Increases in Arousal are More Long-Lasting than Decreases in Arousal: On Homeostatic Failures During Emotion Regulation in Infancy

Article *in* Infancy · May 2018

DOI: 10.1111/infa.12243

CITATIONS
2

READS
159

3 authors, including:

Sam Wass
University of East London
90 PUBLICATIONS **1,004** CITATIONS

SEE PROFILE

INFANCY

THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES

Infancy, 1–22, 2018

ISSN: 1525-0008 print / 1532-7078 online
DOI: 10.1111/infa.12243

WILEY Blackwell

Increases in Arousal are More Long-Lasting than Decreases in Arousal: On Homeostatic Failures During Emotion Regulation in Infancy

Sam V. Wass
School of Psychology
University of East London

Kaili Clackson
Cambridge University

Vicky Leong
Cambridge University and
Nanyang Technological University

In emotion regulation, negative or undesired emotions are downregulated, but there are also opponent processes to emotion regulation—in which undesired emotions are exacerbated dynamically over time by processes that have an amplifying or upregulating impact. Evidence for such processes has been shown in adults, but little previous work has examined whether infants show similar patterns. To examine this, we measured physiological arousal in 57 typical 12 month olds while presenting a 20-min mixed viewing battery. Fluctuations in autonomic arousal were measured via heart rate, electrodermal activity, and movement. We reasoned that if transitions in autonomic arousal are random (stochastic), then (1) arousal would be normally distributed across the session, and (2) episodes where arousal exceeded a certain threshold above the mean should be as long-lived as those where arousal exceeded the same threshold below the mean. In fact we found that (1) heart rate and movement (but not electrodermal activity) were positively skewed, and (2) that increases in arousal have a lower extinction probability than decreases in arousal. Our findings may suggest that increases in arousal are self-sustaining. These patterns are the opposite of the homeostatic mechanisms predicted by naïve approaches to emotion regulation.

Emotion regulation is the adjustment of emotional state or expression to meet goals or to maintain homeostasis or allostasis (Davis, 1958; Eisenberg, Spinrad, & Eggum, 2010; Gross, 1998; Nigg, 2017). It involves both intrinsic (self-regulation) and extrinsic

Correspondence should be sent to Sam V. Wass, School of Psychology, University of East London, Stratford Campus, Water Lane, London E15 4LZ, UK. E-mail: s.v.wass@uel.ac.uk

(interpersonal) processes (Feldman, Greenbaum, & Yirmiya, 1999; Fogel, 1993). Emotion regulation operates via an ongoing interplay between mechanisms of excitation and inhibition as the result of "vertical integration processes" between brainstem, limbic, and cortical brain systems, such that higher systems integrate, elaborate, fine-tune, and serve an inhibitory function for hierarchically lower systems (Posner & Rothbart, 1998; Tucker, Derryberry, & Luu, 2000). In adults, and children, emotion regulation is thought to involve techniques such as cognitive reappraisal (Gross, 1998), attentional redeployment (Gross, 1998), and response modulation (Gross & Thompson, 2007). Although much more primitive in form, emotion regulation also takes place in infants and young children via techniques such as gaze aversion (Aksan & Kochanska, 2004; Rothbart, Ellis, Rueda, & Posner, 2003; Sheese, Rothbart, Posner, White, & Fraundorf, 2008) and self-soothing (Calkins, Dedmon, Gill, Lomax, & Johnson, 2002; Stifter & Fox, 1990). Extrinsic factors, including parent–child affect synchrony, are considered important components of emotion regulation in infancy (Ruth Feldman et al., 1999; Porges & Furman, 2011; Cohn & Tronick, 1988; Fox & Calkins, 2003).

Emotion regulation also has its opposite processes through which negative or undesired emotions can be exacerbated dynamically over time by processes that have an amplifying or upregulating impact (Gross & Thompson, 2007). Although no generic term exists to describe these processes, they might be described as "metastatic" (from the Greek word meta, meaning "beyond"). "Metastatic" processes would be ones in which a small initial perturbation becomes progressively greater over time, taking on a self-sustaining character (Gross & Thompson, 2007). Many of the specific processes involved in homeostasis may also be involved in "metastasis." Thus, cognitive reappraisal (Ray et al., 2008), attentional redeployment (Pine et al., 2005), and response modulation (Salkovskis, 1991) can all, when performed unsuccessfully, lead to amplification or upregulation of a negative or undesired emotion. Attempts to suppress negative emotions can also lead, counterproductively, to increased sympathetic nervous system activation (Gross, 1998; Gross & John, 2003), as well as to paradoxical increases in negative mood if cognitive load is high (Wegner, Erber, & Zanakos, 1993). Self-sustaining increases might also be observed extrinsically, in interpersonal interactions. For example, research has shown that high levels of expressed emotions during parent–child interactions (such as parental criticism of a child misbehaving) can lead to increases in child cortisol, and that elevated physiological stress associates with increased misbehavior, thus creating a positive, mutually reinforcing cycle (Christiansen, Oades, Psychogiou, Hauffa, & Sonuga-Barke, 2010; Kashdan et al., 2004; Snyder, Edwards, McGraw, Kilgore, & Holton, 1994; Sonuga-Barke, 2005).

Virtually no previous work has investigated the early developmental correlates of these "metastatic" processes during infancy and early childhood. Understanding these early developmental precursors is, however, important, for two reasons. First, development is a dynamical, interactive hierarchical process, in which later-developing skills build on foundations laid down early on (Karmiloff-Smith, 1998). This means that understanding early emotional development is important from a practical perspective, for identifying typical and atypical development (Beauchaine, 2003). Second, studying early development is conceptually important and may inform our understanding of adult processes. This is particularly important given the hierarchical perspectives often taken on interactions between brainstem, limbic, and cortical brain systems in emotion regulation, such that higher systems integrate, elaborate, fine-tune, and serve an inhibitory function for hierarchically lower systems (Ruth Feldman et al., 1999). In infants,

brainstem and limbic areas are relatively well-developed at a time when frontal and temporal cortices are not (Johnson, 2015), and so studying emotion regulation during early development may offer some clues to help understand mature emotion regulation.

Our aim for this study was, therefore, to ascertain whether evidence consistent with the presence of such "metastatic" processes during infancy would be observed. Rather than by attempting to identify specific individual processes that might cause such an effect, we chose to do this by studying time series data, to examine fluctuations in emotionality over time. We reasoned that, if evidence consistent with the presence of such processes was observed in our data, then future work should investigate the (likely multiple) different individual factors that might contribute to this phenomenon.

Research into emotion regulation in infants has traditionally used one of three methods: First, parent questionnaires (Rothbart et al., 2003), which show high validity but provide only a single, time-invariant "snapshot," and so was considered unsuitable for this study. Second, video coding of facial affect, which can be performed up to a high temporal resolution (several frames per second—Cohn & Tronick, 1988; Feldman & Mayes, 1999), but which is highly labor intensive, which poses practical limitations on the amount of data that can be analyzed in this way. Third, which is the approach that we have used here, by measuring the physiological correlates of emotion, by recording the autonomic nervous system (ANS; Cacioppo, Tassinary, & Berntson, 2000). Decades of research have shown that the ANS is involved in emotional responding, such that moments of more intense autonomic arousal are associated with greater emotionality (Cacioppo et al., 2000; Cannon, 1915; Kreibig, 2010). Although the ANS can be fractionated and contains multiple, separable, subcomponent processes (e.g., Lacey, 1967), it also, in infants, shows significant covariation between ANS measures (Wass, in press; Wass, Clackson, & de Barbaro, 2016; Wass, de Barbaro, & Clackson, 2015). This justifies its treatment as a unitary construct here.

Our study used 57 typical 12-month-old infants. We presented a continuous, changing stream of static and dynamic age-appropriate pictures and animations, while we recorded infants' heart rate, electrodermal activity (EDA), and two measures of movement: head velocity and foot movement. Some of these, such as heart rate, are thought to be influenced by both the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS; McCabe, Schneiderman, & Field, 2000), whereas others, such as EDA, are thought to be primarily influenced by SNS (Shields, Macdowell, Fairchild, & Campbell, 1987). However, previous research has pointed to significant tonic and phasic covariation between these peripheral indices, consistent with the idea that the SNS and PNS operate largely (albeit not exclusively) in opposition (Kahneman, Tursky, Shapiro, & Crider, 1969; Wass et al., 2015, 2016). Other measures that were originally included were found not to show the same pattern (see Wass et al., 2015, 2016), and so have not been included here. Pupil size, for example, did not show the same pattern of covariation. This is due to practical difficulties in controlling for luminance sufficiently precisely in order to allow for accurate measurement (Wass et al., 2015).

Our intention for this study was to examine slow-varying patterns of "baseline" or "tonic" autonomic arousal, corresponding to emotional arousal, during the presentation of our mixed static and dynamic stimulus testing battery. Our primary aim was to examine whether the hypothesized patterns of "longer than expected," or "self-

sustaining" increases in autonomic arousal could be observed. Specifically, we examined whether increases in arousal were more long-lasting than decreases in arousal.

