Biological Psychology 165 (2021) 108174

Contents lists available at ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

On the physiology of interruption after unexpectedness

Marret K. Noordewier [a,*], Daan T. Scheepers [a,b], John F. Stins [c], Muriel A. Hagenaars [d]

[a] Social, Economic and Organisational Psychology, Social and Behavioural Sciences, Leiden University, Leiden, the Netherlands
[b] Social, Health and Organizational Psychology, Social and Behavioural Sciences, Utrecht University, Utrecht, the Netherlands
[c] Human Movement Sciences, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
[d] Clinical Psychology, Social and Behavioural Sciences, Psychology, Utrecht University, Utrecht, the Netherlands

ARTICLE INFO

Keywords:
Surprise
Unexpectedness
Orienting
Freezing
Physiology
Heart rate
Body sway
Blood pressure
Temperature

ABSTRACT

We tested whether surprise elicits similar physiological changes as those associated with orienting and freezing after threat, as surprise also involves a state of interruption and attention for effective action. Moreover, because surprise is primarily driven by the unexpectedness of an event, initial physiological responses were predicted to be similar for positive, neutral, and negative surprises. Results of repetition-change studies (4 + 1 in Supplemental Materials) showed that surprise lowers heart rate (Experiments 1-4) and increases blood pressure (Experiment 4). No effects on body movement (Experiment 2) or finger temperature (Experiment 4) were found. When unexpected stimuli were presented more often (making them less surprising) heart rate returned to baseline, while blood pressure remained high (Experiment 4). These effects were not influenced by stimulus valence. However, second-to-second analyses within the first (surprising) block showed a tendency for a stronger increase in systolic blood pressure after negative vs. positive surprise.

1. Introduction

People constantly make predictions about the world around them, to be able to respond efficiently and effectively to relevant stimuli in their environment (Clark, 2016; Miceli & Castelfranchi, 2015). It is, however, not possible to have a perfect predictive mental model of the environment and people are regularly faced with unexpected events. Unexpectedness results in surprise—a feeling associated with the interruption of ongoing thoughts and activities and attention to the surprising stimulus to make sense of it (Horstman, 2006, 2015; Meyer, Reisenzein, & Schützwohl, 1997; Noordewier, Topolinski, & Van Dijk, 2016; Noordewier & Van Dijk, 2019; Reisenzein, Meyer, & Niepel, 2012; Reisenzein, Horstmann, & Schützwohl, 2017). Surprise is thus the initial interrupted state after an unexpected event and only after people have made sense of the unexpected, other affective states may follow depending on the valence of the event. This means that it takes some time to experience, for example, joy after a positive surprise or sadness after a negative surprise (Noordewier et al., 2016).

The goal of the current research is to map the physiological correlates of surprise to better understand the initial interruption after unexpectedness. Following literature on orienting and freezing, we tested whether surprise results in a reduction of heart rate and movement. This would fit with the characterization of surprise as interruption for attention and effective action. We also differentiated between positive and negative surprises, to test to what extent the physiology of surprise is influenced by the valence of the stimulus.

1.1. The physiology of interruption

A stimulus that violates people's predictive mental model of the world is registered as a change to the anticipated flow of stimuli through novelty detection (Brosch, 2009). This novelty check is, according to the Components Process Model of Emotion, the first appraisal in the relevance check of events, and it includes evaluations of suddenness, familiarity, and predictability (Ellsworth & Scherer, 2003; Scherer, 2001; Soriano, Fontaine, & Scherer, 2015). Surprise is thus a sign that one has failed to anticipate an event (Huron, 2006; Miceli & Castelfranchi, 2015). Because people are, by definition, unprepared for the unexpected event to occur, surprise involves a transient state during which it is unclear what has happened and whether immediate action is required (Ekman, 2003; Noordewier & Breugelmans, 2013; Scherer, Zentner, & Stern, 2004). This interferes with the need to be able to predict, prepare, and understand outcomes (Miceli & Castelfranchi, 2015; Noordewier et al., 2016; Proulx, Sleegers, & Tritt, 2017; Topolinski & Strack, 2015;

* Corresponding author.
E-mail address: m.k.noordewier@fsw.leidenuniv.nl (M.K. Noordewier).

