In infancy, it's the extremes of arousal that are 'sticky': naturalistic data challenge purely homeostatic approaches to studying self-regulation

Wass, Samuel Vincent (1)(*), Smith, Celia Grace (2), Clackson, Kaili (3), Mirza, Farhan Umar (4)

1 - University of East London, UK

2 - Institute of Psychiatry, Psychology and Neuroscience, King's College, London, UK

3 - University of Cambridge, UK

4 - University of Plymouth, UK

Corresponding author: Dr Sam Wass, University of East London, Water Lane, London E15 4LZ. Telephone: +44(0)7725369189. Email: s.v.wass@uel.ac.uk

Acknowledgements: This research was funded by ESRC grant number ES/N017560/1, by ERC grant number ONACSA 853251 and by a Project Grant RPG-2018-281 from the Leverhulme Trust. We wish to thank all families who participated in the research; and Joan Eitzenberger and Caitlin Gibb for performing data analysis.

Data availability: The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Research highlights

- we recorded day-long spontaneous fluctuations in autonomic arousal in 12-month-old infants
- we found that both low- and high-arousal states were more persistent than intermediate arousal states
- one possible explanation for these findings is that extreme arousal states have intrinsically greater hysteresis
- another is that, through 'metastatic' processes, small initial increases and decreases in arousal may become progressively amplified over time.

Abstract

Most theoretical models of arousal/regulatory function emphasise the maintenance of homeostasis; consistent with this, most previous research into arousal has concentrated on examining individuals' recovery following the administration of experimentally administered stressors. Here, we take a different approach: we recorded day-long spontaneous fluctuations in autonomic arousal (indexed via electrocardiogram, heart rate variability and actigraphy) in a cohort of 82 typically developing 12-month-old infants while they were at home and awake. Based on the aforementioned models, we hypothesised that extreme high or low arousal states might be more short-lived than intermediate arousal states. Our results suggested that, contrary to this, both low- and high-arousal states were more persistent than intermediate arousal states. The same pattern was present when the data were viewed over multiple epoch sizes from 1 second to 5 minutes; over 10-15-minute time-scales, high-arousal states were more persistent than low- and intermediate states. One possible explanation for these findings

is that extreme arousal states have intrinsically greater hysteresis; another is that, through ‘metastatic’ processes, small initial increases and decreases in arousal can become progressively amplified over time. Rather than exclusively studying recovery, we argue that future research into self regulation during early childhood should instead examine the mechanisms through which some states can be maintained, or even amplified, over time.

Keywords: self-regulation, arousal, emotion regulation

Introduction

Animal research suggests that neural control of our arousal/regulatory systems involves regions from the brainstem (the medulla, pons (locus coeruleus) and midbrain) to the forebrain via both the hypothalamus and the thalamus (Pfaff, 2018), as well as neurotransmitter systems including noradrenaline (norepinephrine) (Aston-Jones & Bloom, 1981; Aston-Jones & Cohen, 2005; Aston-Jones, Rajkowski, & Cohen, 1999) and acetycholine (Trofimova & Robbins, 2016). Although in recent years the tendency has been to emphasise the fractionation of different arousal/regulatory subsystems (Janig & Habler, 2000; Lacey, 1967; Roubinov, Boyce, Lee, & Bush, 2020; Schneirla, 1946; Trofimova & Robbins, 2016), there is also evidence that they operate at least partially as a unitary system (Calderon, Kilinc, Maritan, Banavar, & Pfaff, 2016; Pfaff, 2018). For example, behavioural animal researchers quantified three types of behaviour traditionally associated with arousal/regulatory systems: motor activity (distance travelled, total movement duration); sensory responses to external stimuli (e.g. auditory, vestibular, tactile and olfactory); and emotional responses (e.g. to conditioned fear paradigms) (Calderon et al., 2016; Pfaff, 2018). They applied Principle Components Analyses to high-throughput analyses of behaviour. Their results suggested that a significant Generalised Arousal component accounts for between 29 to 45% of the variance in behaviour across studies (reviewed (Calderon et al., 2016; Pfaff, 2018)). These three behavioural attributes also show unitary covariance in studies of selective animal breeding (Weil, Zhang, Hornung, Blizard, & Pfaff, 2010). In physiology, the different peripheral measures commonly used to measure arousal (including heart rate, heart rate variability, electro-dermal activity and movement), all show significant tonic and phasic covariance in infants (Wass, Clackson, & de Barbaro, 2016; Wass, de Barbaro, & Clackson, 2015), despite the fact that they are also thought to tap partially

different systems (Cacioppo, Tassinary, & Berntson, 2000). Again, this merits the treatment of arousal as a construct that shows a single common factor as well as more fine-grained sub-factors.

The function of our arousal/regulatory systems is to perform reflexive, adaptive changes so that we can maintain stability in the face of changing environmental demands – a process known as allostasis (Cannon, 1929; Fiske & Maddi, 1961; Gunnar & Quevedo, 2007; McEwen & Seeman, 1999; McEwen & Wingfield, 2003; Porges, 1995; Selye, 1951; Ulrich-Lai & Herman, 2009). To do this, they act over diverse time-scales and using diverse methods, including both slow-acting endocrine systems (such as the Hypothalamic-Pituitary-Adrenal (HPA) axis (Gunnar & Quevedo, 2007), and fast-acting neural systems (such as the Autonomic Nervous System) (Cacioppo et al., 2000). Recent research has started to uncover in detail the mechanisms through which the different subcomponents of our arousal systems interact in humans (Atkinson, Jamieson, Khoury, Ludmer, & Gonzalez, 2016; Berntson, Cacioppo, Quigley, & Fabro, 1994; Del Giudice, Ellis, & Shirtcliff, 2011), and to investigate how these interactions differ between individuals (Quas et al., 2014; Roubinov et al., 2020). This research is uncovering important links between, for example, different profiles of change across different subcomponents of our arousal systems and socioemotional outcomes (Beauchaine & Thayer, 2015; Cole, Ramsook, & Ram, 2019; Kolacz, Holochwost, Gariepy, & Mills-Koonce, 2016; Roubinov et al., 2020).

Most of our understanding about the mechanisms through which arousal/regulatory systems respond to environmental influences has taken place in laboratory settings (Atkinson et al., 2016; Cole, Bendezú, Ram, & Chow, 2017; Obradović, Bush, & Boyce, 2011; Wass, 2014). To do this, the majority of studies take the same approach: they administer a discrete

experimental stimulus, chosen as a simulacrum of a real-world stressor, within a controlled laboratory setting. And they measure regulation and reactivity by indexing the change between a baseline period and the stimulus.

This baseline-stimulus-baseline method may appear intuitive given theoretical approaches that emphasise the role of arousal/regulatory systems in allostasis (McEwen & Wingfield, 2003). However, it has a number of fundamental limitations (Cole, Ram, & English, 2019; Cole, Ramsook, et al., 2019). First, the pattern of a discrete exogenous stimulus that suddenly appears and then disappears after a short interval only occurs rarely in the real world. Second, and relatedly, in the real world we continuously recalibrate, actively selecting what aspects of the social environment that we attend to from one second to the next (Cole, Ram, et al., 2019). The fact that both the appearance of the stimulus, and its disappearance, are outside participant control in experimental settings again lacks ecological validity. Third, almost all research has used experimental stimuli intended to elicit increases in arousal, to examine how we downregulate following increased arousal; the question of how decreases in arousal (whether exogenously or endogenously triggered) are upregulated has been discussed (Fiske & Maddi, 1961) but rarely studied.