Building on this, we also wished to address three secondary questions. These were to help us understand the causation of the hypothesized "self-sustaining" increases in autonomic arousal, if they were observed. Our first secondary question was as follows: If such cycles are detected are they more likely to occur later in the session, when the infant is fatigued? Previous research suggests that emotion regulation and irritability both increase with increasing fatigue, and soothability declines (Dahl, 1996). Based on this, we hypothesized that the patterns in question might be observed more strongly later in the testing session.

Second, we examined when within the stimulus battery these long-lasting increases in autonomic arousal were more likely to be observed. Our battery was a mixed battery, constituting both static and dynamic visual stimuli interspersed. Previous research has documented a number of differences between infants' reactions to static and dynamic stimuli (e.g., Shaddy & Colombo, 2004). For example, attentional inertia—the finding that looks to the screen during TV viewing become progressively more long-lasting, the longer the look continues—is thought to be more commonly observed during the presentation of static, than of dynamic stimuli (Richards & Anderson, 2004). Based on this, and given the negative relationships previously documented between autonomic arousal and attention (Bacher & Robertson, 2001; de Barbaro, Clackson, & Wass, 2016), we hypothesized that long-lasting increases in arousal would be less prevalent during the presentation of dynamic, infant-friendly TV clips than they were during the other components of the battery.

Our third question was as follows: Can stable individual differences be identified? In other words, is an individual who is more likely to show a pattern of more long-lasting increases than decreases in arousal in one context also more likely to show them in another? Given previous research suggesting that other aspects of arousal such as heart rate variability are stable over time (Kleiger, 1991), we predicted that stable individual differences would be observed.

To address our primary question, we conducted two analysis. The first (Analysis 1) examines whether the distributions obtained for our four autonomic measures across the entire testing session were, overall, positively or negatively skewed (see schematics in Figure 1). Prior to conducting the analysis, data were *z*-scored on a per-participant basis. Thus, our analyses did not examine episodes of increased and decreased autonomic arousal relative to the group mean; rather, they examine the number and duration of episodes of increased and decreased autonomic arousal relative to the average levels recorded for that participant, at that session. We hypothesized that, following Figure 1, if self-sustaining increases were discernible, then a positive skew would be observed.

The second (Analysis 2) examines episodes of especially elevated and especially decreased arousal. Again we hypothesized that if positive self-sustaining cycles were discernible, then increases in autonomic arousal would be more long-lasting than decreases in arousal.

We addressed our secondary questions using three further analyses. First, to examine whether skewedness emerges over the testing session, we repeated the analyses in Analysis 1, but dividing our analysis into four chronological quartiles (Analysis 1b). We hypothesized that the patterns of self-sustaining increases would be observed more strongly during later quartiles. Second, to assess when within the stimulus battery these

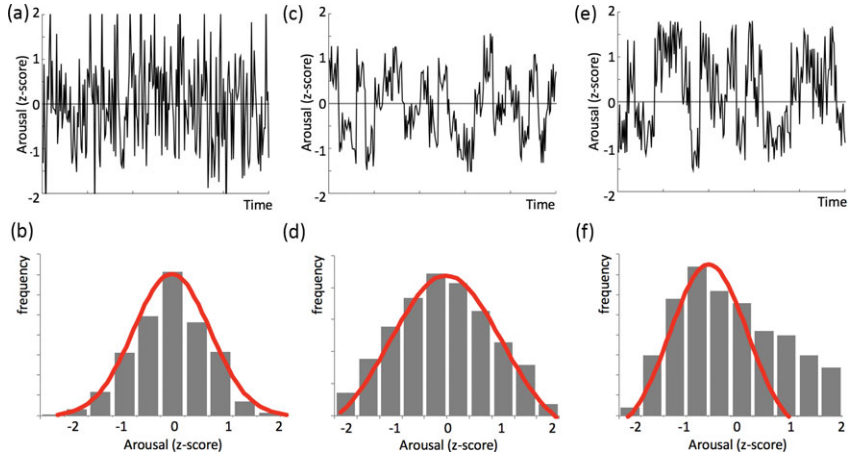

Figure 1 Schematic illustrating different ways of conceptualizing spontaneous changes in arousal. (a) Model 1—a random, stochastically varying model. (b) Histogram showing arousal levels across the entire sample in Model 1. A Gaussian distribution has been superimposed in red for illustrative purposes. (c) Model 2—an oscillatory "inertia" model, in which arousal at one moment is contingent on arousal at the previous moment, together with random, stochastic changes. (d) Histogram showing arousal levels across the entire sample in Model 2. Again, these are normally distributed. (e) Model 3 —similar to Model 2, but with the addition that episodes of high arousal are more likely to be maintained than episodes of low arousal. (f) Histogram showing arousal levels across the entire sample in Model 3. A Gaussian distribution has been superimposed in red, for illustrative purposes. The histogram shows that results are positively skewed.

types of patterns were more likely to be observed, we compared the distributions observed during the different sections of our testing battery (Analysis 3). We hypothesized that the patterns would be observed less strongly during the presentation of infant-friendly TV clips than at other times. Third, we examined evidence for consistent individual differences (Analysis 4). We hypothesized that consistent individual differences would be observed.

METHODS

Participants

Sixty-three infants were recruited for the study from a volunteer research participant pool at the Medical Research Council, Cambridge, UK. No explicit exclusion criteria were applied, but no participating families reported any major psychiatric or other clinical conditions within the immediate family. All children were from homes based in the Cambridge region of the UK, which is a wealthy university town, and participants were predominantly Caucasian.

This study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Cambridge Psychology Research Ethics Committee at the University of Cambridge.

Of the 63 infants, data collection was impossible for six due to equipment malfunction. In total, therefore, usable data were collected from 57 infants. The average age of these infants was 12 months (mean age in days: 377, *SD*: 40, range 315–455 days); the sample comprised 29 males and 28 females.

All participants viewed a 20-min stimulus battery consisting of a mixture of dynamic and static viewing materials (Figure 2). Of note, this battery constituted the first visit of a repeat-visit study, the results of which have been presented elsewhere (de Barbaro et al., 2016; Wass, Cook, & Clackson, in press). The present analyses and research questions are completely different to those reported previously. All testing materials were presented using a Tobii TX300 eye tracker recording at 120 Hz and MATLAB, Psychtoolbox, and the MATLAB SDK. The monitor subtended c. 30° of visual angle. All external sources of lighting were occluded during testing, and participants were lit with a single, diffuse light source above the tracker, at c. 300 lux.

Behaviorally, the protocol for this study was identical to that we used in several previous studies and is, to our knowledge, standard experimental procedure for infant testing elsewhere (Oakes, 2010). Infants were seated on their caregivers' laps during recording. Caregivers were requested not to talk during recording, but otherwise, no explicit instructions were given. Stimuli were presented continuously. Breaks were permitted during testing if the child became agitated (defined as sustained crying, lasting for more than 10 sec) or if the parent requested it. This occurred rarely (<10% of cases). In these cases, recording and stimulus presentation were discontinued and recommenced when the experimenter and parent agreed it was possible. See Supplementary Materials for further analyses addressing this point.

Stimuli were presented in a number of different blocks, constituting three different categories—static, mixed static/dynamic, and dynamic. The order of the blocks was identical for all participants. Blocks generally lasted between 60 and 120 sec. The static category constituted pictures of a child's face. The same picture was presented continuously for an entire block. The mixed static/dynamic category constituted animations and shapes, such as a cartoon monkey face shown moving between two different rectangles. The dynamic category constituted clips from age-appropriate television programs from the BBC such as Baby Jake and Abadas. Stimuli were presented with concurrent audio.

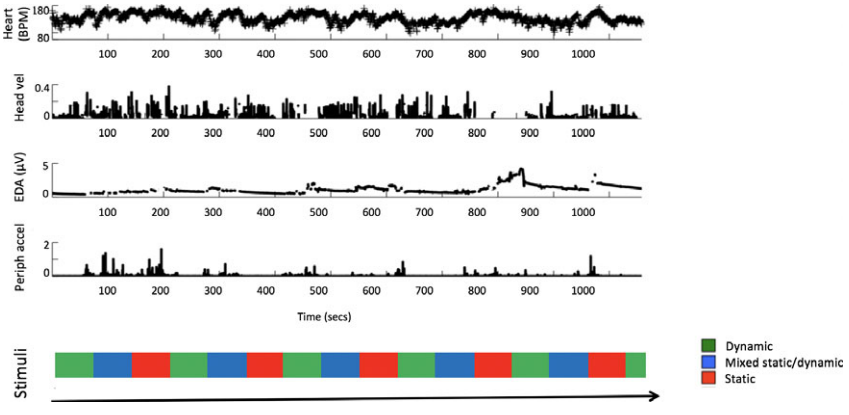

Figure 2 Schematic showing the testing battery administered to all participants.