https://doi.org/10.1016/j.biopsycho.2021.108174
Received 12 April 2021; Received in revised form 7 July 2021; Accepted 20 August 2021
Available online 25 August 2021

see also Abelson et al., 1968; Gawronski & Strack, 2012; Harmon-Jones & Mills, 1999; Harmon-Jones, Amodio, & Harmon-Jones, 2009; Proulx, Inzlicht, & Harmon-Jones, 2012). This makes surprise an "undecided state" in which it is not yet clear whether one should approach, avoid, or ignore the surprising stimulus (Scherer et al., 2004). To move beyond this indeterminate situation as soon as possible, surprise facilitates sense-making by a shift in attention to the surprising stimulus, accompanied by an interruption of ongoing thoughts and activities (Horstmann, 2015; Reisenzein et al., 2017). This interruption is particularly pronounced when the surprise is part of goal-directed behavior (Meyer et al., 1997)—even though people will also detect surprises that are not specifically relevant to a task, as it is key to identify changes to one's environment to be able to respond effectively (e.g., Harmon-Jones et al., 2009).

To assess which physiological markers are associated with this state of interruption and attention, a first relevant connection is the orienting response. Orienting is a "what-is-it" response (Lang, Simons, & Balaban, 1997; Pavlov, 1927) that occurs in response to novel stimuli (Packer, Siddle, & Tipp, 1989; Parmentier, Vasilev, & Andrés, 2018). A key physiological correlate of the orienting response is a reduction of heart rate (bradycardia; Cook & Turpin, 1997). This is assumed to enhance the sensitivity to input from the environment (Campbell, Wood, & McBride, 1997; Graham & Clifton, 1966; Parmentier et al., 2018), to facilitate perception and preparation for (defensive) action (Lang, Bradley, & Cuthbert, 1997). Orienting has also been connected to motor inhibition (e.g., slowing of responses or action stopping); an inhibitory response that is often related to increased attention and cognitive control or processing of the event (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Notebaert et al., 2009; Parmentier et al., 2018; Wessel, 2017).

Orienting may be followed by "freezing"—a defensive response that occurs under (the expectancy of) threat (Hagenaars, Oitzl, & Roelofs, 2014). Freezing is also characterized by a reduction of heart rate and reduction of movement, and in a laboratory setting it can be observed when watching aversive pictures/films (Azevedo et al., 2005; Hagenaars, Stins, & Roelofs, 2012; Hagenaars, Roelofs, & Stins, 2014) or during social threat (Roelofs, Hagenaars, & Stins, 2010; see also Noordewier, Scheepers, & Hilbert, 2020). Similar to orienting, freezing is also thought to facilitate perception and action preparation (Roelofs, 2017). It for instance relates to increased perception of coarse information (Lojowska, Gladwin, Hermans, & Roelofs, 2015) and freezing responses were shown to be stronger when participants had the opportunity to act in the threatening situation (vs. when they did not; Gladwin, Hashemi, van Ast, & Roelofs, 2016; see also Hashemi et al., 2019). While freezing thus also results in a reduction of heart rate and movement, it is assumed to be of longer duration, stronger, and more specific to (possible) threat than orienting (Hagenaars, Oitzl et al., 2014; Hagenaars, Roelofs et al., 2014; Roelofs, 2017).

Both orienting and freezing can be conceptualized as a form of inhibition ("a break on the system") as well as an active anticipatory state relevant for perception and action preparation (see Gladwin et al., 2016; Hashemi et al., 2019; Lang, Bradley et al., 1997; Lang, Simmons et al., 1997; Parmentier et al., 2018; Roelofs, 2017; Walker & Carrive, 2003). As such, it involves activation of both the sympathetic and parasympathetic nervous system (Hagenaars, Oitzl et al., 2014; Hagenaars, Roelofs et al., 2014; Roelofs, 2017). Importantly, these processes relate to the key hallmark of surprise; the interruption of ongoing thoughts and activities to attend to the unexpected stimulus, to facilitate sense-making and preparation for effective action. Moreover, this interruption after unexpectedness may even have threatening aspects, because as long as one does not fully understand what has happened, the unexpectedness violates consistency and meaning maintenance motives (Abelson et al., 1968; Gawronski & Strack, 2012; Harmon-Jones & Mills, 1999; Harmon-Jones et al., 2009; Miceli & Castelfranchi, 2015; Noordewier et al., 2016; Proulx et al., 2012, 2017; Topolinski & Strack, 2015).