As a result of these limitations, an increasing number of studies with humans are starting to use different methods to measure self-regulation (Cole et al., 2017; Cole, Lougheed, Chow, & Ram, 2020; Morales et al., 2018). For example, in one previous study, rather than taking an event-locked approach, they instead analysed spontaneous fluctuations in arousal in 12-month-old infants during the presentation of age-appropriate (and not deliberately anxiogenic) static and dynamic viewing materials (Wass, Clackson, & Leong, 2018). Arousal was measured via a composite of peripheral measures of physiological arousal including

electrocardiography (inter-beat interval), heart rate variability, electro-dermal activity and movement. They reasoned that if transitions in autonomic arousal are purely random, with deviations above and below the mean corrected for via allostasis, then arousal should be normally distributed across the session. In fact, they found that: 1) arousal was positively skewed, and 2) increases in arousal had a lower extinction probability than decreases in arousal. They argued that their results suggested that increases in arousal might be intrinsically self-sustaining, even in the absence of identifiable environmental stressors. Thus, for example, oppositional inter-personal interactions might be both a consequence, and a cause, of elevated arousal (Potegal, Carlson, Margulies, Gutkovitch, & Wall, 2009) – operating not as allostatic processes, but as 'metastatic' ones (from the Greek word meta, meaning 'beyond'), in which small fluctuations become progressively amplified over time.

This previous analysis was, however, based on relatively small segments of data (c. 20 minutes per participant), collected during the administration of an attention battery in laboratory settings. Further, the analyses did not directly examine how the time series fluctuated, but merely examined the distributional properties of the dataset. Lastly, since during viewing parents were asked to keep their infant in a seat or on their lap, they also do not provide a good proxy for real-world settings, where infants are allowed to roam freely (Zubek & MacNeill, 1966).

Here, we examine for the first time how 12-month-old infants' autonomic arousal spontaneously fluctuates in real-world, fully naturalistic settings, across day-long recordings. To measure this, we developed wireless miniaturised wearable autonomic monitors to record electrocardiogram (inter-beat interval), heart rate variability and actigraphy, along with cameras and GPS sensors to allow us to observe where the infant was and what they were

doing. Experimenters visited participants' homes in the morning to fit the equipment and returned in the evening to pick it up; otherwise, participants were encouraged to behave exactly as they would on a normal day. Because, when they were not home, infants were generally strapped into car seats or prams, which influenced their physiological recordings, we concentrated on segments where they were at home, and awake. During these periods, infants were generally able to move as they liked. Motivated by previous research suggesting that arousal-behaviour interactions are scale-free (Proekt, Banavar, Maritan, & Pfaff, 2012) we used the same approach to examine change across multiple time-frames (Cole et al., 2020; Ram & Diehl, 2015), from 1-second (which was the highest time resolution typically used in autonomic data analyses) through to 15-minute epoch durations (which was the longest epoch duration that gave sufficient epochs for analysis, given that on average we had c.3-4 hours of available data per participant).

Specifically, we examined a question that builds on this previous research, but which has received relatively little attention in the literature: how the stability of infants' arousal varies contingent on their fluctuating levels of tonic (baseline) arousal, in naturalistic settings. In the absence of previous research that has looked at naturalistic arousal fluctuations we made the naïve prediction that, if spontaneous arousal fluctuations are random, with fluctuations above and below the mean corrected for via allostasis, then, across multiple time-scales, extreme high or low arousal states would be more short-lived than intermediate states.

Methods

Participants

Participants consisted of a socio-economically diverse cohort of 93 12-month-old infants recruited from the London, Essex, Hertfordshire and Cambridge regions of the UK. Due to equipment errors (see SM section 1.2), no usable ECG data was obtained from eleven participants, and so the total usable sample size for both Analysis 1 and Analysis 2 is 82. Demographic details and exclusion criteria and are given in the Supplementary Materials (section 1.1, Table S1).

Experimental Protocol

Participating parents selected a day for which they would be spending the entire day with their child but that was otherwise, as far as possible, typical. The researcher visited the participants' homes in the morning (between 7.30 and 10am) to fit the equipment and explain its use, and then returned in the late afternoon (between 4 and 7pm) to remove it. Mean *(std)* overall recording time per day was 7.3 *(1.4)* hours.

The equipment consisted of two wearable layers (see Figure S1). A specially designed baby-grow was worn next to the skin, containing a single, integrated recording device consisting of an ECG recorder, Actigraph, GPS, and microphone. A T-shirt, worn on top of the device, contained a pocket to hold the microphone and a miniature video camera. The camera was commercially available (a Narrative Clip 2); the remaining equipment was specially manufactured for this study. The clothes were comfortable when worn and, other than a request to keep the equipment dry, participants behaved as they would on a normal day.

Data pre-processing

Autonomic data. See Figure 1 for an example of the raw data recorded. Three peripheral (non-invasive) measures of autonomic arousal in humans were recorded:

i) *electro-cardiography (ECG)*, from which heart rate measured as inter-beat intervals (IBI) in Beats Per Minute (BPM) was derived. ECG was recorded using three standard Ag-Cl electrodes, placed in a modified lead II position.

ii) *Heart Rate Variability (HRV)*, which was derived from the IBI data. HRV was taken as the Root Mean Square of Successive Differences (RMSSD).

iii) *Actigraphy* was recorded using a tri-axial accelerometer that was positioned approximately on the solar plexus of the infants, and held securely in place in the baby-grow. Movement data were low-pass filtered at 0.1Hz.

Further details of how these three measures were recorded and parsed are given in the Supplementary Materials (section 1.2).

INSERT FIGURE 1 HERE

The preliminary analyses suggested that the three autonomic measures showed strong patterns of tonic and phasic covariation, consistent with previous research (Wass et al., 2016; Wass et al., 2015). See further details, and discussion, in the Supplementary Materials (SM) (section 1.4 and Figure S4). Motivated by this, we collapsed the autonomic indices into a single composite measure for Analysis 1.

Awake/sleeping segments

Our analyses suggested that when infants were outside they were often strapped into a buggy or car seat, which strongly affected their autonomic data. Because of this, only segments of the data when infants were at home were included. Since all aspects of autonomic arousal differ markedly between waking and sleeping (see e.g. Figure 1), sections where infants were asleep were also excluded. Details of the criteria through which home/not-home and sleeping/waking segments were identified are given in the SM (section 1.5). Following these exclusions, the mean (std) amount of data entered into the analyses was 3.7 hours (1.7 hours) per participant (see further details in Results section, below).

Epoching and binning

We epoched our data at multiple time-scales using median windowing into the following epoch durations: 1 second, 10 seconds, 30 seconds, 60 seconds, 300 seconds, 600 seconds, 900 seconds. The choice of these epoch durations was constrained by necessity: 1 second is the most fine-grained time resolution with which autonomic data can reliably be examined, and 15 minutes is the most coarse-grained time resolution that allowed a sufficient number of epochs to be available for analysis, based on a rule that a minimum of 15 epochs should be available overall. If more than 25% of the sample was missing or unavailable (for example, if the infant was not home, or asleep) then the entire epoch was excluded; otherwise, it was retained. Following the exclusions described above, the total mean (sterr) number of epochs available for each analysis ranged from 13264 (691) epochs per participant for the 1-second epoch analysis through to 15.4 (0.8) epochs per participant for the 15-minute analysis.

Following epoching, the data were sorted into evenly sized bins, so that the number of elements within in bin was the same. This was done using the Matlab function tiedrank.m.

(Of note, this means that the absolute cut-off values therefore differed between bins - see the third paragraph of the Discussion for a more detailed discussion of this point.)

Comparison of real data with control data

For our analyses, we wished to compare the dynamics of the observed arousal data with surrogate time series that matched the original data as closely as possible, except for the parameter of interest – which, in this case, was the temporal interdependencies within the time series. To do this, we took two approaches to generating a surrogate dataset. For the first (control 1), the mean and variance of the available data was calculated separately for each participant, and these parameters were used to generate a random time series with the same length, mean and variance as the original data using the 'normrnd' function in Matlab (see Figure 2a, 2b). For the second (control 2), the real data were simply temporally translocated, participant by participant (see SM section 2.1).

Of note, however, there are two features of the original data that neither of these two control analyses matches. The first is that the original data were generally mildly negatively skewed (see Figure 2c and SM section 1.6, Figure S5). However, we can see no way in which this feature of the data might have contributed to the observed results. The second is that autocorrelation was present in the real data (Figure 2d), but not in the surrogate data. In the SM (section 2.3), we present an additional analysis in which we removed the autocorrelation present in the arousal data by fitting autoregressive models and repeated our main analysis, in order to investigate how this feature of the data may have influenced our results.