Arousal monitoring

Arousal was measured using heart rate, electrodermal activity, and two measures of movement: head velocity and foot movement (measured using an accelerometer). In previous research, we found significant tonic and phasic covariation between these measures (Wass et al., 2015, 2016). We also recorded pupil size but found that it did not show the same phasic covariation with the other measures, which we considered likely to be due to the impossibility of controlling adequately for luminance in our setup, and so this measure has not been included here.

Electrocardiogram (ECG), electrodermal activity (EDA), and accelerometry were recorded using a BioPac™ (Santa Barbara, CA) recording at 1000 Hz. 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 were recorded. In addition, head velocity data were derived from the head position estimates that are automatically generated during heads-free eye tracking. They were recorded by a Tobii TX300 (Tobii Technology, Stockholm, Sweden) eye tracker. The process used to extract these data is described in detail elsewhere (Wass et al., 2015).

Further details on data reduction protocols are given in the Supplementary Materials, and in previous papers (Wass et al., 2015, 2016). Brief summaries are presented here.

Heart rate

R-peak identification was performed by the Acknowledge commercial software package. Artifact rejection was then performed by excluding those beats showing an interbeat interval of <330 or >750 msec, and by excluding those samples showing a rate of change of interbeat interval of greater than 80 ms between samples. This measure was set following extensive visual inspection of both of the IBI time series (see Supplementary Materials) and of the raw ECG data to ensure that genuine, vagally mediated oscillations were excluded erroneously. Our analyses suggested that, in infants, vagally mediated oscillations occur at a substantially slower rate than 80 ms between individual beats. Finally, HR data were *z*-scored on a per-participant basis and epoched into one-second epochs. In the Supplementary Materials, we report on a comparison of these cleaning techniques with traditional hand coding which shows a close comparison between the two approaches. Two of the 57 infants failed to provide any usable data. Of those that did, valid data were obtained from 88.9% (*SEM* 0.02%) of samples recorded.

Head velocity

First, data samples showing a change in position of more than 0.025 screen units between 120 Hz iterations were excluded as being above the maximum possible threshold at which head movement can take place and therefore likely to be artifactual (see Wass et al., 2015, for more details). Second, data were downsampled to 12 Hz 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 on a per-participant

basis and epoched into one-second epochs. One of the 57 infants failed to provide any usable data. Of those that did, valid data were obtained from 98.1% (*SEM* 0.01%) of samples recorded.

Electrodermal activity (EDA)

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 0.1 μV. Second, data were log transformed to remove positively skewed values. Third, data were *z*-scored on a per-participant basis and epoched into one-second epochs. In total, seven of the 57 infants failed to provide any usable data. Of those that did, valid data were obtained from 92.0% (*SEM* 0.03%) of samples recorded. Of note, the linear detrend that is typically applied to EDA prior to analyzing the results (e.g., Panju, Brian, Dupuis, & Anagnostou, 2015) was not applied in this instance, as we wished to examine slow-moving changes. In the Supplementary Materials, however, for comparison, we present a version of the results based on detrended data. The linear detrend was calculated by computing the least-squares fit and subtracting the function from the data.

Actigraph

Our approach was similar to that previously used with developmental populations (Robertson, Bacher, & Huntington, 2001). First, data were filtered to remove high-frequency noise using a Butterworth filter with a cutoff of 0.5 Hz. Second, three-dimensional movement data were summed to create a one-dimensional estimate of total movement. Third, data were *z*-scored on a per-participant basis and epoched into one-second epochs. One of the 57 infants failed to provide any usable data. Of those that did, valid data were obtained from 98.7% (*SEM* 0.01%) of samples recorded.

To assess skewness, the Kolmogorov-Smirnov (KS) test was first conducted to examine whether the distribution of results obtained differed significantly from normal. Second, the skewness was calculated using the standard formula:

$$s = \frac{E(x-\mu)^3}{\sigma^3},$$

where μ is the mean of x, σ is the standard deviation of x, and $E(t)$ represents the expected value of quantity t. The variance in skewness was also calculated using the formula:

$$\text{var(s)} = \frac{6n(n-1)}{(n-2)(n+1)(n+3)},$$

where n is the sample size. S was calculated as $S = s/\text{var(s)}$. When $S > 1.96$ or $S < -1.96$ the skewness differs significantly (at $\alpha = .05$) from zero. Statistical analyses were conducted on the data from each individual participant separately.

Hazard function analysis (Analysis 2)

Hazard function analyses were conducted to examine the survival likelihood of episodes of increased and decreased autonomic arousal. The start of an increased-, or decreased-arousal episode was treated as the first epoch in which values obtained for that participant exceeded ±1 *SD* from the mean value recorded for that participant, across the entire session. The time elapsed between the moment when arousal first exceeded this threshold and the moment when it returned below that threshold was measured. Based on the average episode durations, a Kaplan–Meier plot was then calculated to estimate the survival probability of each episode, using epoch-by-epoch data pooled across all participants. This analysis was conducted separately for each measure. In addition to the ±1 *SD* threshold used in the main analyses, analyses were repeated with different levels for this threshold and have been included in the Supplementary Materials.

RESULTS

Analysis 1—are distributions positively or negatively skewed?

Figure 3(a–d) shows histograms of all raw scores obtained. Individual epochs were pooled across all participants. For EDA, the mean raw score value is close to zero, and the distribution does not appear positively skewed. For heart rate, head velocity and actigraph two complementary aspects can be seen. First, the distributions appear

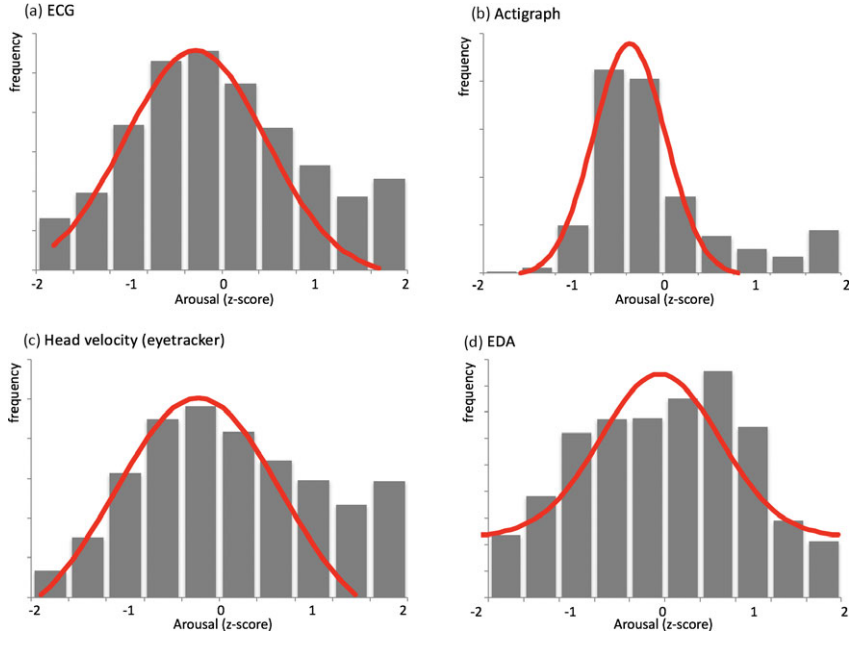

Figure 3 (a–d) Frequency distributions of all raw scores obtained. *X*-axis shows the arousal level (*z*-score). *Y*-axis shows the frequency. Gaussian distributions have been superimposed for illustrative purposes in red. (a) Heart rate; (b) Actigraph; (c) Head velocity; (d) EDA.

to show a positive skew. Second, and relatedly, the modal *z*-score values (calculated on a per-participant basis) are less than zero.

Results are presented in Table 1. In brief, results for all tests conducted for heart rate, head velocity, and actigraph showed strong positive skew. Those obtained for EDA did not differ from chance.

Analysis 1b—does skewness change over time?

To examine how skewedness changes over time, we divided the experimental session into chronological quartiles and repeated the analyses described above. The same *z*-scored data (calculated once per participant) were used as in Analysis 1.

Figure 4 and Table 2 show the results of this analysis. These results show that the results observed were largely consistent across the quartiles. At quartile 1 (the first 5 min of the testing session), significant positive skews were observed for the heart rate ($p < .001$), head velocity ($p < .001$), and actigraph ($p < .001$) data, but not for EDA.

Although the results from early quartiles are skewed, they initially appear to be less skewed than the equivalent results from later quartiles. To assess whether this difference reached significance, repeated measures ANOVAs were applied with skewness (calculated on a per-participant basis) as the DV and the factor quartile. No significant main effects of quartile were observed for heart $F(1,52) = 1.3$, $p = .27$; head $F(1,48) = .003$, $p = .96$, EDA $F(1,43) = 1.6$, =.21; or actigraph $F(1,54) = .03$, $p = .87$. This suggests that skewness did not systematically change over the course of the testing session.