Based on this, we reasoned that surprise is likely to share physiological and behavioral features with orienting and freezing—i.e., primarily a reduced heart rate and reduced movement. Moreover, given that it is primarily the unexpectedness that drives surprise irrespective of the valence of the surprising outcome, these responses should occur for all types of surprises, including positive ones.

1.2. Preliminary evidence

Unexpectedness is indeed assumed to trigger an orienting response (e.g., Reisenzein et al., 2012, 2017), but only limited work has tested the link between surprise and reduced heart rate directly (see Niepel, 2001). Some studies have tested responses to rare events (i.e., infrequent stimuli, such as in oddball studies), but it is important to note that rarity is not the same as surprise (for a more elaborate discussion on this, see Horstmann, 2015). That is, people often predict the occurrence of rare events and rarity has different effects on attention and distraction during task-performance than surprise has (e.g., surprise is always distracting, while for rarity this depends on the context; Horstmann & Ansorge, 2006). Therefore, in the current studies, we systematically tested heart rate reductions after one critical surprise trial, i.e., the presentation of one deviant stimulus, rather than multiple deviant stimuli.

Regarding movement, there is some evidence for a connection between surprise and reduced movement from a study on expectancy violation in infants, where Scherer and colleagues (2004) argued that behavioral freezing is a natural response to unexpectedness. Freezing allows "to await further information, or hope for the danger situation to clear, or to discover that there was a false alarm." (p. 399). Supporting this, studies showed that infants respond with facial and behavioral stilling in response to a so-called "impossible event" (i.e., a toy switch or a sudden change in the voice of the experimenter; Camras et al., 2002; Scherer et al., 2004). Infants thus show reduced movement after (neutral) unexpectedness, presumably as a correlate of an attempt to make sense of the new situation. In addition, studies with adults support the link between deviance and motor inhibition. For example, when participants were presented with deviant task-irrelevant sounds (i.e., sounds that occurred in 20–25 % of the cases), they responded slower in an oddball task (adding extra delay to post-error slowing; Parmentier et al., 2018; see also Meyer et al., 1997), they show stronger inhibition when asked to withhold a response in a Go/NoGo-task (Wessel, 2017; see also Leiva, Andrés, & Parmentier, 2015; Schröger, 1996), and they fixate longer on a word when reading (Vasilev, Parmentier, Angele, & Kirkby, 2019). A recent study including TMS and EEG confirmed that such deviant stimuli facilitate action stopping and activate brain regions for motor inhibition (i.e., Dutra, Waller, & Wessel, 2018).

1.3. Hypotheses

Taken together, we predicted that surprise would share physiological and behavioral features with orienting and freezing. Following this, we focused on a well-established physiological indicator in all studies, namely a lowered heart rate. Moreover, we incorporated other indicators of orienting and/or freezing in individual studies. Specifically, in Experiment 2, we focused on a reduction of movement and in Experiment 4, we focused on finger temperature and blood pressure. Temperature and blood pressure have received less attention in research on orienting and freezing, but they are relevant to include because a lowered heart rate *in combination with* lowered temperature and increased blood pressure would be suggestive of an inhibitory but active

state (Reyes del Paso, Vila, & García, 1994; Sawada, 2003; see also Carrive, 2000; Hagenaars, Oitzl et al., 2014; Hagenaars, Roelofs et al., 2014; Roelofs, 2017; Vianna & Carrive, 2005). This would fit with the conceptualization of surprise as a state of interruption for effective action. We elaborate on this further in the introduction of Experiment 4.

