**General methodology**

**Apparatus**

The events were presented on a lit stage (grey) presenting an aperture 64 cm wide x 40 cm high x 43 cm deep, in a dimly lit testing room. An 8.5 cm high black screen located 31.5 cm behind the front of the stage could be rotated to conceal the toys, and a black blind could be lowered to conceal the whole stage. The objects were two 12.5 (H) x 9.5 (W) cm toy hedgehogs and 13 (H) x 9.5 (W) cm snails that squeaked when pressed. A 30 (W) x 10 (D) cm rotating platform with four surfaces arranged in a square and with magenetic mounts, allowed different configurations of objects to be rotated into place rapidly. There was a 1.5cm gap in between the screen and the rotating platform. The experimenter presenting the toys wore long maroon gloves. A video camera, placed at the top center of the stage recorded infants’ head and eyes for live recording of preferential looking and for subsequent reliability testing by a naïve observer. A remote-optics corneal reflection eye-tracker (Tobii X60), located below and at the mid-point of the stage, was used to collect fixation data. A 21 inch plasma display was mounted immediately behind the stage and, prior to testing, each infant’s point of gaze at 5 locations was calibrated in standard fashion by presenting attention-getting videos. A scene camera was also placed behind and just above the infants’ head to record image on the stage.

**Procedure**

Infants sat either in an infant high seat or on a caregiver’s lap, approximately 60 cm from the front edge of the stage. In the latter case, the caregiver’s eyes were above the stage and s/he could not see the displays. One researcher controlled the eye-tracker and recorded preferential looking during familiarization and test trials, and a second researcher manipulated the objects. After gaze calibration, the procedure was close to Wynn (1992) and Bremner et al. (2017). Infants saw three pretest trials representing the correct outcome and the two incorrect outcomes tested in the pair of conditions run, in counterbalanced order across infants. The blind was raised to reveal either one or two toys and the observer recorded looking at the toy(s). Toys were placed 33 cm behind the front edge of the stage. When one was presented it was placed with midline 7.5 cm to the right of platform centre, and when two were presented, the other was 7.5 cm to left of platform centre.. The trial continued until the infant had accumulated at least 2 s looking time and looked away from the display for 2 s or more. The blind was then lowered and the procedure was repeated for the other two displays. Two blocks of four test trials were then presented, involving addition or subtraction, and in which the outcome alternated between correct and incorrect for the two conditions in question. The test trial sequences are illustrated in Figure 1. In cases in which two objects were placed close together, one was placed with mdline 7.5 cm from the centre of the platform and the other was presented with midline at the centre of the platform. In addition conditions the experimenter’s gloved hand emerged at one side above the screen (conterbalanced across infants) holding a toy that she squeaked to capture the infant’s attention. She then moved the toy, still squeaking, and placed it on one of the locations used during correct outcome familiarisation. She then slowly withdrew her hand whereupon the screen was raised to hide the toy. This event, from appearance of the toy to withdrawal of the hand, took approximately 5 s. Her other hand then reappeared from the opposite side, above the screen, clutching another identical squeaking toy. When she had the infant’s attention the experimenter placed the toy in the other location used during appropriate outcome familiarization, raised her hand, clasped and unclasped it to emphasize that it was empty, and then slowly withdrew it, whereupon the screen was lowered to reveal the appropriate or inappropriate outcome. The period from appearance to disappearance of her hand took approximately 6 s. Again, test trials continued until the infant had accumulated at least 2 s looking time and looked away from the display for 2 s or more.

In subtraction conditions, on test trials the experimenter placed two squeaking toys, consecutively, on the stage in the locations used in the two toy familiaristation trial (with LR order conterbalanced across infants), an event that took approximately 9 s. Following the raising of the screen her empty hand reappeared above the screen, lowered to one side of the screen and reappeared holding one toy which she squeaked above the screen and withdrew screen left, an event that took approximately 6 s, followed by lowering of the screen to reveal the appropriate or inappropriate outcome for the condition.

Incorrect outcomes were accomplished by rotation of the triangular platform wheel so that the desired outcome appeared when the screen was lowered. The observer who recorded infants’ looking was aware of which condition (addition or subtraction) the infant was in but was unaware on each test trial of whether the outcome was possible or impossible. Preferential looking (violation of expectancy) data were recorded on a Mac G4 using Habit software (Cohen, Atkinson, & Chaput, 2004).

Eye tracker data are dwell times in three areas of interest 6cm wide and 14 cm high extending upwards from the object platform and centred on the three possible positions of the objects (labelled 'left', 'centre', & 'right') in data files.