Supplementary Methods

Arousal measures

Equipment

Electro-cardiogram (ECG), electro-dermal activity (EDA) and accelerometry were recorded using a BioPac™ (Santa Barbara, CA) recording at 1000Hz. ECG was recorded using disposable Ag-Cl electrodes placed in a modified lead II position. EDA was recorded using two EDA (Isotonic Gel) snap electrodes placed on the plantar surface of the foot (Ham & Tronick, 2008). A triaxial accelerometer 5G was attached to the same foot from which EDA data was recorded. In addition, head velocity data was derived from the head position estimates that are automatically generated during heads-free eyetracking. They were recorded by a Tobii TX300 eyetracker. The process used to extract this data is described in ([blinded for review]).

Experimental stimuli

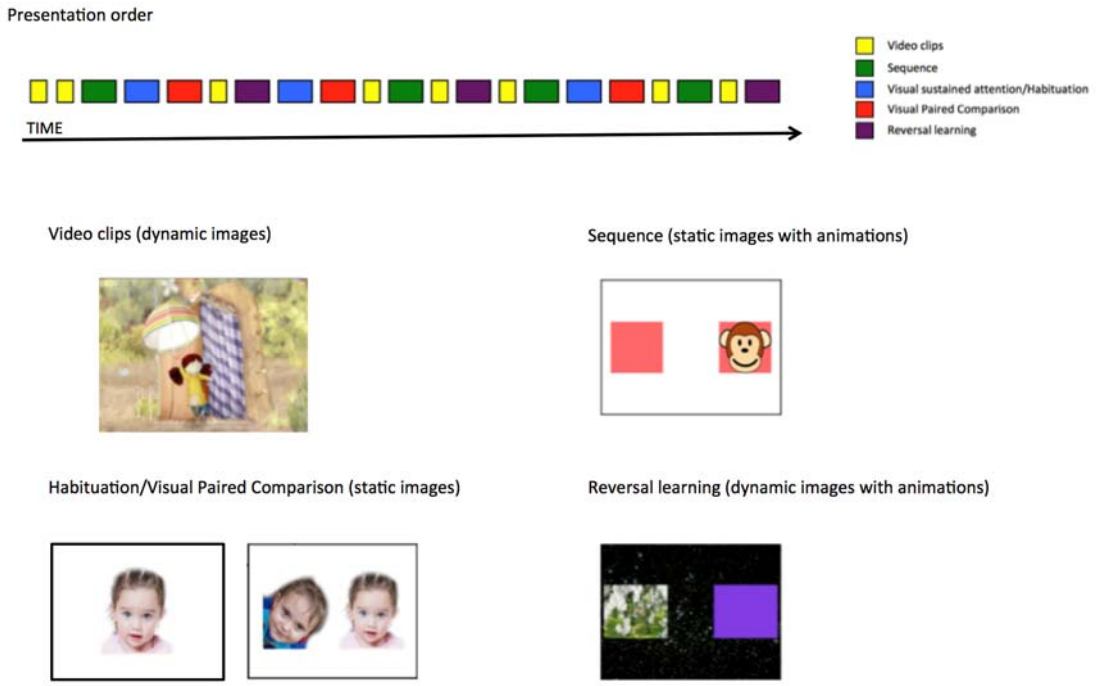

Figure S1 – schematic showing the viewing materials presented.

Data reduction

All data reduction techniques were identical to those used in our previous publications, in which we evaluated patterns of covariance between these different measures ([blinded for review]). More details on data processing techniques are given in these papers. To summarise, briefly:

Heart rate (HR). Automatic r-peak identification was performed by the Acknowledge commercial software package. Automatic artifact rejection was then performed by excluding those beats showing an inter-beat interval of <330 or >750 ms, and by excluding those samples showing a rate of change of inter-beat interval of greater than 80ms between samples. In another paper ([blinded for review]) we report on a comparison of these cleaning techniques with traditional hand-coding which shows a close comparison between the two approaches. Finally, HR data were z-scored and epoched into one-second epochs.

Electrodermal activity (EDA). Again, our approach was similar to that previously used with developmental populations (Ham & Tronick, 2008; Hernes et al., 2002). First, null values were removed from the data using a threshold of 0.1μV. Second, data were log transformed to remove positively skewed values. Third, data were z-scored and epoched into one-second epochs.

Head velocity (HV). First, data samples showing a change in position of more than 0.025 screen units between 120Hz iterations were excluded as being above the maximum possible threshold at which head movement can take place and therefore likely to be artifactual. This threshold corresponds to 2.5% of the screen, representing approximately 1.25 cm in our set-up. Second, data were downsampled to 12Hz by calculating a moving median window. Third, position data were converted to velocity data by taking the first derivative. Fourth, six data streams (three dimensions, two eyes) were collapsed to a single stream. Fifth, data were z-scored and epoched into one-second epochs.