Importantly, the initial physiological responses after a surprise are predicted to be similar for positive, negative, or neutral stimuli, as it is the *unexpectedness* that primarily drives initial responses rather than the *valence* of the stimulus (Noordewier & Van Dijk, 2019; Noordewier et al., 2016). This perspective on surprise fits a sequential appraisal perspective on emotion, where novelty evaluations occur before pleasantness evaluation (i.e., whether a stimulus results in pleasure or pain; Delplanque et al., 2009; Ellsworth & Scherer, 2003). Moreover, previous research using facial expression coding showed that after a positive or a negative surprise, expressions were initially similar, and only after some time the expressions started to differentiate depending on the valence of the event (e.g., more positive when the surprise was positive; Noordewier & Van Dijk, 2019; see also Noordewier & Breugelmans, 2013). If initial behavioral expressions to surprise are indeed primarily driven by the unexpectedness of the stimulus, then it is likely that initial physiological responses after surprise are also similar for different types of surprises.

Thus, rather than finding variation in physiology for positive or negative *stimuli* (e.g., Bradley, Codispoti, Cuthbert, & Lang, 2001; Bradley, Lang, & Cuthbert, 1993; Hagenaars, Oitzl et al., 2014; Hagenaars, Roelofs et al., 2014), we predict that different types of *surprises* should initially (i.e., in the first seconds) result in similar physiological responses. This prediction in seconds (compared to, for example, milliseconds) is based on the temporal pattern found in previous facial expression research (Noordewier & Van Dijk, 2019) and on the logic that surprise takes time to dissipate because it includes not only the perception/detection of the unexpectedness to dissipate, but also the interruption of processing, reallocation of processing resources, and analyses and evaluation of the event, before someone can fully understand what has happened and evaluate it (e.g., Reisenzein et al., 2012). When, however, a stimulus is not surprising anymore, responses are predicted to be influenced by the *valence* of the stimulus (Noordewier et al., 2016): For negative stimuli, a sustained freezing-like response is likely the result of the threatening aspects of the stimuli, while for positive stimuli, responses will likely return to baseline because of habituation to their (non-threatening) occurrence. In sum, we predict that all surprises result in responses similar to orienting and freezing. Yet, when a stimulus is repeatedly presented (and, as a result, not surprising anymore) an effect of valence is predicted to occur: Responses to positive stimuli will return to baseline, while responses to negative stimuli will not.

1.4. The current studies

We conducted four experiments to examine the physiological response profile of reactions to unexpectedness. Furthermore, to test whether the pattern of physiological responses would be similar or different for different types of surprises, we used relatively neutral stimuli in Experiments 1 and 2, positive stimuli in Experiment 3, and we directly compared positive vs. negative stimuli in Experiment 4.

All experiments involved 5-minute baseline period during which we recorded the physiological responses while the participant was in a relative state of rest. Moreover, all experiments employed a repetition-change paradigm; a standardized and well-validated method to induce surprise (e.g., Camras et al., 2002; Meyer et al., 1991; Niepel, 2001; Noordewier & Van Dijk, 2019; Reisenzein, Bördgen, Holtbernd, & Matz, 2006; Schützwohl, 1998). In this paradigm, participants are repeatedly exposed to a series of comparable stimuli, which creates an expectancy about what will follow. After a series of trials, this is changed by presenting a completely different stimulus, which is surprising because it does not fit the expectancy that was induced in the repetition phase. To strengthen the engagement in the task, we asked participants to respond to the repetition stimuli (e.g., name the color). Moreover, to strengthen the expectancy about the content of the repetition trials, each trial was programmed such that stimulus presentation durations and inter-display intervals were fixed at 1 and 0.5 s, respectively. This creates a fixed tempo for the stimulus presentation (see also Noordewier & Van Dijk, 2019).

In all studies, we focused on (reduced) heart rate. In Experiment 2, we also measured body movement using a balance board (e.g., Roelofs et al., 2010). In Experiment 4, we included temperature and blood pressure measurements. This final study also tested the time course of responses to positive and negative stimuli, by presenting the stimuli multiple times, allowing us to compare the first surprising block (i.e., testing the response to the *unexpectedness*) to the non/less surprising block later (i.e., testing the response to the *valence* of the stimulus). Moreover, we measured the subjective evaluation of surprising stimuli in various ways. In Experiments 1, 3, and 4, we incorporated a surprise and valence-check (which for Experiment 4 also separates between initial and general feelings). In Experiment 2, we included more elaborated self-reports. In addition to gaining more insight into the contents of their emotional experience, we wanted to check whether and how participants would differentiate between initial and later responses (using an open question format) and whether and how they would differentiate their evaluations of the unexpectedness vs. the valence of the stimulus.