INSERT FIGURE 2 HERE

Results

Analysis overview

Our aim was to examine how the stability of infants' arousal varied contingent on their fluctuating levels of tonic (baseline) arousal. We hypothesised that, across all time scales, extreme high or low arousal states would be more short-lived than intermediate states. To test this, we conducted two analyses: Analysis 1 and Analysis 2.

In analysis 1, we binned the arousal data participant by participant into 5 equally sized bins, in order to control for individual differences in average arousal. We then calculated, separately for each arousal bin at time t, the likelihood of the bin at time t+1 being the same as the bin at time t. We then plotted this 'change likelihood' score against the arousal bin at time t. In the Supplementary Materials, we repeat the same analyses, but based on data from the different individual arousal measures (heart rate, heart rate variability, actigraphy) rather than the arousal composite (SM section 2.2), and based on data from which the autocorrelation had been removed (SM section 2.3).

One possibility left open by analysis 1 is that the observed results were due to the fact that a composite measure of arousal was used. For example, it is possible that individual constituent measures may have contributed inconsistently to the composite measure across low, intermediate and high values (although see also SM, section 2.2). A second possibility left open by analysis 1 is that the results observed are in some way an artefact of the fact that the arousal data were binned participant by participant prior to conducting the analysis (although it is unclear how this could have created the pattern of observed results).

To examine these possibilities, we conducted analysis 2, in which we examined two individual measures (actigraphy and ECG), after they were binned separately into 10 equally sized bins (i.e., not controlling for differences in average levels). We then plotted bivariate vector plots to examine how the patterns of change on the two measures varied both independently, and in combination with each other.

Both Analysis 1 and Analysis 2 address the same question, which is to assess how the stability of infants' arousal varies contingent on their fluctuating levels of tonic (baseline) arousal.

Analysis 1

Data were downsampled into epochs of different sizes. The time scales used varied from 1 second epochs (i.e., examining change in arousal between consecutive 1-second epochs) through to 15-minute epochs (i.e., examining change in arousal between consecutive 15-minute epochs), in the following increments: 1 second, 10 seconds, 30 seconds, 60 seconds, 300 seconds, 600 seconds, 900 seconds. Further details of the epoching are given in the Methods.

After the data were epoched, they were binned into five equally sized epochs on a per-participant basis (see Figure 2b). Figure 3a shows an adapted Poincaré plot in which arousal bin at time t (x-axis) is plotted against arousal at time t+1 (y-axis). It can be seen that the most densely populated area is around the 1:1 line, indicating that the data show autocorrelation. This is as expected based on the autocorrelation plot in Figure 2d. The

possibility that this aspect of the data may have investigated our main results is investigated in SM section 2.3.

INSERT FIGURE 3 HERE

To address our primary research question we performed the following analysis. The results of this analysis are illustrated in detail using just one, intermediate epoch duration (60 seconds) in Figures 3b-3f. Summary results based on the same analysis, but using variable epoch durations, are shown in Figure 4.

We calculated, separately for each arousal bin at time t, the likelihood of the bin at time t+1 being the same as the bin at time t. We then plotted this 'change likelihood' score against the arousal bin at time t. Figure 3b shows the results of this analysis for the real and the control data. It can be seen that, for the control 1 data, this likelihood is flat at 0.2 – as expected given that the data are randomly distributed. Control 2 data, shown in SM section 2.1, are highly similar. In the real data, in comparison, two features can be seen:

The first is that, for all bins, the values for the real data appear greater than those for the control data (see Figure 3b and Figure S6). If true, this would reflect merely that the arousal show autocorrelation, as is already shown in Figures 2d and 3a. In order to test the significance of this, the values obtained for each participant and for each arousal bin were compared between the real and the control data. Since the between-participant variance was not always normally distributed, the more conservative nonparametric Wilcoxon Signed Rank Test was used throughout. Multiple comparisons were corrected for using the Benjamini Hochberg False Discovery Rate (FDR) procedure (Benjamini & Hochberg, 1995).

In the example shown (based on 60-second epochs), all bins of the real data are significantly ($p<.001$) higher than those for the control data.

The second feature of interest is that the extreme bins (bins 1 and 5) appear to show values that are higher than those in intermediate bins. This is the primary feature of interest in this paper. If true, this would reflect that extreme low- and high-arousal states tend to be more persistent than intermediate arousal states. In order to test the significance of this, the values obtained for each participant and for each arousal bin were compared directly between bins; this was repeated pairwise for each possible combination of bins (bin 1 vs 2, 3, 4, 5; bin 2 vs 3, 4, 5 etc). Since not all results were parametrically distributed, the more conservative Wilcoxon Signed Rank Test was used throughout, and multiple comparisons were corrected for using the Benjamini Hochberg False Discovery Rate (FDR) procedure (Benjamini & Hochberg, 1995). The significance of these tests is shown in Figures 3c and 3d. In the real data (Fig 3c) all possible combinations of bins differ, with the exception that bin 3 does not differ significantly from bin 4. No equivalent comparisons are significant in the control data (Fig 3d).

As an additional visualisation, Fig 3e and 3f show the likelihood of arousal bin at t+1 being less (Fig 3e) or greater than (Fig 3f) arousal bin at t. Given that the control data are randomly distributed, it is to be expected that at bin 5 the likelihood of the next sample being lower is 0.8, and that at bin 2 it is 0.2. In comparison, the real data show that, at bin 2, it is more likely than chance that the next sample will be lower; and that, at bin 5, it is less likely than chance that the next sample will be lower (Fig 3e). The opposite pattern of results is shown in Fig 3f. These results are consistent with the findings in Fig 3b, that both extremes show higher stability rates than expected.

Figure 4 shows the results of the same analysis as shown in Fig 3b, but using variable epoch durations. Figures 4b and S6a shows the results of the same analysis applied to two different versions of the control data. Figures 4c and 4d show how the likelihood of decreases (4c) or increases (4d) in arousal varies as a function of arousal, relative to control data (drawn in grey). Figures 4e and S6d show the results of significance tests, conducted as described in the Methods, to examine whether the results shown in Figure 4a differed significantly from the control analysis. Figure 4f shows the results of significance tests to examine how the results shown in Fig 4a differed pairwise between bins.

INSERT FIGURE 4 HERE

From these figures, the following conclusions can be drawn. First, results obtained at all epoch durations from 1 second to 300 seconds (Fig 4a) showed consistently higher stability in arousal than the control data (Fig 4b/S6a). This reflects the fact that, as shown in Figure 2d, autocorrelation is present. Second, arousal stability was higher at shorter epoch durations than at longer epoch durations. Again, this reflects the autocorrelation. Third, at longer epoch durations (600/900 seconds), less consistent differences in arousal stability between the real and control data are observed. Again, this reflects that autocorrelation would have a smaller influence on analyses based on longer epoch durations.

Of primary interest, however is the comparison between bins, which shows how arousal stability varies as a function of arousal. At epoch durations from 1 to 300 seconds, consistent significant differences between all bins were observed, such that bin 1>bin2>bin3<bin4<bin5 (Fig 4f). At shorter epoch durations (1 to 60 seconds), differences are observed between all

bins; at 300 seconds, bins 1 and 5 (extreme high and low arousal) differ from other bins (bin 1>bins 2-4; bin 5>bins 2-4), but otherwise no significant differences are observed. At 600 and 900 seconds, bin 5 (extreme high arousal) shows higher stability than other bins (bin 5>bins 1-4). Overall, these results suggest that, when the data were viewed over multiple epoch sizes from 1 second to 5 minutes, both low- and high-arousal states were more persistent than intermediate arousal states. Over 10-15-minute time-scales, high-arousal states were more persistent than low- and intermediate states.

In the Supplementary Materials (section 2.2, Figures S7, S8 and S9) we present the identical analyses to those shown in Figure 4, but conducted based on data from the different individual measures included in the arousal composite. Although a number of differences between these measures can be seen, contingent on the varying degrees of autocorrelation present in the different datasets, the chief feature of interest for this analysis is observed highly consistently across different measures, and different epoch durations.
In the Supplementary Materials (section 2.3, Figure S10) we also present identical analyses to those shown in Figure 4, but conducted based on data from which the auto-correlation had been removed by fitting autoregressive models. Results obtained from this analysis were similar, indicating that the effect is not an artifact of the autocorrelation in the data.