Of note, Figure 4d shows a marked change between the quartiles in EDA values. Studies using EDA usually report an increase in EDA values over the course of the testing session, which is usually thought to be due to artifactual reasons (the gel on the pads warming up). Because of this, a linear detrend is normally applied to the data prior to analysis (Panju et al., 2015). In the Supplementary Materials (Figure S2, Tables S1 and S2), we show equivalent analyses applied to the detrended EDA data. Results do not materially differ to those presented here: no significant skewness was observed in the EDA data.

TABLE 1
Summary Results of Statistical Analyses from Analyses 1 and 2

	Heart rate	*Head velocity*	*EDA*	*Actigraph*
Analysis 1—Mean skewness (range)	48.9 (−47.6 to 183.6)	26.0 (−70.3 to 126.0)	−18.8 (−469.3 to 264.9)	297.8 (11.1 to 663.4)
Analysis 1—Proportion of individual participants showing positive skew $S > 1.96$ (N/total N)	45/55 (81.8%)	38/56 (69.1%)	20/50 (40.0%)	54/56 (98.2%)
Analysis 1—Do population skewness values differ from 0	$t(54) = 6.65$, $p < .001$	$t(55) = 3.68$, $p < .001$	$t(49) = -1.0$, $p = .32$ (ns)	$t(55) = 17.5$, $p < .001$
Analysis 2—Are increased-arousal episodes or decreased-arousal episodes more long-lasting?	PA > NA $t(54) = 12.5$, $p < .001$	PA > NA $t(55) = 9.4$, $p < .001$	No diff. $t(49) = 1.3$, $p = .19$	PA > NA $t(55) = 6.2$, $p < .001$

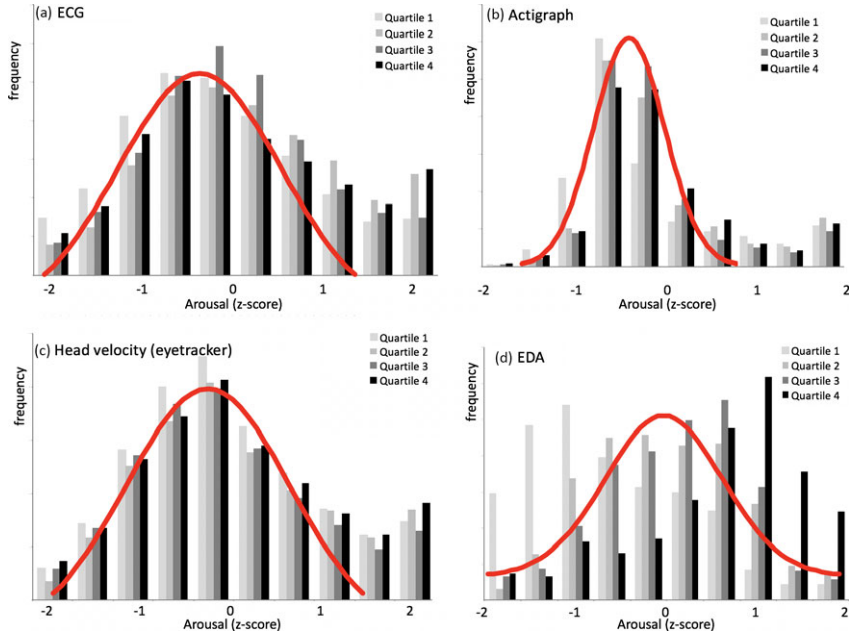

Figure 4 Frequency distributions of all raw scores obtained, with the results broken down by the quartile of the testing session in which the data were collected. *X*-axis shows the arousal level (*z*-score). *Y*-axis shows the frequency. Gaussian distributions have been superimposed for illustrative purposes in red. (a) Heart rate; (b) Actigraph; (c) Head velocity; (d) EDA. Figure calculated identically to that in Figure 2, with the addition that results were broken down by quartile. Of note, the large difference in EDA is likely to be artifactual, and due to the fact that no linear detrend was applied to the data.

TABLE 2
Summary Results of Statistical Analyses from Analysis 1, Subdivided by Quartile

	Heart rate	*Head velocity*	*EDA*	*Actigraph*
Analysis 1—Mean skewness (Q1/Q2/Q3/Q4)	12.1/6.9/ 11.4/14.2	15.1/12.7/ 16.5/14.1	8.6/9.3/10.6/−3.5	80.3/77.3/79.5/78.1
Analysis 1—Percentage of individual subjects showing positive skew $S > 1.96$ (Q1/Q2/Q3/Q4)	69/59/69/71	80/85/82/80	55/53/59/35	93/95/96/98
Analysis 1—Do population skewness values differ from 0. p value of t-test (Q1/Q2/Q3/Q4)	<.001/.01/ <.001/ <.001	<.001/<.001/ <.001/<.001	.35/.13/.07/.63	<.001/<.001/ <.001/<.001

The second aspect that we wished to examine was the fact that, as described in the Methods, above, breaks were permitted during testing for a subset ($N = 6$) of participants who became upset. To examine whether these participants showed a different pattern to the remaining participants, we repeated the analysis, subdividing our sample between those participants for whom breaks were permitted ($N = 6$) and those who

did not require breaks ($N = 50$; see Figure S2 and Table S1). In sum, these results suggested that both groups showed skewed distributions for heart rate, head velocity, and actigraph but not EDA. The group for whom breaks were permitted appeared to show more skewed distributions, but this difference was not significant for all but the EDA measure.

Analysis 2—are episodes of increases in autonomic arousal more long-lasting than episodes of decreases in arousal?

Our main hypothesis is that the survival likelihood of episodes of increased autonomic arousal will be greater than episodes of decreased arousal. To assess this, the average duration of the episodes of especially increased, and decreased, autonomic arousal was calculated on a participant by participant basis. Figure 5(a–d) shows the results from these analyses. The two results were compared using paired samples *t*-test analysis to examine whether the mean duration of episodes of increases in arousal exceeded the mean duration of episodes of decreases in arousal. Significantly longer increases than decreases in arousal were observed for heart rate $t(54) = 9.6$, $p < .001$; head velocity $t(55) = 9.0$, $p < .001$; and actigraph $t(55) = 7.6$, $p < .001$ but not EDA $t(50) = 1.6$, $p = .11$.

One possibility we wished to address was that the arbitrary threshold setting using to define increases and decreases in arousal may have influenced our results. To assess this possibility, we repeated our analyses using two different thresholds: ± 1.5 *SD* and

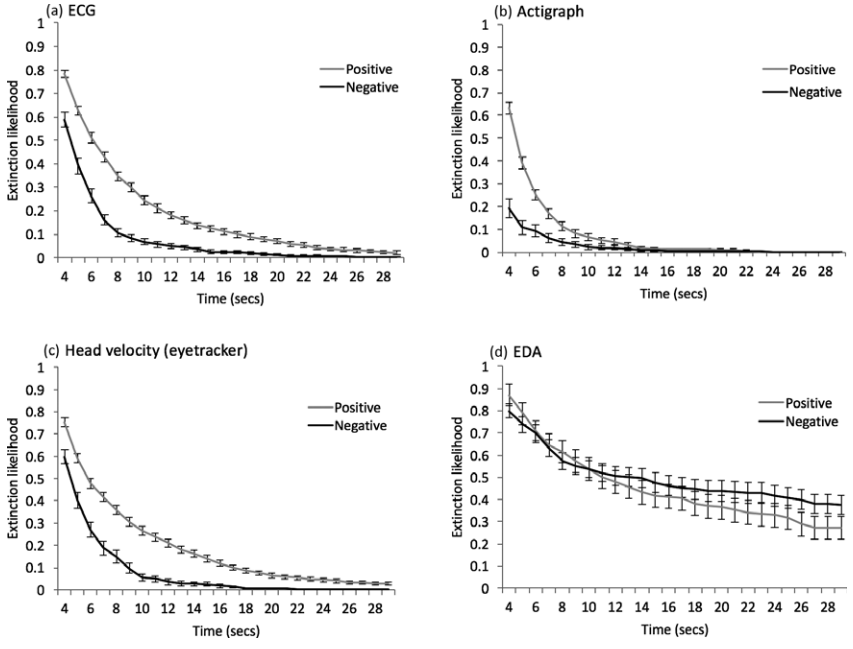

Figure 5 (a–d) Hazard plots showing the extinction likelihood of episodes of positive (grey) and negative (black) arousal. *X*-axis shows the epoch length (in sec). *Y*-axis shows the extinction probability. Error bars show standard error. (a) Heart rate; (b) Actigraph; (c) Head velocity; (d) EDA.