Participants received course credits or a monetary reward for their participation, except in Experiment 2, where all participants received a monetary reward. Studies were performed in Dutch and the items described below are translations from the original Dutch texts. All studies were approved by the Psychology Research Ethics Committee at Leiden University (The Netherlands).

1.4.1. Data analyses and reporting

We report all manipulations, all measures, and all data exclusions. For ease of presentation, we include some analyses and one additional Experiment (i.e., Experiment 2b; a replication and extension of Experiment 2) in a Supplement, which is available on OSF via https://osf.io/9rvgu/. Data and materials of all studies are available on https://dataverse.nl/dataverse/leidensocial.

Sample sizes are at least 40 per cell (Simmons, Nelson, & Simonsohn, 2013), yet, when possible, we collected more data to be able to account for data exclusion as a result of unexpected measurement issues, such as recording errors. We report sensitivity power analyses for each study, which were calculated using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). We based the sensitivity analysis on repeated measures MANOVA's with $\alpha = .05$ and 80 % power and the sample size after exclusion as input (where relevant). In addition, to match the effect-size produced by G*Power to those produced by SPSS, we un-checked the mean-correlation in the calculation and used the Muller and Peterson algorithm (1984; see Faul et al., 2007; Lakens & Caldwell, 2021). Moreover, we report the η_p^2, calculated from the *f*-value as provided by G*Power (see Lakens & Caldwell, 2021).

Data processing and analyses was the same for all studies. The ECG was recorded using Acqknowledge software and scored in Matlab using the "Physiodata toolbox" (Sjak-Shie, 2018; see https://physiodatatoolbox.leidenuniv.nl; processing details for movement and blood pressure are described in Experiments 2 and 4, respectively). In the ECG, R-peaks were automatically detected and then manually checked. Mean heart

rate was then calculated for the last minute of the baseline and for the 10 or 15 s (depending on the study) during which the surprising stimulus was shown (surprise block).

In addition to comparing the surprise block to baseline levels, we also examined the temporal unfolding of heart rate changes after the surprise trial by means of "second-to-second" analyses. For these analyses we compared the heart rate one second before the surprise stimulus to the heart rate of each second during the 10/15-second surprise block. The reasons for these, more fine-grained, analyses were threefold (see also Noordewier et al., 2019): First, it can potentially rule out the possibility that decreased heart rate is (partly) explained by participants becoming simply more relaxed during the course of the experiment and that this would explain differences between the surprise block and baseline. Second, it can rule out any effect of passive viewing during baseline vs. the more active responses during the task, as the second just before the surprise is part of the more active repetition-task. Finally, the second-to-second analyses can inform us about how long a possible heart rate change after a surprising stimulus persists (e.g., whether heart rate would be consistently lower during the surprise block or only during the first few seconds).

We tested our hypothesis regarding *differences* (e.g., lower heart rate after surprise) with repeated measures MANOVAs. We first report the multivariate test-output, before reporting simple contrasts—where relevant. Multivariate statistics are suitable for designs with one within-subject factor and they have the advantage that they have more power than (corrected) univariate statistics, as they are not affected by sphericity violations (Algina & Keselman, 1997). Next, we tested our hypothesis regarding *similarity* with equivalence testing (e.g., similar reduction of heart rate for positive and negative surprise; Experiment 4). More details on equivalence testing are provided in Experiment 4.

In all studies, we defined statistical outliers as values more than 3.3 standard deviations above or below the mean (i.e., the 1% cut-off point in a normal distribution, see e.g., Seery, Leo, Lupien, Kondrak, & Almonte, 2013), which were calculated with Z-scores per participant. Outliers were Winsorized by recoding them into the first non-extreme value + 1% (e.g., Seery et al., 2013). This procedure reduces the impact of extreme values, while keeping the participant included in the sample and preserving the rank-order of the distribution (see Seery et al., 2013). In each study, we report whether and how many values were corrected; we also report when results were different without winsorizing (which only occurred on one of the BP measures in Experiment 4). In addition, we present confidence intervals for effect sizes, which are 95 % for Cohen's *d* and 90 % for η_p^2 (Steiger, 2004).