Analysis 2

In order to confirm the validity of the findings from Analysis 1, we also conducted a separate analysis to address the same question, by using bivariate vector plots. Instead of being based on a composite measure of arousal, this analysis examines patterns of change in the heart rate

and the actigraphy measures both individually and in combination with one another, thus demonstrating that any conclusions from Analysis 1 are not an artifact of the composite arousal measure used. And instead of being based on data binned on a per-participant measure, it is based on data binned across all participants, thus demonstrating that any conclusions from Analysis 1 are not an artifact of the binning procedure used. As with Analysis 1, the analysis was conducted only on segments when the infant was at home and awake.

To calculate the vector plots, all heart rate and actigraphy data were averaged separately. In the analyses presented in the main text, an intermediate epoch duration of 60 seconds was selected; analyses in the SM section 2.4 (Figure S11) show the same analyses repeated with the same variable epoch durations as used in Analysis 1.

After epoching, data were concatenated across all participants; heart rate and actigraphy data were then separately divided into ten equally sized bins, based on the entire concatenated data (see Figure 5a-c). Each epoch was then classified according to what bin it fell into for both heart rate and actigraphy. This is represented as a two-dimensional matrix – so all epochs that were bin 3 for heart rate and bin 4 for actigraphy are drawn at location [x -3, y - 4]. The size of each blue dot within the matrix indicates what proportion of the total available samples was located within each bin – so a larger blue dot at [3,4] than at [6,1] indicates that epochs at bin 3 for heart rate and bin 4 for Actigraphy were more commonly observed than epochs at bin 6 for heart rate and bin 1 for Actigraphy.

In addition, for each bin, we calculated the average change between all epochs in that bin and the epoch immediately following. This change score is drawn on the vector plot as a red line.

Thus, for the point located at [10,10] on the vector plot, which represents all epochs that were classified as in bin 10 for both heart rate and actigraphy, the vector extends -0.9 on the x-axis (representing change in heart rate), and -1.1 on the y-axis (representing change in actigraphy). This indicates that, across all epochs starting from [10,10], the average change to the next epoch was a reduction of 0.9 bins in heart rate, and 1.1 bins in Actigraphy.

Figure 5a-c gives an illustrative sample of three individual datasets to which this analysis has been applied. All data are grouped around the 1:1 equivalence line as heart rate and actigraphy are strongly correlated (see SM section 1.2). In addition, and because, as described above, binning was conducted on the entire concatenated dataset, it can be seen that participant #1 showed higher average levels for actigraphy and heart rate relative to other participants, that participant #3 showed lower heart rate but not lower actigraphy levels than other participants, and so on.

Our primary aim was to examine how, across the entire dataset, the change vectors for both heart rate and actigraphy varied as a function of bin. To assess this, we conducted a similar calculation as for Analysis 1: the change vectors obtained for each participant and for each bin were directly compared; this was repeated pairwise for each possible combination of bins (bin 1 vs 2, 3, 4, 5; bin 2 vs 3, 4, 5 etc). Multiple comparisons were corrected for using the Benjamini Hochberg False Discovery Rate (FDR) procedure (Benjamini & Hochberg, 1995).

INSERT FIGURE 5 HERE

Figure 5d shows the full vector plot for the control data; Figure 5e shows the full vector plot for the real data. Figure 5f shows arousal change as a function of arousal bin for the control

data; Figure 5g shows the same plot for the real data. Figure 5h shows the results of pairwise bin comparisons to examine how arousal change varies as a function of arousal bin for the control data; Figure 5i shows the same plot for the real data.

From these plots the following points can be seen. First, the control plot shows long vectors pointing towards the centre of the plot. This indicates regression to the mean: the data have been randomly generated (see Method) and so it is likely that an epoch that is high will be followed due to random fluctuation by an epoch that is lower. Second, the fact that the vectors are shorter in the real data (Fig 5e) is because the real data show autocorrelation; the fact that they nevertheless radiate inwards shows that the real data also show regression to the mean. Third, the centre point of the inwards-radiating vectors is bin [6,6], rather than [5,5], which is because both measures show a mild negative skew (see SM section 1.5). Fourth, the slight swirling pattern in the data is because Actigraphy has a lower autocorrelation (faster rate of change) than heart rate, and so that changes in actigraphy tend to anticipate changes in heart rate (Wass et al., 2016).

The key point of interest for the present paper, however, is how the change vectors vary as a function of bin. Fig 5f shows that, in the control data, these decrease linearly with increasing bin, due to regression to the mean: epochs where arousal is high are likely to be followed by a decrease in arousal, and epochs where arousal is low are likely to be followed by an increase in arousal. The significance plots (Fig 5h) indicate this as a highly significant pattern in the data (bin 1>bin 2>bin 3, etc). In the real data, in contrast (Fig 5g) the same overall pattern of regression to the mean is present in the data, but falls off at the extreme high and low bins. This is particularly true for lower bins: the significance plots indicate that, for both heart rate and actigraphy, bin 1 shows a significantly *lower* rate of change than some intermediate bins

(bin 3/4) – a pattern not seen in the control data. The same pattern can also be seen in the higher bins: there is no significant difference between bins 9 and 10 in change scores for either actigraph or heart rate (Fig 5i), whereas there is in the control data. In the SM section 2.4 Figure S11 we see that consistent patterns are observed across variable epoch durations, from 1 second to 15 minutes. Overall, these results are consistent with Analysis 1, insofar as they suggest that extreme low- and high-arousal states were more persistent than intermediate arousal states.

Discussion

We recorded spontaneous fluctuations in autonomic arousal across day-long data segments in naturalistic settings. Our participants were 12-month-old infants from mixed demographic backgrounds (see Table S1). Only segments of the data when the infant was at home and awake were analysed. In the absence of any previous data on how infant arousal fluctuates in naturalistic settings we made the naïve prediction that, if transitions in arousal are random, with deviations above and below the mean corrected for via allostasis, then extreme high- and low-arousal states might be more short-lived than intermediate states. In fact, we found the opposite: consistently across two separate analyses, we found that extreme high and low states were more persistent than intermediate states. The same pattern was observed at multiple time-scales, from 1-second to 5-minute epochs (Nakamura et al., 2007; Proekt et al., 2012), although at longer epoch durations (10- and 15-minute) only extreme high arousal states were more persistent than intermediate and low arousal states.

In supplementary analyses we showed that the effect is not an artifact of how the control data were generated (SM section 2.1), is observed consistently across different individual autonomic variables (SM section 2.2 and Analysis 2), and is still observed even when the autocorrelation is removed from the data (SM section 2.3). We also showed that the same pattern is observed both when data are binned separately for each participant (in order to control for differences in average arousal) (Analysis 1) and when they are not (Analysis 2).

One remaining possible artifactual cause for our finding could be that the raw arousal data are from a Gaussian distribution, but divided into five equal-sized bins. In a Gaussian distribution, between-samples variance would naturally be greater in the extreme bins; it is

plausible that this might cause our finding that extreme high and low states are more persistent. However, the Control analysis 1 is designed to preclude this possibility, as the control data also come from a Gaussian distribution with matched variance (see SM section 2.1).

Previous research has shown that, when infants are sitting on their parents' lap watching an experimental test battery, with the parents asked to restrain the infant if they try to get down, then only high arousal states are more persistent than expected (Wass et al., 2018). Our findings are consistent with this, but suggest that, in naturalistic home settings, both low- and high arousal states are more persistent than expected. Indeed, of the two extremes, low arousal states seemed even more stable than high arousal states (Fig 5g, 5i).

There are two possible explanations for our finding. The first is that extreme arousal states are intrinsically slower-decaying – or show different hysteresis. This possibility has certainly been acknowledged – for example, Hobson and Steriade noted that "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). Sleep is a more stable arousal state than others (Saper, Fuller, Pedersen, Lu, & Scammell, 2010); although sleeping sections were excluded from our data in this study, it is possible that other arousal states (including both extreme high and low states) may also show differing intrinsic hysteresis.