±0.5 *SD* (see Figures S3 and S4). In sum, the results suggested that the same set of results is observed when different thresholds are used: Increased-arousal episodes were significantly longer than decreased-arousal episodes for heart rate, head velocity, and actigraph, but not EDA. The sole exception for this was actigraph at a threshold of 0.5 *SD*, which showed a different pattern.

In addition to our main analysis, a number other differences between our measures can be seen. EDA has a slow decay rate, and actigraph a fast decay rate, relative to the other two measures. This corresponds to previous findings from autocorrelation analyses, suggesting that, relative to other measures, EDA is a slower-changing measure, and actigraph a faster changing measure (Wass et al., 2016).

Analysis 3—is the same pattern observed across all types of viewing materials?

The analyses hitherto have examined arousal across the entire testing session. Analysis 3 examines whether the same pattern, which increases in arousal were more long-lasting that decreases in arousal, was observed consistently across the mixed testing battery presented. To do this, we compared three sections of our experimental test battery (see Figure 2): (1) static, (2) mixed static and dynamic, and (3) dynamic (Figure 6).

To maintain consistency with our previous analyses, we wished to base this analysis on data that were *z*-scored once, on a per-participant basis. We were concerned, however, that when examining reduced sections of our data; there was a possibility that between-section differences in mean arousal level might influence the skewness of the

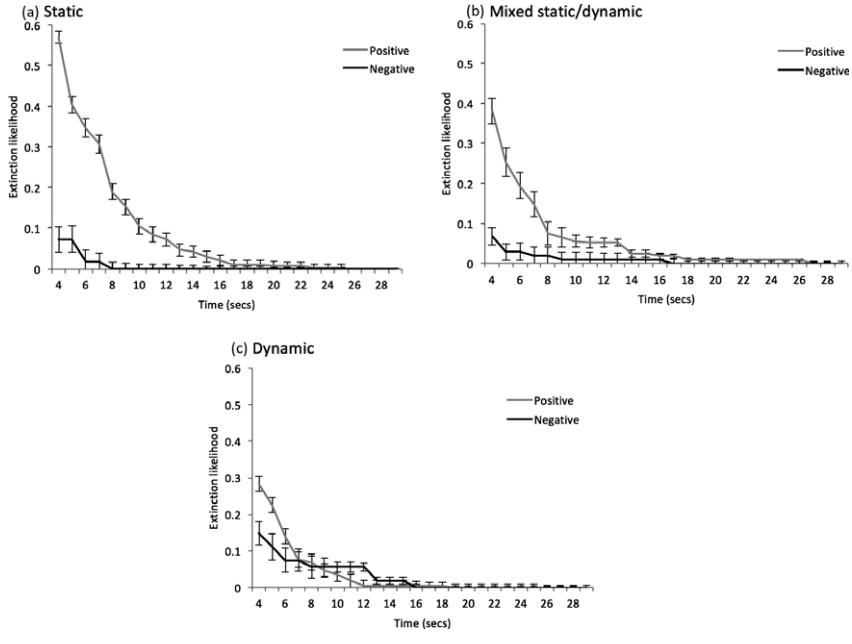

Figure 6 Hazard plots showing the extinction likelihood of episodes of positive (grey) and negative (black) arousal. Calculations were performed based on a composite arousal measure, derived by averaging the *z*-scores of the four arousal measures used hitherto. (a) Static; (b) Mixed static/dynamic; (c) Dynamic.

distributions obtained. Therefore, the approach of analysis 2 (the hazard function analysis) was used. Because we were subdividing our data into three, and to avoid this leading to a problem of multiple comparisons, just a composite arousal measure was used, rather than the four individual arousal measures. As in previous work (de Barbaro et al., 2016), this was obtained by averaging together the *z*-scored values obtained from the four measures independently.

Paired samples *t*-tests were calculated to estimate whether, for each subsection of the task independently, episodes of increases in autonomic arousal were more long-lasting than episodes of decreases in arousal. This was true for all three categories: static $t(54) = 18.3$, $p < .001$; mixed static/dynamic $t(54) = 3.9$, $p < .001$; and dynamic $t(54) = 2.9$, $p = .006$. This suggests that, for all three sections considered independently, episodes of increases in arousal were more long-lasting than episodes of decreases in arousal. Paired-sample *t*-tests also suggested that the episodes of increases in arousal were more long-lasting during the static than during the mixed static/dynamic stimuli $t(54) = 4.7$, $p < .001$ and during the static than the dynamic stimuli $t(54) = 6.7$, $p < .001$. However, the difference between the mixed static/dynamic and the dynamic stimuli was not significant $t(54) = .96$, $p = .34$.

Analysis 4—individual differences

The final question was—are there stable individual differences in the frequency of sustained increases in arousal? To address this, two separate analyses were performed. First, we repeated Analysis 1. A split-half analysis was conducted, and the distributions obtained for each individual for each half, and for each measure, were compared. Our analyses examined whether the skewness obtained for that individual, for that measure, during the first half of the testing session was significantly associated with the skewness obtained for that individual, for that measure, during the second half. All four relationships observed were positive. The heart rate measure was found to show significantly split-half reliability $\rho(54) = .40$, $p = .003$. Head velocity did not show significant split-half reliability $\rho(54) = .15$, $p = .39$. EDA also did not show significant split-half reliability $\rho(54) = .23$, $p = .13$. Actigraph did show significant split-half reliability $\rho(54) = .48$, $p < .001$.

One limitation of this approach is that only a single data point, representing the population skewness, is generated per half, per individual. We wished, therefore, to repeat this analysis, but using a more conventional approach in which a number of different samples were obtained from each individual, and these numbers were averaged to create a single, more accurate, index, before test–retest reliability was calculated. To do this, we adopted a different approach to identify episodes of sustained increases in autonomic arousal. This approach is analogous to the raster plot approach used to identify spike firing in electrophysiology (Lewicki, 1998). An episode of sustained increase in autonomic arousal was identified when consecutive increases were observed between three consecutive 2-sec epochs (see Figure 7a). These consecutive increases all had to be above a threshold of .03 standard deviations. The criterion was applied consistently across the entire testing session. Only episodes contained entirely within each segment were included. Comparative analyses were also conducted, which for the sake of brevity are not reported here, which suggested that small changes to the levels of these thresholds did not have a large influence on results. Using this criterion, we identified an average of 1.6 incidents of sustained

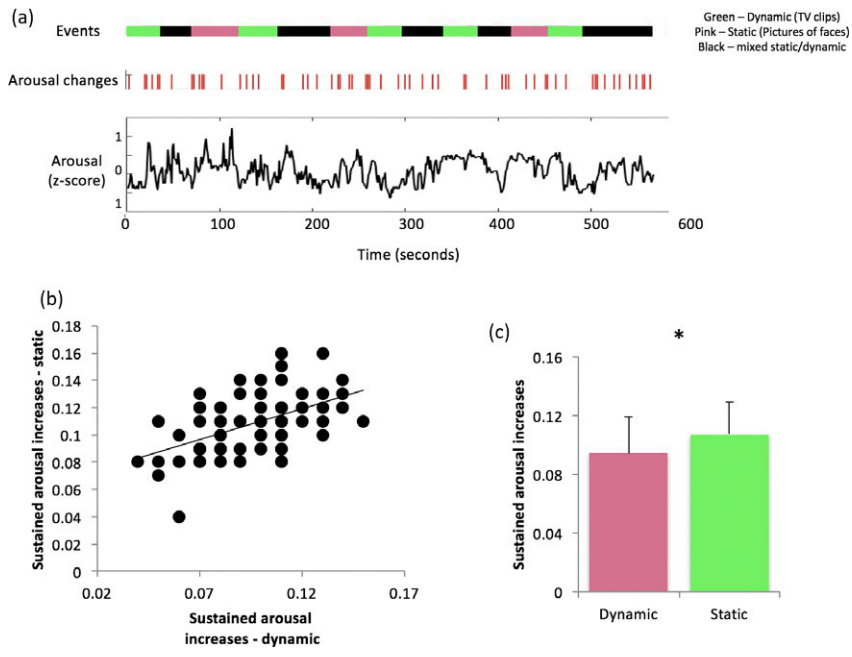

Figure 7 (a) Illustrative data sample showing from a single infant viewing our test battery. Illustrating the raster plot approach used to identify incidents of sustained increases in arousal. The events presented are shown on the top line. In the second line, arousal increases are shown, as defined in the text below. In the third line, continuous arousal is shown. (b) Scatterplot showing the relationship between the number of increases in arousal observed across the dynamic and static conditions. (c) Bar chart showing the number of increases in arousal observed in the dynamic and static conditions.

increases in arousal per minute of testing time, across the entire battery. As this analysis was conducted on data that had been *z*-scored on a per-participant basis beforehand, this analysis examines the frequency of episodes of sustained increases in autonomic arousal whilst controlling for individual differences in tonic arousal (mean) and in arousal variability (*SD*).

