7) = 0.38, *p* = 0.71, *r* = 0.14).

**Statistical analyses with failure-to-watch trials included (log-transformed)**

*Combined analysis*

A mixed-model ANOVA was first conducted on looking time duration for both the pitch and formant conditions, with trial type as a within-subjects factor (inconsistent, consistent) and presentation order (inconsistent first, consistent first) and condition (pitch, formant) as between-subjects factors. These analyses revealed a significant main effect of trial type such that, as predicted, infants looked longer at the inconsistent sound/size pairing trials (*M* = .94, *SD* = .38) than the consistent sound/size pairing trials (*M* = .83, *SD* = .35; *F*(1,28) = 4.30, *p* = .047, *η*p2 = .13). There was no difference found between the pitch and formant conditions *F*(1,28) = .003, *p* = .956, *η*p2 < .001), and likewise no main effect of presentation order (*F*(1,28) = .809, *p* = .376, *η*2 = .028). However, there was also a trial type x presentation order interaction (*F*(1,28) = 9.27, *p* = .005, *η*2 = .249), meaning that the effect of trial type differed depending on which trial order infants saw.

Because of the significant interaction, it was necessary to examine the effects of trial type broken up by each of the presentation orders. Infants who saw the inconsistent trial block first looked substantially longer at the inconsistent size/sound pairings than at the consistent size/sound pairings (inconsistent trials *M* = 1.07, *SD* = .33, consistent trials *M* = .80, *SD* = .31; *t*(15) = 3.28, *p* = 0.005, *r* = .65). In contrast, infants who saw the consistent trial block first did not show a looking time difference across the inconsistent versus consistent pairings (inconsistent trials *M* = .81, *SD* = .39, consistent trials *M* = .86, *SD* = .38; *t*(15) = -.79, *p* = 0.444, *r* = .20). Thus, the interaction reflects that infants who saw the inconsistent events first drove the difference between inconsistent and consistent event types.

*Analysis broken up by pitch and formant*

To examine the independent effects of pitch versus formant, the two conditions were also analyzed separately using a 2 x 2 mixed model ANOVA with trial type as a within subjects factor and order as a between subjects factor. Again, a trial type X presentation order interaction was evident in the pitch condition, *F*(1, 14) = 11.92, *p* = .004, *η*p2 = 0.46, and in the formant condition, although the interaction effect was weaker for formant, *F*(1, 14) = 1.25, *p* = .28, *η*p2 = .08).

Paired-samples t-tests on the log-transformed looking time data showed that infants who saw the inconsistent trial block first looked longer at the inconsistent size/sound pairings than the consistent size/sound pairings in the pitch condition (inconsistent trials *M* = 1.17, *SD* = .22, consistent trials *M* = .85, *SD* = .22; *t*(7) = 3.40, *p* = 0.011, *r* = .79) and marginally longer in the formant condition (inconsistent trials *M* = .96, *SD* = .40, consistent trials *M* = .76, *SD* = .39; *t*(7) = 1.54, *p* = 0.168, *r* = .50). No looking time differences were found for infants who saw the consistent trial block first, (*pitch*: inconsistent trials *M* = .70, *SD* = .36, consistent trials *M* = .83, *SD* = .40; *t*(7) = -1.42, *p* = 0.198, *r* = 0.47; *formant*: inconsistent trials *M* = .91, *SD* = .41, consistent trials *M* = .91, *SD* = .39; *t*(7) = 0.30, *p* = 0.78, *r* = 0.11).

**Exploratory analyses of counterbalanced stimuli variables not of theoretical interest (null results):**

A mixed-model ANOVA was conducted on looking time duration for both the pitch and formant conditions, with trial type as a within-subjects factor (inconsistent, consistent) and the following as a between-subjects factor:

*Order in which creature appeared first:*

Whether the larger or smaller creature was first presented had no main effect on looking time, *F*(1,30) = .195, *p* = .662, *η*p2 = .006), and did not interact with trial type, *F*(1,30) = 1.42, *p* = .243, *η*p2 = .045).

*Box color and creature size pairing:*

Whether the larger creature was assigned to the green box, and the smaller to the blue box, or vice a versa, had no main effect on looking time, *F*(1,30) = .003, *p* = .957, *η*p2 = .000), and did not interact with trial type, *F*(1,30) = 1.37, *p* = .251, *η*p2 = .044).

*Order of sound presentation:*

Whether the lower frequency sound was presented first or second had no main effect on looking time, *F*(1,30) = 2.38, *p* = .134, *η*p2 = .073), and did not interact with trial type, *F*(1,30) = .296, *p* = 591, *η*p2 = .010).

*Order of which box color moved first:*

Whether the green box was shown moving first or second had no main effect on looking time, *F*(1,30) = .004, *p* = .952, *η*p2 = .000), and did not interact with trial type, *F*(1,30) = .353, *p* = .557, *η*p2 = .012).

*Switch type:*

Whether size/sound consistency and inconsistency was achieved by switching the creature or the sound had no main effect on looking time, *F*(1,30) = .003, *p* = .957, *η*p2 = .000), and did not interact with trial type, *F*(1,30) = 1.37, *p* = .251, *η*p2 = .044).