An alternative possibility is that both extreme low and high arousal states may lead to changes in how we interact with the external environment and with people (Cole et al., 2017) - changes that may in turn lead to extreme arousal states becoming progressively amplified

over time. Adult research into emotion regulation offers several models from adult psychopathology for cognitive reinforcement loops within anxiety disorders such as panic disorder, for example - showing how processes such as rumination, attention biases or maladaptive safety-seeking behaviours (Gross & Feldman Barrett, 2011; Pine et al., 2005) can lead to 'metastatic' processes, in which small increases in arousal become progressively amplified over time (Margraf, 1993; Salkovskis, 1991). Similar ideas are, however, less well developed within the context of child self-regulation (Cole, Ram, et al., 2019; Cole, Ramsook, et al., 2019).

To illustrate the difference between allostatic processes (in which increases and decreases in arousal are corrected for over time) and metastatic processes (in which increases and decreases in arousal become amplified over time), we can consider the case of toy removal. This is well-studied as an experimental test of self-regulation in children (Gagne, Van Hulle, Aksan, Essex, & Goldsmith, 2011). Most researchers hitherto have studied childrens' reactions to toy removal within the context of allostatic mechanisms. Thus, when an exogenous stressor occurs that increases our arousal, we respond with behaviours intended to downregulate arousal, to compensate (a negative feedback process). Widely studied downregulatory behaviours include self-soothing and gaze aversion, which has been shown even by 5-month-old infants in some contexts (Buss & Goldsmith, 1998; Stifter & Braungart, 1995).

Consider, though, an alternative scenario for how a child might respond when they have fixed their eyes on a toy they want while out shopping. Although they might respond with self-soothing, and gaze aversion, they might also show a very different type of response (Cole, Ramsook, et al., 2019; Potegal et al., 2009). A child might refuse to let go of the toy, and start

crying; their parent, tired and in a hurry, might abruptly say ‘no’, and attempt to take the toy off them, leading to a physical tug of war. The child might lose this, sit down with a bump, and burst out crying. The child might smash the toy on the floor, breaking it. Others in the shop might turn around to look at the noise. This series of events – being abruptly told ‘no’, a tug of war, the toy breaking, being stared at by strangers – are all triggered by the child’s behaviour, as a function of their arousal state (Cole et al., 2017). But they are also independent, exogenous causes of further increases arousal – a positive feedback look (Potegal et al., 2009; Wass et al., 2018).

Similar ‘metastatic’ processes might also explain how *decreases* in arousal might become amplified over time. Thus, for example, a child’s arousal state can influence how they react when a complex or slow-paced new stimulus is presented (Richards, 1987; Van der Meere & Sergeant, 1988) – either engaging with it, or not. At the same time, comprehensible stimuli elicit greater decreases in arousal during presentation compared with incomprehensible ones (e.g. TV programs with the shots correctly ordered vs randomly re-shuffled - (Pempek et al., 2010; Richards, 2010)). This suggests that arousal decelerations can be a *consequence* of understanding in children. Thus, a decrease in arousal might lead to increased endogenous engagement with more complex or slow-paced external stimuli, causing further decreases in arousal – another positive feedback loop.

Relatedly, other research has shown that our susceptibility to external influences fluctuates systematically contingent on our tonic arousal. For example, in animals, tonic activity in the Locus Ceruleus predicts phasic, stimulus-evoked responsiveness, with greatest phasic responsiveness observed at intermediate tonic activity (Aston-Jones & Cohen, 2005; Aston-Jones et al., 1999). Broadly consistent results have been observed with humans (Wass, 2018),

suggesting that phasic autonomic responsiveness to attention-eliciting stimuli is greatest in infants and young children at intermediate states of tonic autonomic arousal. This relationship has, however, only been observed for attention- and not for emotion-eliciting stimuli (Wass, 2018). Furthermore, it has only been demonstrated for phasic responsiveness over the second-by-second scale, and not across the multiple time-scales we used here. However, the suggestion that sensitivity to the external environment is greatest at intermediate arousal, and that both extreme high and low arousal shows less sensitivity to external stimuli, could be consistent with our present findings.

Future research should use more sophisticated approaches to differentiate between these two possibilities: a) that extreme high and low arousal states show different intrinsic stability and b) that extreme states trigger different environmental and interpersonal interactions that then take on a self-sustaining character. Understanding this will help us to enrich our understanding of self-regulation. Instead of relying solely on approaches that minimise environmental influences using an experimenter-controlled baseline-stimulus-baseline approach, we can instead recognise that self-regulation involves ongoing, dynamic interplay between exogenous and endogenous factors, which operate via a process of continuous, dynamic recalibration. Such an approach may offer new insights into how arousal and self-regulation develop over time, from infancy into adulthood, as well as how these processes develop atypically in developmental psychopathology.

References

Aston-Jones, G., & Bloom, F. E. (1981). Activity of norepinephrine-containing locus coeruleus neurons in behaving rats anticipates fluctuations in the sleep-waking cycle. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 1*(8), 876-886.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience, 28*, 403-450. doi:10.1146/annurev.neuro.28.061604.135709

Aston-Jones, G., Rajkowski, J., & Cohen, J. (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biological Psychiatry, 46*(9), 1309-1320. doi:10.1016/s0006-3223(99)00140-7

Atkinson, L., Jamieson, B., Khoury, J., Ludmer, J., & Gonzalez, A. (2016). Stress physiology in infancy and early childhood: cortisol flexibility, attunement and coordination. *Journal of neuroendocrinology, 28*(8).

Beauchaine, T. P., & Thayer, J. (2015). Heart rate variability as a transdiagnostic biomarker of psychopathology. *Int J Psychophysiol, 98*(2 Pt 2), 338-350. doi:10.1016/j.ijpsycho.2015.08.004

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Berntson, G. G., Cacioppo, J. T., Quigley, K. S., & Fabro, V. T. (1994). Autonomic space and psychophysiological response. *Psychophysiology, 31*(1), 44-61. doi:10.1111/j.1469-8986.1994.tb01024.x

Buss, K. A., & Goldsmith, H. H. (1998). Fear and anger regulation in infancy: Effects on the temporal dynamics of affective expression. *Child Development, 69*(2), 359-374.

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

Calderon, D., Kilinc, M., Maritan, A., Banavar, J., & Pfaff, D. (2016). Generalized CNS arousal: an elementary force within the vertebrate nervous system. *Neuroscience & Biobehavioral Reviews, 68*, 167-176.

Cannon, W. B. (1929). Organization for physiological homeostasis. *Physiological reviews, 9*(3), 399-431.

Cole, P. M., Bendezú, J. J., Ram, N., & Chow, S.-M. (2017). Dynamical systems modeling of early childhood self-regulation. *Emotion, 17*(4), 684.

Cole, P. M., Lougheed, J. P., Chow, S.-M., & Ram, N. (2020). Development of Emotion Regulation Dynamics Across Early Childhood: a Multiple Time-Scale Approach. *Affective Science*, 1-14.

Cole, P. M., Ram, N., & English, M. S. (2019). Toward a unifying model of self-regulation: A developmental approach. *Child Development Perspectives, 13*(2), 91-96.

Cole, P. M., Ramsook, K. A., & Ram, N. (2019). Emotion dysregulation as a dynamic process. *Development and Psychopathology, 31*(3), 1191-1201.

Del Giudice, M., Ellis, B. J., & Shirtcliff, E. A. (2011). The adaptive calibration model of stress responsivity. *Neuroscience & Biobehavioral Reviews, 35*(7), 1562-1592.

Fiske, D. W., & Maddi, S. R. (1961). Functions of varied experience.

Gagne, J. R., Van Hulle, C. A., Aksan, N., Essex, M. J., & Goldsmith, H. H. (2011). Deriving childhood temperament measures from emotion-eliciting behavioral episodes: Scale construction and initial validation. *Psychological assessment, 23*(2), 337.

Gardner, J. M., & Karmel, B. Z. (1995). Development of Arousal-Modulated Visual Preferences in Early Infancy. *Developmental Psychology, 31*(3), 473-482.