The number of increases in autonomic arousal and the skewness of the distribution were only weakly inter-related ($r(56) = .14$, $p = .29$) (such that more increases in arousal were associated with a more positively skewed arousal distribution). This suggests that the association between this criterion and the previous one is not a strong one. However, using this method, we were able to replicate the previous finding that sustained increases in arousal were more likely to occur during the static condition than the dynamic condition $t(56) = 5.2$, $p < .001$ (Figure 7c).

We also examined the association between the number of sustained increases in arousal observed in the static and dynamic conditions, using a Pearson's bivariate correlation. A significant relationship was identified between the number of sustained increases in arousal observed in the static and dynamic conditions $r(54) = .51$, $p < .001$. This suggests that stable individual differences were observed.

DISCUSSION

Emotion regulation is the adjustment of emotional state or expression to meet goals or to maintain homeostasis or allostasis (Gross & Thompson, 2007). These processes are generally considered in contexts in which negative or undesired emotions are downregulated, or deamplified. In addition to these downregulatory processes, though, there also exist opposite processes, in which negative or undesired emotions can be exacerbated over time through a process of upregulation, or amplification. Previous research has identified both downregulatory and upregulatory processes in adults (Gross & Thompson, 2007). Early precursors of downregulatory processes have been identified previously in infants but, to our knowledge, no previous work has examined these upregulatory, or "antihomeostatic", processes in infants.

We recorded four measures of autonomic arousal (heart rate, head velocity, EDA, and actigraph) in 57 typically developing 12-month-old infants while they viewed a mixed experimental testing battery (Figure 2). We examined how ANS activity fluctuated dynamically within the testing session. Prior to performing our calculations, data were *z*-scored on a per-participant basis. Therefore, our analyses examined autonomic arousal levels relative to the average arousal levels recorded for that participant, for that session. We found that, for three of our measures—heart rate, head velocity and actigraph—results were significantly positively skewed (Figure 3 and Table 1). For EDA, no skew was observed (Figure 3 and Table 1). Patterns were observed consistently across different individual participants (Table 1), and so were not simply an effect of pooling across individuals. We also used a survival probability analysis to estimate whether increased-arousal episodes were more long-lasting than decreased-arousal episodes (Analysis 2). Consistent with analysis 1, we found that, for heart rate, head velocity, and actigraph, but not EDA, increased-arousal episodes were more long-lasting than decreased-arousal episodes (Figure 5).

Previous research has suggested that stimulus batteries of this kind tend to evoke a number of predictable, reflexive autonomic responses: For example, systematic decreases in heart rate tend to be observed in infants relative to the onsets of new stimulus events, such as pictures and TV clips (Richards, 1997). In addition, cyclical oscillatory activity will also have been present in our data—due, for example, to periodic changes in heart rate with the respiration cycle (known as Respiratory Sinus Arrhythmia; e.g., Porges, 1974), and cyclical fluctuations at c. 0.1 Hz (known as Mayer waves; Julien, 2006). However, it is unclear how the pattern of results we observed across the first two analyses could be attributed to the summative effects of any of these factors.

One alternative possible cause for a skew in our data might be ceiling and floor effects. It is possible that, for example, mean heart rate was closer to the minimum than to the maximum possible heart rate value, thus causing the positive skew. This would be despite the fact that, as described in the Methods, a ± 2 *SD* cutoff was used, with values more than two standard deviations from the mean truncated. Two arguments count against this possibility. First, the same pattern was observed consistently across heart rate and the two different movement measures, despite the fact that ceiling and floor levels will have differed between the three measures. Second, some (albeit limited) evidence for consistent individual differences was observed, as discussed further below.

An alternative possibility is that our results could have been due to artifact. As described in the methods, sections of the data where the participant was moving were not systematically excluded. This approach is consistent with that used in other studies

with infants (Beauchaine, 2001; Colombo, Shaddy, Richman, Maikranz, & Blaga, 2004), but at odds with those sometimes used in psychophysiological recordings with adults (e.g., Goudriaan, Oosterlaan, de Beurs, & van den Brink, 2006; see also Porges et al., 2007; Bush, Alkon, Obradović, Stamperdahl, & Boyce, 2011). The presence of movement may have contributed to increased artifact in our heart rate data, leading to the false impression that heart rate was increasing (see Supplementary Materials). However, the fact that the same pattern was also observed on other measures including actigraph, which is relatively artifact free, suggests that this is unlikely to have been the cause of the results observed.

Electrodermal activity did not show the same pattern as the other three measures. In previous research, we found that EDA shows strong phasic covariation with other measures when general arousal is high, but not when general arousal is low—in contrast to other measures, which showed covariation both at high, and low, general levels of arousal (Wass et al., 2015). This may explain the present results.

Sustained increases in autonomic arousal appeared to be observed more commonly during the static than the dynamic sections of the stimulus battery (Frankenhaeuser, Nordheden, Myrsten, & Post, 1971). In future, it will be interesting to investigate this question in more detail in order to ascertain whether there is an optimum level between over- and understimulation (cf Gardner & Karmel, 1984; Kidd, Piantadosi, & Aslin, 2012) and how this differs between individuals. This may relate to previous research that has investigated early individual differences in infants' preferences for high and low external stimulation from both a state (Gardner & Karmel, 1984; Geva, Gardner, & Karmel, 1999) and trait (Cohen et al., 2013; Gardner, Karmel, Magnano, Norton, & Brown, 1990; Gardner et al., 2006; Geva, Schreiber, Segal-Caspi, & Markys-Shiffman, 2013) perspective.

We also observed some, albeit limited, evidence for consistent individual differences (Analysis 4, Figure 7), despite the fact that per-participant *z*-scoring had been conducted to control for individual differences in tonic arousal (mean) and arousal variability (standard deviation). This might indicate that in some children, an episode of temporarily elevated arousal is more likely to lead onwards to a more sustained period of elevated arousal, whereas in other children, it is less likely (cf Potegal, Carlson, Margulies, Gutkovitch, & Wall, 2009). Of note, though, the fact that, for the second analysis, the two testing sections compared were presented interleaved with one another means that the stable individual differences that we identified could be due to temporary, short-lived state differences rather than stable trait differences (Cattell, 1966).

Our motivation for this study was to examine whether evidence of dynamical upregulation, or self-sustaining increases in autonomic arousal, could be observed during infancy. The results were consistent with the existence of such patterns. However, it should be noted that this is not the only pattern that would account for these findings. An alternative hypothesis is that increases in autonomic arousal may be slower-decaying or have greater hysteresis. Similar dynamical asymmetries between increases and decreases in arousal have been noted previously. Thus: "a transition from a motorically activated state (waking) to a motorically inactivated state (sleep) is always slower than the reverse: thus the ascending limb of the rest–activity cycle is steeper than the descending limb" (Hobson & Steriade, 1986).

The question of whether the phenomena that we have documented here are best considered as active processes in which small increases are positively amplified, or passive processes in which increases are slower to decay than decreases, is a question for

future research. It is informative to note, though, that, although it is little discussed, the same question applies equally to positive (successful) emotion regulation. When an un-wished-for emotion appears and then disappears, is this due to *active* emotional regulatory processes (i.e., an active process of reducing emotion)? Or is it due to "regression to the mean," in which the effects of a small initial perturbation gradually dissipate and reduce over time?

As we discussed in the Introduction, it seems likely that multiple different individual processes may have contributed to the pattern of results we observed; the question of whether these processes are best understood in terms of active amplification or not may differ depending on which of these different contributory processes we examine. For example, when considering interpersonal interactions, the positive amplification of increases in arousal (both positive and negative in valence (Feldman, 2003)) seems intuitive. For example, we have noted in our laboratory that, sometimes, when positioned on their parents' lap during testing, 12-month-old infants try to get down. When the parent attempts to restrain them, they often become oppositional and start thrashing and crying. Seen in this light, it seems easy to contemplate how a small initial increase in arousal (leading to the urge to get down) might become progressively amplified over time. It would be interesting to observe these processes in more detail, to examine whether they are observed in some dyads more than others, whether they are influenced by the arousal level of the parent, and at what age they start. Relatedly, some research has examined "stress contagion" (Waters, West, & Mendes, 2014), and similar patterns of interpersonal emotional exacerbation in older children (Christiansen et al., 2010).

But what of the *intra*-individual level? Adult research offers a model of how increases in arousal can become endogenously amplified over time, through processes such as rumination and attention biases (Pine et al., 2005; Gross & Feldman Barrett, 2011). Can similar patterns be observed in infants? Anecdotally, many a parent has observed that a child can, by crying, "work themselves up into a stew," but these types of behavior have, to our knowledge, received little formal investigation. It is also unclear whether such processes also exist at a lower threshold: Whether other types of negative emotion that are less intense than crying also show the same pattern.