Peripheral accelerometer (PA). Our approach was similar to that previously used with developmental populations (Robertson, et al., 2001)). First, data were filtered to remove high-frequency noise using a Butterworth filter with a cut-off of 0.5Hz. Second, three-dimensional movement data were summed to create a one-dimensional estimate of total movement. Third, median windowing was performed.

Behavioural measures

Look duration (Analyses 1 and 3). For analyses 1 and 3 we wished to record continuous looking data, in 20-minute segments. Infants' looks to and away from the screen were measured during the administration of a mixed testing battery (described in the main text). Look duration was automatically coded based on the eyetracker footage recorded. Looks were treated as starting when the child first looked towards the screen, and ending when the child looked away from the screen. This was derived as the time interval between the moment when the eyetracker data first became available, and it becoming unavailable. Very short sections of missing data (<2 seconds) were interpolated, to cover short periods of missing data due to blinks and other artifactual causes (see further discussion of this in [blinded for review]).

Look duration (Analysis 2). Look duration data presented in Analysis 3 are taken from an infant-controlled habituation task. This was presented in 3 blocks at different stages of the testing protocol. Each block featured one picture (a child's face). Trials commenced with a small (3°) fixation target, presented concurrently with an attention-getter sound; once the infant had looked to this target, the image (subtending

c.10°) was presented. An experimenter, behind a curtain, viewed a live video feed and a feed showing live eyetracking data. When the infant had looked away from the screen, the experimenter pressed a key to signal the end of a trial. The same image was re-presented consecutively until two consecutive looks had taken place that were less than 50% of the longest look recorded that block. The block also ended if the child had accumulated either 12 looks or 120 seconds' looking time without reaching habituation criteria.

Calculation of cross-correlation and auto-correlation (Analysis 1)

For analysis 1 our data consist of continuous time-series data, recorded over a 20-minute testing session. First, these data were time-synchronised and epoched into 1-second epochs. A 1-second epoch duration was selected based on previous work ([blinded for review]). Z-scores were calculated relative to baseline. This was calculated based on the average score obtained across the entire testing session.

The procedure for calculating the cross-correlations between measures was as follows. First, we calculated the average correlation between values obtained for those measures across all epochs, using a Spearman's rank order correlation. This correlation value was calculated independently for each participant, based on all epochs available for that participant (c.1200 per individual). A single average correlation value was then calculated by averaging across participants. This average correlation value is shown as the value at Time 0. Next, correlations were calculated in the same manner at each time-lag by shuffling one measure forwards and backwards in time, relative to the other. The procedure was identical for auto-correlations – except that instead of examining the relationship of two different measures at variable time intervals, the relationship of each measure *to itself* at variable time intervals was assessed.

Calculating the significance levels of the auto-correlations is straightforward, and is done by averaging the significance values of the Spearman's correlations conducted at each time interval. Calculating the significance levels of the cross-correlations is, however, non-trivial, since the values obtained for the cross-correlation (the degree to which the relationship between two measures is present if a time-lag is introduced between them) is confounded by the degree of auto-correlation (the degree to which each measure, considered individually, is fast- or slow-changing) (Clifford et al., 1989; Thiebaux & Zwiers, 1984). This potential problem in cross-correlations can be solved by first calculating the Effective Sample Size (Clifford, et al., 1989; Thiebaux & Zwiers, 1984): at each time interval, the cross-correlation (i.e. the relationship between the two variables) was first calculated, and then the auto-correlation value for each variable (i.e. the relationship of that variable to itself, at that time-lag) was then calculated. The higher of these two values was used to calculate the Effective Sample Size, using the standard formula: $N^* = \frac{N(1-r)}{(1+r)}$, where N^* is the Effective Sample Size, N is the actual sample size and r is the higher of the two auto-correlation values obtained at that time interval for each of the two measures independently (Thiebaux & Zwiers, 1984). The significance level of the cross-correlation obtained was then adjusted based on the Effective Sample Size. In this way we calculated the significance level of the relationship between two variables at a particular time-lag,

independent of the relationship of each variable to itself at that time-lag. An alternative potential solution to this problem is to perform pre-whitening to remove auto-correlation in the data prior to analysing the cross-correlation (Martens et al., 2003); however, this radical procedure can sometimes have unpredicted effects on results, and so we have preferred the Effective Sample Size method here.