Gardner, J. M., Karmel, B. Z., & Flory, M. J. (2003). Arousal modulation of neonatal visual attention: Implications for development. *Perspectives on fundamental processes in intellectual functioning, 2*, 125-153.

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.

Gunnar, M., & Quevedo, K. (2007). The neurobiology of stress and development. *Annu. Rev. Psychol., 58*, 145-173.

Hobson, J. A., & Steriade, M. (1986). Neuronal basis of behavioral state control. *Comprehensive Physiology*.

Janig, W., & Habler, H. (2000). Specificity in the organization of the autonomic nervous system: a basis for precise neural regulation of homeostatic and protective body functions. *Progress in brain research, 122*, 351-368.

Kolacz, J., Holochwost, S. J., Gariepy, J. L., & Mills-Koonce, W. R. (2016). Patterns of joint parasympathetic, sympathetic, and adrenocortical activity and their associations with temperament in early childhood. *Developmental psychobiology*. doi:10.1002/dev.21429

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

Margraf, J. (1993). Hyperventilation and panic disorder: a psychophysiological connection. *Advances in Behaviour Research and Therapy, 15*(1), 49-74.

McEwen, B. S., & Seeman, T. (1999). Protective and damaging effects of mediators of stress: elaborating and testing the concepts of allostasis and allostatic load. *Annals of the New York Academy of Sciences, 896*(1), 30-47.

McEwen, B. S., & Wingfield, J. C. (2003). The concept of allostasis in biology and biomedicine. *Horm Behav, 43*(1), 2-15.

Morales, S., Ram, N., Buss, K. A., Cole, P. M., Helm, J. L., & Chow, S. M. (2018). Age-related changes in the dynamics of fear-related regulation in early childhood. *Developmental Science, 21*(5), e12633.

Nakamura, T., Kiyono, K., Yoshiuchi, K., Nakahara, R., Struzik, Z. R., & Yamamoto, Y. (2007). Universal scaling law in human behavioral organization. *Physical review letters, 99*(13), 138103.

Obradović, J., Bush, N. R., & Boyce, W. T. (2011). The interactive effect of marital conflict and stress reactivity on externalizing and internalizing symptoms: The role of laboratory stressors. *Development and Psychopathology, 23*(01), 101-114.

Pempek, T. A., Kirkorian, H. L., Richards, J. E., Anderson, D. R., Lund, A. F., & Stevens, M. (2010). Video Comprehensibility and Attention in Very Young Children. *Developmental Psychology, 46*(5), 1283-1293. doi:10.1037/a0020614

Pfaff, D. (2018). *How Brain Arousal Mechanisms Work: Paths Toward Consciousness* (Vol. 1): Cambridge University Press.

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. (1995). Cardiac vagal tone: a physiological index of stress. *Neurosci Biobehav Rev, 19*(2), 225-233.

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 Hum Dev., 40*(4), 621-636.

Proekt, A., Banavar, J. R., Maritan, A., & Pfaff, D. W. (2012). Scale invariance in the dynamics of spontaneous behavior. *Proceedings of the National Academy of Sciences, 109*(26), 10564-10569.

Quas, J. A., Yim, I. S., Oberlander, T. F., Nordstokke, D., Essex, M. J., Armstrong, J. M., . . . Boyce, W. T. (2014). The symphonic structure of childhood stress reactivity: patterns of sympathetic, parasympathetic, and adrenocortical responses to psychological challenge. *Dev Psychopathol, 26*(4 Pt 1), 963-982. doi:10.1017/s0954579414000480

Ram, N., & Diehl, M. (2015). Multiple-time-scale design and analysis: Pushing toward real-time modeling of complex developmental processes.

Richards, J. E. (1987). Infant visual sustained attention and respiratory sinus arrhythmia. *Child Development*, 488-496.

Richards, J. E. (2010). The development of attention to simple and complex visual stimuli in infants: Behavioral and psychophysiological measures. *Developmental Review, 30*(2), 203-219. doi:10.1016/j.dr.2010.03.005

Roubinov, D. S., Boyce, W. T., Lee, M. R., & Bush, N. R. (2020). Evidence for discrete profiles of children's physiological activity across three neurobiological system and their transitions over time. *Developmental Science*, e12989.

Salkovskis, P. M. (1991). The Importance of Behaviour in the Maintenance of Anxiety and Panic: A Cognitive Account. *Behavioural Psychotherapy, 19*(1), 6-19.

Saper, C. B., Fuller, P. M., Pedersen, N. P., Lu, J., & Scammell, T. E. (2010). Sleep state switching. *Neuron, 68*(6), 1023-1042.

Schneirla, T. C. (1946). Problems in the bio-psychology of social organization. *The Journal of Abnormal and Social Psychology, 41*(4), 385.

Selye, H. (1951). The Physiology and Pathology of Exposure to Stress. *The Journal of Bone and Joint Surgery, 33-A*, 818-819.

Stifter, C. A., & Braungart, J. M. (1995). The regulation of negative reactivity in infancy: function and development. *Developmental Psychology, 31*(3), 448.

Trofimova, I., & Robbins, T. W. (2016). Temperament and arousal systems: a new synthesis of differential psychology and functional neurochemistry. *Neuroscience & Biobehavioral Reviews, 64*, 382-402.

Ulrich-Lai, Y. M., & Herman, J. P. (2009). Neural regulation of endocrine and autonomic stress responses. *Nature Reviews Neuroscience, 10*(6), 397-409. doi:10.1038/nrn2647

Van der Meere, J., & Sergeant, J. (1988). Controlled processing and vigilance in hyperactivity: Time will tell. *Journal of abnormal child psychology, 16*(6), 641-655.

Wass, S. V. (2014). Comparing methods for measuring peak look duration: are individual differences observed on screen-based tasks also found in more ecologically valid contexts? . *Infant Behavior and Development., 37*(3), 315-325.

Wass, S. V. (2018). How orchids concentrate? The relationship between physiological stress reactivity and cognitive performance during infancy and early childhood. *Neuroscience & Biobehavioral Reviews*.

Wass, S. V., Clackson, K., & de Barbaro, K. (2016). Temporal dynamics of arousal and attention in infants. *Developmental psychobiology, 64*, 1-17. doi:DOI 10.1002/dev.21406

Wass, S. V., Clackson, K., & Leong, V. (2018). Increases in Arousal are More Long-Lasting than Decreases in Arousal: On Homeostatic Failures During Emotion Regulation in Infancy. *Infancy*.

Wass, S. V., de Barbaro, K., & Clackson, K. (2015). Tonic and phasic co-variation of peripheral arousal indices in infants. *Biological Psychology, 111*, 26-39.

Weil, Z. M., Zhang, Q., Hornung, A., Blizard, D., & Pfaff, D. W. (2010). Impact of generalized brain arousal on sexual behavior. *Proceedings of the National Academy of Sciences, 107*(5), 2265-2270.

Zubek, J. P., & MacNeill, M. (1966). Effects of immobilization: Behavioural and EEG changes. *Canadian Journal of Psychology/Revue canadienne de psychologie, 20*(3), 316.

Zuckerman, M. (1979). *Sensation seeking*: Wiley Online Library.