Differentiating between the different processes that might cause the patterns that we observed should be a target for future research. To this end, research should consider the importance of wider context in which infants' regulatory capacities were being assessed (Fogel, 1993). In our experimental setup, infants were placed on their parent's lap in front of a computer monitor. The intention was that they should view a set amount of viewing materials, following a pre-prescribed program. Future work should investigate whether similar patterns would be observed in situations where the infant is unconstrained and allowed to play freely. It should also include video coding to address one limitation of our work here, which is that we were unable to distinguish changes in arousal that were positive, and negative, in valence. In future, it will also be interesting to investigate possible relationships to other well-researched aspects of emotionality such as stress reactivity, negative affect, and soothability (Aksan & Kochanska, 2004; Morasch & Bell, 2012; Rothbart et al., 2003; Shaw, Stringaris, Nigg, & Leibenluft, 2014; Sheese et al., 2008). Of note, other research has suggested that excessive (maladaptive) reflexive changes in heart rate variability in response to challenge may be a cross-diagnostic marker of psychopathology (Beauchaine & Thayer, 2015). It may be that similarities exist between these excessive (maladaptive) reflexive changes and the patterns we have documented here. Finally, future work should also

use other, more advanced quantitative methods such as nonlinear time series analyses to examine whether the patterns are best understood as actively amplificatory processes (in which a small initial perturbation becomes amplified over time), or more passive processes (in which increases in arousal are slower to decay than decreases).

The interpretability of our present findings must remain limited, for the reasons given above. Our use of emotional arousal as a unidimensional construct for multivalent autonomic activity and our equivocation of autonomic arousal and emotional arousal are both simplifications (albeit necessary ones; see e.g., Kreibig, 2010). Nevertheless, we contend that, given the fact that the majority of published research into emotion regulation in infancy has examined positive processes, these opposite processes should be more investigated.

Specifically, we have argued that we must understand in more detail the specific factors that led to the patterns we observed here: Whether they are limited to the situation (a confined, structured testing environment) or whether they would also be observed during naturalistic play; whether they are best understood as intrinsic (self-regulatory) or extrinsic (due to oppositionality with the parents); and whether they should be understood as "active" processes (in which small initial increases in arousal become amplified over time), or "passive" processes (in which increases in arousal are slower to decay than decreases).

ACKNOWLEDGMENTS

This research was funded by intramural Medical Research Council funding at the MRC Cognition and Brain Sciences Unit in Cambridge and by ESRC Grant numbers ES/N006461/1 and ES/N017560/1. Thanks to Edmund Sonuga-Barke, Kaya de Barbaro, and Emily Jones for useful discussions.

REFERENCES

Aksan, N., & Kochanska, G. (2004). Links between systems of inhibition from infancy to preschool years. *Child Development*, *75*(5), 1477–1490.

Bacher, L. F., & Robertson, S. S. (2001). Stability of coupled fluctuations in movement and visual attention in infants. *Developmental Psychobiology*, *39*(2), 99–106.

de Barbaro, K., Clackson, K., & Wass, S. V. (2016). Infant attention is dynamically modulated with changing arousal levels. *Child Development*, *88*(2), 629–639.

Beauchaine, T. P. (2001). Vagal tone, development, and Gray's motivational theory: Toward an integrated model of autonomic nervous system functioning in psychopathology. *Development and Psychopathology*, *13*(2), 183–214.

Beauchaine, T. P. (2003). Taxometrics and developmental psychopathology. *Development and Psychopathology*, *15*(3), 501–527.

Beauchaine, T. P., & Thayer, J. (2015). Heart rate variability as a transdiagnostic biomarker of psychopathology. *International Journal of Psychophysiology*, *98*(2 Pt 2), 338–350.

Bush, N. R., Alkon, A., Obradović, J., Stamperdahl, J., & Boyce, W. T. (2011). Differentiating challenge reactivity from psychomotor activity in studies of children's psychophysiology: Considerations for theory and measurement. *Journal of Experimental Child Psychology*, *110*(1), 62–79.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2000). *Handbook of psychophysiology* (2nd ed.). Cambridge, UK: Cambridge University Press.

Calkins, S. D., Dedmon, S. E., Gill, K. L., Lomax, L. E., & Johnson, L. M. (2002). Frustration in infancy: Implications for emotion regulation, physiological processes, and temperament. *Infancy*, *3*(2), 175–197.

Cannon, W. B. (1929). The sympathetic division of the autonomic system in relation to homeostasis. *Archives of Neurology & Psychiatry*, *22*(2), 282–294.

Cattell, R. B. (1966). Patterns of change: Measurement in relation to state dimension, trait change, lability, and process concepts. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 288–329). Chicago: Rand McNally.

Christiansen, H., Oades, R. D., Psychogiou, L., Hauffa, B. P., & Sonuga-Barke, E. J. (2010). Does the cortisol response to stress mediate the link between expressed emotion and oppositional behavior in Attention-Deficit/Hyperactivity-Disorder (ADHD)? *Behavioral and Brain Functions*, *6*(1), 45.

Cohen, I. L., Gardner, J. M., Karmel, Z., Phan, H. T. T., Kittler, P., Gomez, T. R., … Barone, A. (2013). Neonatal brainstem function and four-month arousal modulated attention are jointly associated with autism. *Autism Research*, *6*(1), 11–22.

Cohn, J. F., & Tronick, E. Z. (1988). Mother-infant face-to-face interaction: Influence is bidirectional and unrelated to periodic cycles in either partner's behavior. *Developmental Psychology*, *24*(3), 386–392.

Colombo, J., Shaddy, D. J., Richman, W. A., Maikranz, J. M., & Blaga, O. M. (2004). The developmental course of habituation in infancy and preschool outcome. *Infancy*, *5*(1), 1–38.

Dahl, R. E. (1996). The regulation of sleep and arousal: Development and psychopathology. *Development and Psychopathology*, *8*(01), 3–27.

Davis, R. C. (1958). The domain of homeostasis. *Psychological Review*, *65*(1), 8.

Eisenberg, N., Spinrad, T. L., & Eggum, N. D. (2010). Emotion-related self-regulation and its relation to children's maladjustment. *Annual Review of Clinical Psychology*, *6*, 495–525.

Feldman, R. (2003). Infant-mother and infant-father synchrony: The coregulation of positive arousal. *Infant Mental Health Journal*, *24*(1), 1–23.

Feldman, R., Greenbaum, C. W., & Yirmiya, N. (1999). Mother–infant affect synchrony as an antecedent of the emergence of self-control. *Developmental Psychology*, *35*(1), 223.

Feldman, R., & Mayes, L. C. (1999). The cyclic organization of attention during habituation is related to infants' information processing. *Infant Behavior and Development*, *22*(1), 37–49.

Fogel, A. (1993). *Developing through relationships: Origins of communication, self and culture*, London, UK: Harvester Wheatsheaf.

Fox, N. A., & Calkins, S. D. (2003). The development of self-control of emotion: Intrinsic and extrinsic influences. *Motivation and Emotion*, *27*(1), 7–26.

Frankenhaeuser, M., Nordheden, B., Myrsten, A.-L., & Post, B. (1971). Psychophysiological reactions to understimulation and overstimulation. *Acta Psychologica*, *35*(4), 298–308.

Gardner, J. M., & Karmel, B. Z. (1984). Arousal effects on visual preferences in neonates. *Developmental Psychology*, *20*(3), 374.

Gardner, J. M., Karmel, B. Z., Freedland, R. L., Lennon, E. M., Flory, M. J., Miroshnichenko, I., … Harin, A. (2006). Arousal, attention, and neurobehavioral assessment in the neonatal period: Implications for intervention and policy. *Journal of Policy and Practice in Intellectual Disabilities*, *3*(1), 22–32.

Gardner, J. M., Karmel, B. Z., Magnano, C. L., Norton, K. I., & Brown, E. G. (1990). Neurobehavioral indicators of early brain insult in high-risk neonates. *Developmental Psychology*, *26*(4), 563–575.

Geva, R., Gardner, J. M., & Karmel, B. Z. (1999). Feeding-based arousal effects on visual recognition memory in early infancy. *Developmental Psychology*, *35*(3), 640.

Geva, R., Schreiber, J., Segal-Caspi, L., & Markys-Shiffman, M. (2013). Neonatal brainstem dysfunction after preterm birth predicts behavioral inhibition. *Journal of Child Psychology and Psychiatry*, *34*, 556–586.

Goudriaan, A. E., Oosterlaan, J., de Beurs, E., & van den Brink, W. (2006). Neurocognitive functions in pathological gambling: A comparison with alcohol dependence. Tourette syndrome and normal controls. *Addiction*, *101*(4), 534–547.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, *2*(3), 271.

Gross, J. J., & Feldman Barrett, L. (2011). Emotion generation and emotion regulation: One or two depends on your point of view. *Emotion Review*, *3*(1), 8–16.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, *85*(2), 348.

Gross, J. J., & Thompson, R. A. (2006). Emotion regulation: Conceptual foundations. In J. J. Gross (Ed.), *Handbook of Emotion Regulation*. New York: Guildford Press.