Figures

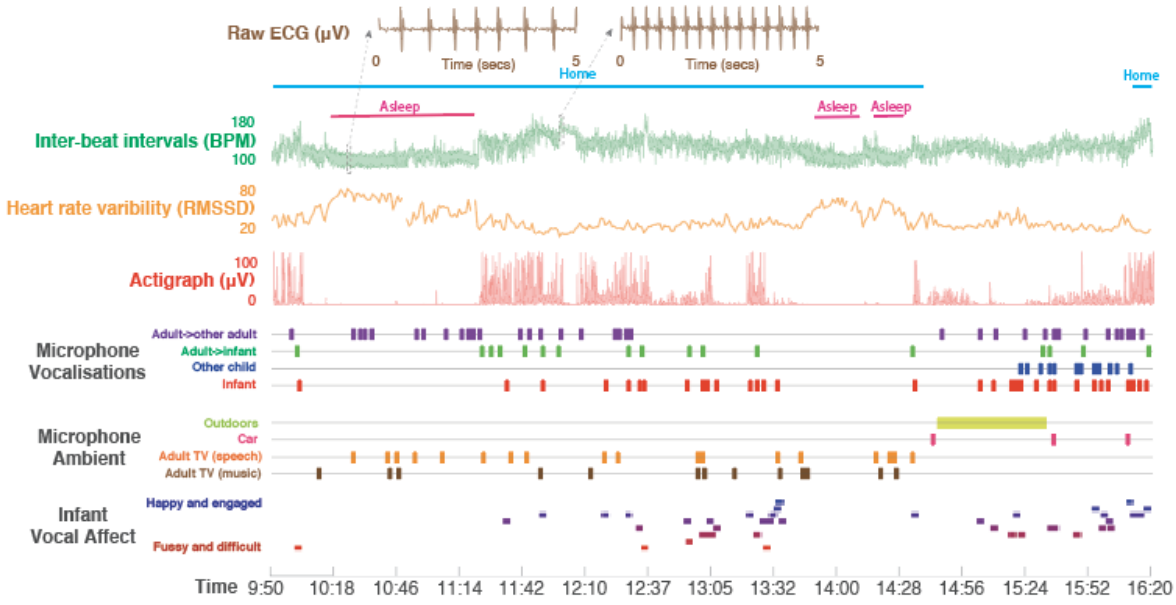

Figure 1: Data sample illustrating the raw data collected from a single participant. Six and a half hours' data is presented (see time axis at the bottom). From top to bottom: the raw electro-cardiography recording; the inter-beat intervals derived from the electro-cardiography recording, indexed as Beats Per Minute (BPM); the heart rate variability (indexed as the Root Mean Square of Successive Differences (RMSSD)); actigraphy (indexed as micro-Volts); the results of coding for of the microphone data (see SM section 1.5).

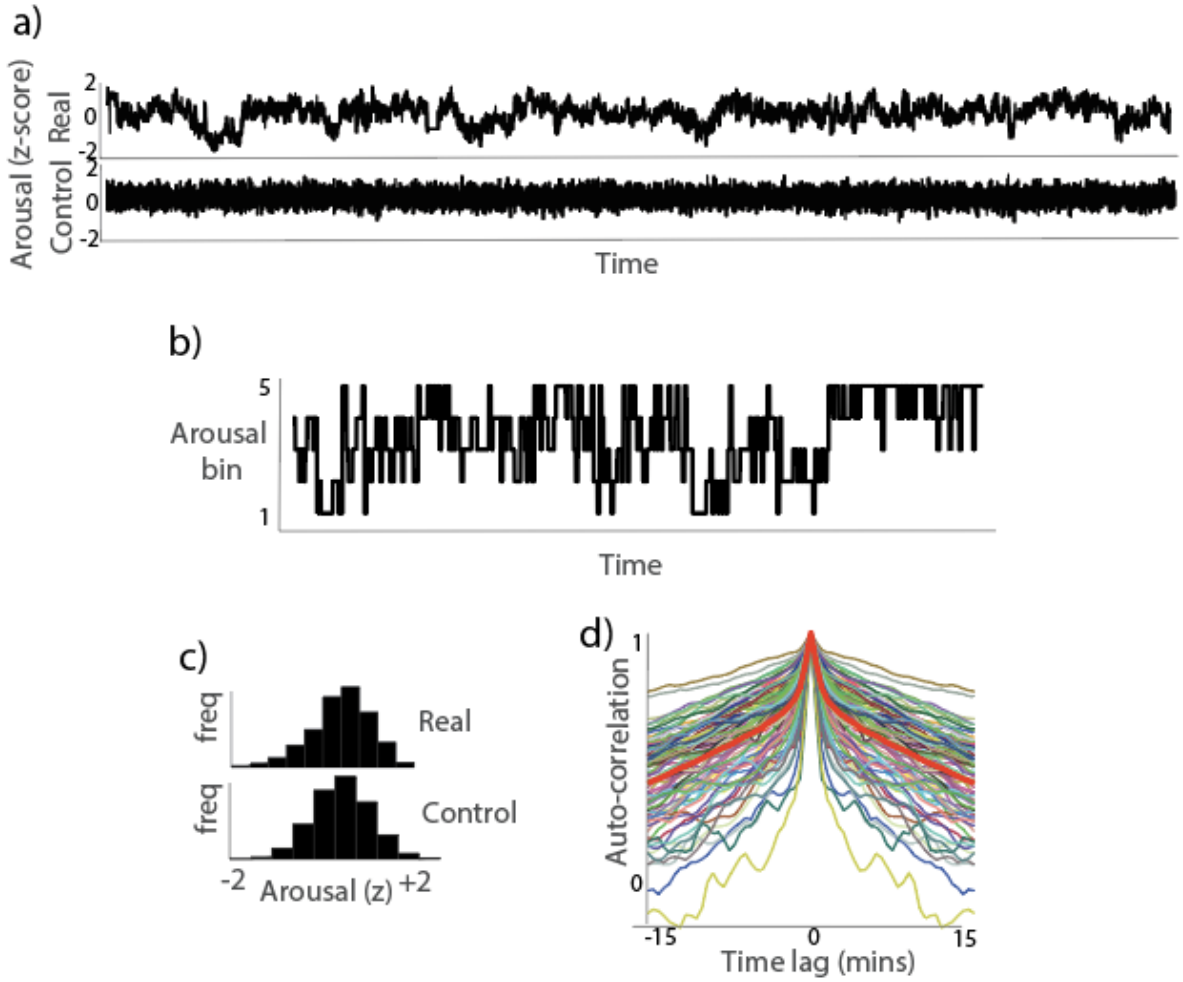

Figure 2: a) example data segment from a single participant showing (top) 14596 samples (=4.05 hours) of real infant arousal data at 1Hz and (below) an equivalent number of samples of Control 1 surrogate data, generated as described in the Methods). b) example excerpt of data from a single participant after data were binned into 5 equally sized bins, based on arousal data downsampled to 60-second epochs. c) histograms showing the distribution of real and control 1 datasets shown in Figure 2a (see also SM section 1.5 and Figure S4). d) Exploratory auto-correlation plot, based on arousal data down-sampled to 60-second epochs. Thick red line shows average autocorrelation; coloured lines show autocorrelation plots for individual participants.

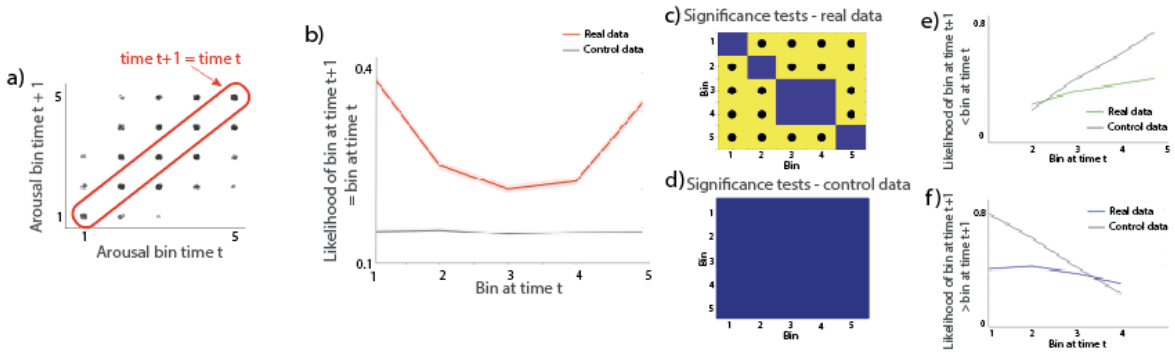

Figure 3: a) Exploratory visualisation of an adapted Poincaré plot in which arousal bin at time t is plotted against arousal bin at time t+1, based on arousal data downsampled to 60-second epochs. b) plot based on arousal data downsampled to 60-second epochs which shows, separately for each arousal bin at time t, the likelihood of time t+1 being the same as time t. Red line shows the real data; grey the control data (see also Figure S6 for alternative control analyses). Shaded areas show Standard Error of the Means. c) and d): results of pairwise significance comparisons between bins for real (c) and control (d) data – yellow squares indicate a significant difference between bins, blue indicate no significant difference. e) plot based on arousal data downsampled to 60-second epochs in which the likelihood that: (bin at time t+1)<(bin at time t) is shown (y-axis) against bin at time t (x-axis). f) plot based on arousal data downsampled to 60-second epochs in which the likelihood that: (bin at time t+1)>(bin at time t) is shown (y-axis) against bin at time t (x-axis).

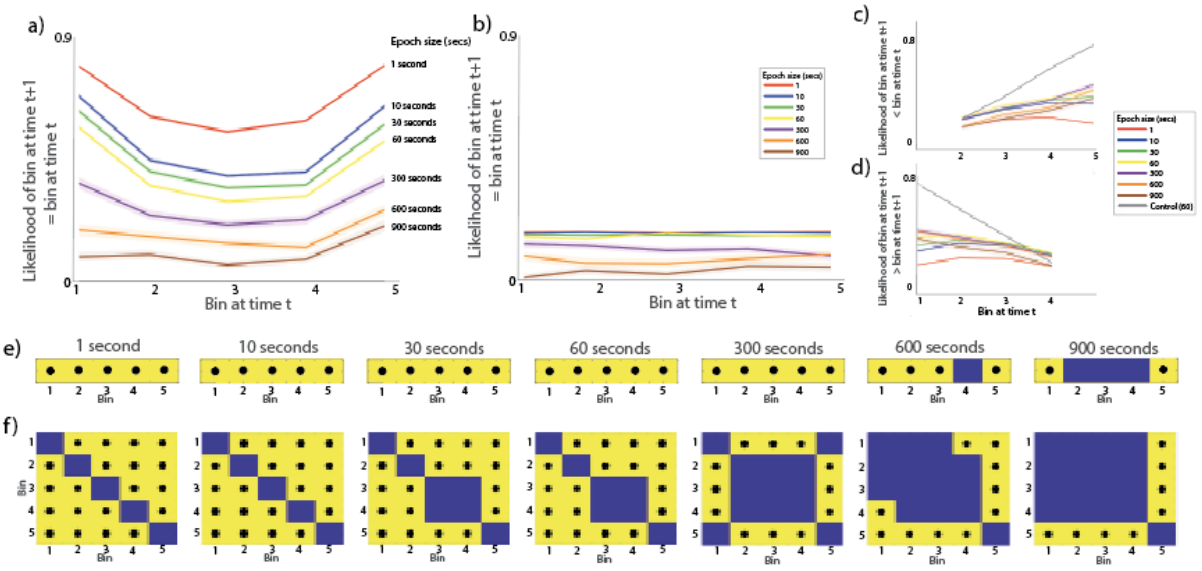

Figure 4: a) Results of analysis 1, examining how stability in arousal varies as function of arousal. Analysis was conducted as described in the Methods. Different coloured lines show the different epoch durations used. Shaded areas show the Standard Error of the Means (SEM). b) Results of identical analysis to a, but conducted on control data. See Methods for description of how this was generated. c) and d): Results of analysis 1, showing how the likelihood of a decrease (c) or an increase (d) in arousal varies as a function of arousal. Different coloured lines show the different epoch durations used. Grey line shows the control data; for clarity, only one epoch duration for the control data is shown (60 seconds). Shaded areas show the SEM. e) Results of analysis to compare the observed values within each bin to chance. f) Results of pairwise significance analyses to compare the observed results between bins. For both plots, yellow squares indicate a significant difference between bins, blue indicate no significant difference.

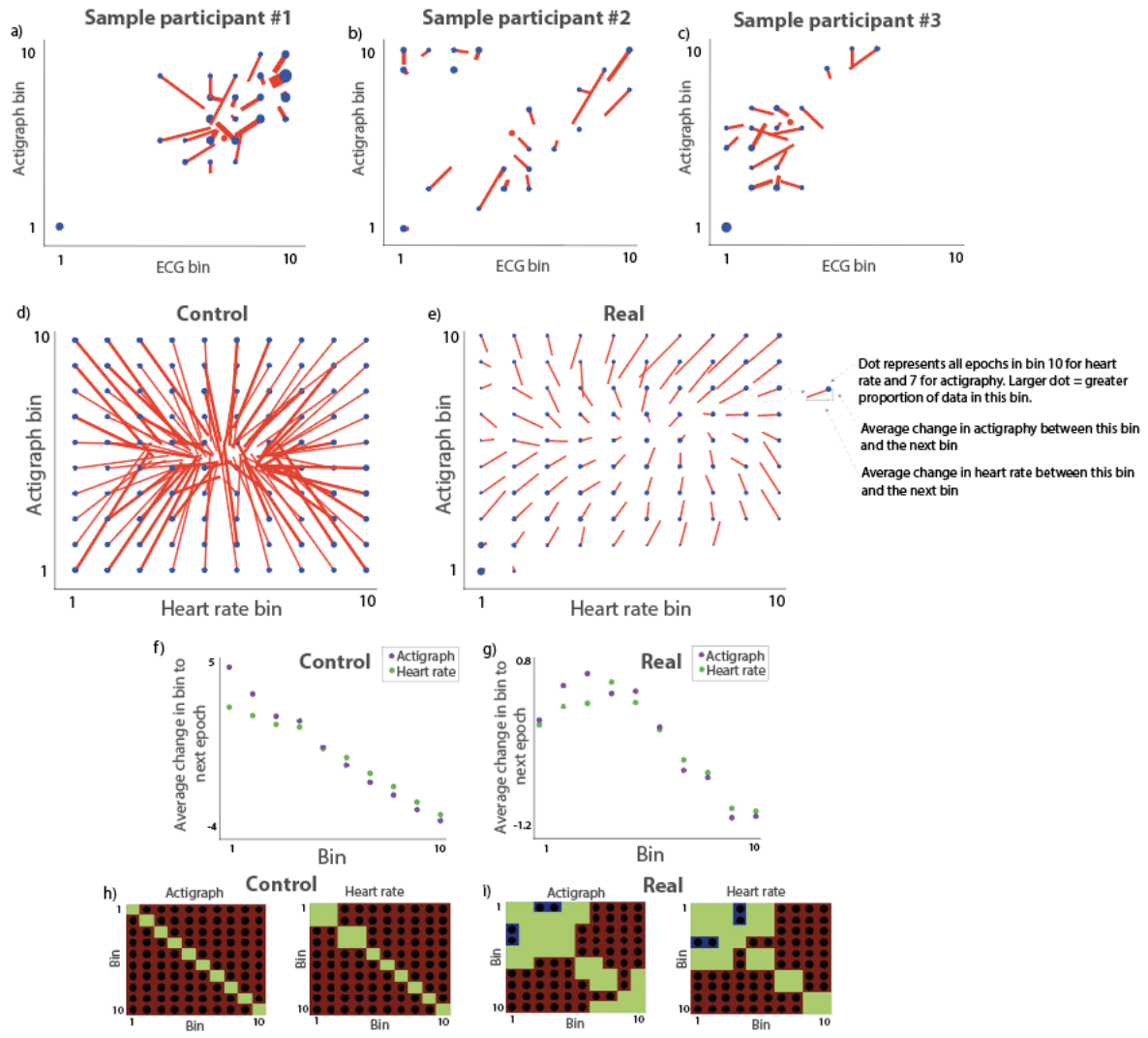

Figure 5: Vector plots conducted for analysis 2. See detailed description in the Methods of how these were calculated. a) to c): sample dataplots for three individual participants. d) Vector plot for entire control dataset. e) Vector plot for the entire real dataset. f) and g): Scatterplots showing how arousal change (equivalent to the length of the vectors in the vector plots) varies contingent on arousal. Purple dots show change in actigraphy (the x-dimension of the vectors on plots d and e). Green dots show change in heart rate (the y-dimension of the vectors on plots d) and e)). h) and i): Results of pairwise significance analyses to assess how arousal change (as shown in figures f and g) differs between bins. Green indicates no difference; red and blue indicate significant differences (blue – x axis bin < y axis bin; red – x axis bin > y axis bin).