Ham, J., & Tronick, E. (2008). A procedure for the measurement of infant skin conductance and its initial validation using clap induced startle. *Developmental Psychobiology*, *50*(6), 626–631.

Hernes, K. G., Morkrid, L., Fremming, A., Odegarden, S., Martinsen, O. G., & Storm, H. (2002). Skin conductance changes during the first year of life in full-term infants. *Pediatric Research*, *52*(6), 837–843.

Hobson, J. A., & Steriade, M. (1986). The neuronal basis of behavioral state control. In F. E. Bloom (Ed.), *Handbook of physiology – the nervous system*. Vol. *IV*. Bethesda, Maryland: American Physiological Society.

Johnson, M. H. (2015). *Developmental cognitive neuroscience* (4th ed.). Oxford, UK: Wiley-Blackwell.

Julien, C. (2006). The enigma of Mayer waves: Facts and models. *Cardiovascular Research*, *70*(1), 12–21.

Kahneman, D., Tursky, B., Shapiro, D., & Crider, A. (1969). Pupillary heart rate and skin resistance changes during a mental task. *Journal of Experimental Psychology*, *79*(1 Part 1), 164–167.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*, *2*(10), 389–398.

Kashdan, T. B., Jacob, R. G., Pelham, W. E., Lang, A. R., Hoza, B., Blumenthal, J. D., & Gnagy, E. M. (2004). Depression and anxiety in parents of children with ADHD and varying levels of oppositional defiant behaviors: Modeling relationships with family functioning. *Journal of Clinical Child and Adolescent Psychology*, *33*(1), 169–181.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, *7*(5), e36399 36391–36398.

Kleiger (1991). *Stability over time of variables measuring heart rate variability in normal subjects – ScienceDirect*. Retrieved from http://www.sciencedirect.com/science/article/pii/000291499190355O

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological psychology*, *84*(3), 394–421.

Lacey, J. I. (1967). Somatic response patterning and stress: Some revisions of activation theory. In M. H. Appley, & R. Trumbull (Eds.), *Psychological stress* (pp. 14–37). New York, NY: Appleton-Century-Crofts.

Lewicki, M. S. (1998). A review of methods for spike sorting: The detection and classification of neural action potentials. *Network: Computation in Neural Systems*, *9*(4), R53–R78.

McCabe, P. M., Schneiderman, N., & Field, T. M. (2000). *Stress, coping, and cardiovascular disease*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Morasch, K. C., & Bell, M. A. (2012). The role of inhibitory control in behavioral and physiological expressions of toddler executive function. *Journal of Experimental Child Psychology*, *108*(3), 593–606.

Nigg, J. T. (2017). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *58*(4), 361–383.

Oakes, L. M. (2010). Advances in eye tracking in infancy research. *Infancy*, *17*(1), 1–8.

Panju, S., Brian, J., Dupuis, A., & Anagnostou, E. (2015). Atypical sympathetic arousal in children with autism spectrum disorder and its association with anxiety symptomatology. *Molecular Autism*, *6*, 64.

Pine, D. S., Mogg, K., Bradley, B. P., Montgomery, L., Monk, C. S., McClure, E., … Kaufman, J. (2005). Attention bias to threat in maltreated children: Implications for vulnerability to stress-related psychopathology. *American Journal of Psychiatry*, *162*(2), 291–296.

Porges, S. W. (1974). Heart rate indices of newborn attentional responsivity. *Merrill-Palmer Quarterly-Journal of Developmental Psychology*, *20*(4), 231–254.

Porges, S. W., & Furman, S. A. (2011). The early development of the autonomic nervous system provides a neural platform for social behavior: A polyvagal perspective. *Infant Child Dev*, *20*(1), 106–118.

Porges, S. W., Heilman, K. J., Bazhenova, O. V., Bal, E., Doussard-Rooseveit, J. A., & Koledin, M. (2007). Does motor activity during psychophysiological paradigms confound the quantification and interpretation of heart rate and heart rate variability measures in young children? *Developmental Psychobiology*, *49*(5), 485–494.

Posner, M. I., & Rothbart, M. K. (1998). Attention, self-regulation and consciousness. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *353*(1377), 1915–1927.

Potegal, M., Carlson, G., Margulies, D., Gutkovitch, Z., & Wall, M. (2009). Rages or temper tantrums? The behavioral organization, temporal characteristics, and clinical significance of angry-agitated outbursts in child psychiatry inpatients. *Child Psychiatry and Human Development*, *40*(4), 621–636.

Ray, R. D., Wilhelm, F. H., & Gross, J. J. (2008). All in the mind's eye? Anger rumination and reappraisal. *Journal of personality and social psychology*, *94*(1), 133.

Richards, J. E. (1997). Peripheral stimulus localization by infants: Attention, age, and individual differences in heart rate variability. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 667–680.

Richards, J. E., & Anderson, D. R. (2004). Attentional inertia in children's extended looking at television. *Advances in Child Development and Behavior*, *32*, 163–212.

Robertson, S. S., Bacher, L. F., & Huntington, N. L. (2001). The integration of body movement and attention in young infants. *Psychological Science*, *12*(6), 523–526.
Rothbart, M. K., Ellis, L. K., Rueda, M. R., & Posner, M. I. (2003). Developing mechanisms of temperamental effortful control. *Journal of Personality*, *71*(6), 1113–1143.
Salkovskis, P. M. (1991). The importance of behaviour in the maintenance of anxiety and panic: A cognitive account. *Behavioural Psychotherapy*, *19*(1), 6–19.
Shaddy, D. J., & Colombo, J. (2004). Developmental changes in infant attention to dynamic and static stimuli. *Infancy*, *5*(3), 355–365.
Shaw, P., Stringaris, A., Nigg, J., & Leibenluft, E. (2014). Emotion dysregulation in attention deficit hyperactivity disorder. *American Journal of Psychiatry*, *171*(3), 276–293.
Sheese, B. E., Rothbart, M. K., Posner, M. I., White, L. K., & Fraundorf, S. H. (2008). Executive attention and self-regulation in infancy. *Infant Behavior & Development*, *31*(3), 501–510.
Shields, S. A., Macdowell, K. A., Fairchild, S. B., & Campbell, M. L. (1987). Is mediation of sweating cholinergic, adrenergic or both – A comment on the literature. *Psychophysiology*, *24*(3), 312–319.
Snyder, J., Edwards, P., McGraw, K., Kilgore, K., & Holton, A. (1994). Escalation and reinforcement in mother-child conflict: Social processes associated with the development of physical aggression. *Development and Psychopathology*, *6*(2), 305–321.
Sonuga-Barke, E. J. S. (2005). Causal models of attention-deficit/hyperactivity disorder: From common simple deficits to multiple developmental pathways. *Biological Psychiatry*, *57*(11), 1231–1238.
Stifter, C. A., & Fox, N. A. (1990). Infant reactivity: Physiological correlates of newborn and 5-month temperament. *Developmental Psychology*, *26*(4), 582.
Tucker, D. M., Derryberry, D., & Luu, P. (2000). Anatomy and physiology of human emotion: Vertical integration of brain stem, limbic and cortical systems. In J. C. Borod (Ed.), *The neuropsychology of emotion: Series in affective science* (pp. 56–79). New York, NY: Oxford University Press.
Wass, S. V. (in press). How orchids concentrate? The relationship between physiological stress reactivity and cognitive performance during infancy and early childhood. *Neuroscience and Biobehavioral Reviews*.
Wass, S. V., Clackson, K., & de Barbaro, K. (2016). Temporal dynamics of arousal and attention in infants. *Developmental Psychobiology*, *64*, 1–17.
Wass, S. V., Cook, C., & Clackson, K. (2017). Changes in behaviour and salivary cortisol following targeted cognitive training in typical 12-month-old infants. *Developmental Psychology*, *53*(5), 815–825.
Waters, S. F., West, T. V., & Mendes, W. B. (2014). Stress contagion: Physiological covariation between mothers and infants. *Psychological Science*, *25*(4), 934–942.
Wegner, D. M., Erber, R., & Zanakos, S. (1993). Ironic processes in the mental control of mood and mood-related thought. *Journal of Personality and Social Psychology*, *65*(6), 1093.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1. Comparison of artefact rejection procedures for heart rate data.

Figure S2. Repeat of Analyses 1 and 1b with detrended EDA data.

Figure S3. Frequency distributions of all raw scores obtained, with the results broken down between those infants who did require a break during testing and those who did not.

Figure S4. Identical plots to Figure 3 in the main text, but using a threshold of ± 1.5 *SD* to define episodes of high and low arousal.

Figure S5. Identical plots to Figure 3 in the main text, but using a threshold of ± 0.5 s.d. to define episodes of high and low arousal.

Table S1. Summary results of statistical analysis from Analysis 1.

Table S2. Summary results of statistical analysis from Analysis 1b.

Table S3. Summary results of statistical analysis from Analysis 1.