2. Experiment 1

In the first experiment we tested whether a surprise would result in a lower heart rate. As in the other Experiments we compared heart rate following surprise to baseline levels and we also examined the temporal unfolding of the heart rate changes right after the surprise using 1-second epochs.

2.1. Method

A total of 59 participants were recruited at the Social Science Faculty of Leiden University (48 females, 11 males; $M_{age} = 21.41$, $SD_{age} = 2.63$) to participate in a within-subjects experiment, in which we compared heart rate between two blocks: baseline vs. surprise. Data of five participants were excluded from analyses: For one participant the event marker was not recorded, for one participant the door of the cubicle opened during the study, and three participants had participated in a similar study before (i.e., Experiment 3, which was executed before this experiment). We report analyses on the remaining 54 participants (44 females, 10 males; $M_{age} = 21.46$, $SD_{age} = 2.70$). The sensitivity power analysis showed that the minimum effect size to consider the observed effect as relevant is $\eta_p^2 = 0.13$ in the block-comparison and $\eta_p^2 = 0.31$ in the 1-second epoch comparisons.

2.1.1. Materials and procedure

After signing the informed consent, the experimenter connected the participant to a Biopac MP150 system comprising an ECG100c module. Electrodes were placed on the wrists and ankles following a Lead I configuration.[1] Participants wearing shoes that covered their ankles were asked to take them off, to enable electrode placement. ECG signals were measured at a 1000 Hz sampling rate; a 35 Hz low-pass filter was applied online, and a 1 Hz high-pass filter was applied offline.

Participants were led to an individual cubicle and were asked to stand upright during the study, to match the conditions in other freezing studies (e.g., Roelofs et al., 2010) as well as in Experiment 2 where we used a balance board. The computer monitor was placed on eye-level and participants were asked to stand on marks on the floor to standardize their position. The study was presented as a study on attention, perception, and physiology. We explained that we would first conduct a baseline measure, for which participants were asked to watch a 5-minute video showing underwater scenes. The heart rate during the last minute of the video served as our baseline measure.

After the baseline phase, the actual experiment started. Participants viewed geometrical shapes in different colors. After each set of four geometrical shapes participants were asked to verbally report the color of the last shape. We asked them to verbally report to minimize movement. We recorded the verbal responses using the audio function of a webcam. No video recordings were made, which was also mentioned explicitly to the participants. Each geometrical shape (triangle, circle, square, or rectangle in green, grey, blue, or red) was presented for 1 s, followed by a 0.5 s inter-display interval. Then, the question to indicate the color of the last shape was presented for 1.5 s. After this, the study continued to the next trial. Each trial thus consisted of the stimuli "shape"-"shape"-"shape"-"shape"-"question", which were automatically presented in a fixed timing (for a comparable procedure, see Noordewier & Van Dijk, 2019).

Participants first completed two practice trials, to familiarize themselves with the procedure. Then, the main task followed, which consisted of 16 standard trials and one (critical) surprise trial (for an illustration, see Fig. 1). During the surprise trial the question about the color of the last shape was replaced with a picture of George Clooney. The picture was a Gif-file (i.e., a graphic interchange format picture, containing multiple image-frames in sequence), showing a still of Clooney with some movement in the background, taken from the movie "Up in the Air".[2] The picture was shown for 10 s after which the study continued to background questions.

As background questions, we asked participants to rate the intensity of their surprise using the item "To what extent were you surprised by the last picture?" and a response scale ranging from 1 = *not at all* to 7 = *extremely*. Then, we asked participants to rate the valence of the surprising target using the item "How do you evaluate the last picture?" and a response scale ranging from 1 = *negative* to 7 = *positive*. Next, we checked the familiarity of the target person with the question "Did you recognize the man in the picture as George Clooney?" and the answering option yes/no. In this experiment as well as Experiments 2 and 3, we also included the Ten Item Personality Inventory (TIPI; Gosling, Rentfrow, & Swann, 2003) but we do not report this data in this paper. Finally, we asked for age and gender and checked whether participants had participated before in a similar study (yes/no).

[1] This experiment and Experiment 3 also included a pulse measure. For ease of presentation, results will not be discussed.

[2] Surprising target of Experiment 1: https://images.app.goo.gl/5gHhX2W9cV8o7U3DA.

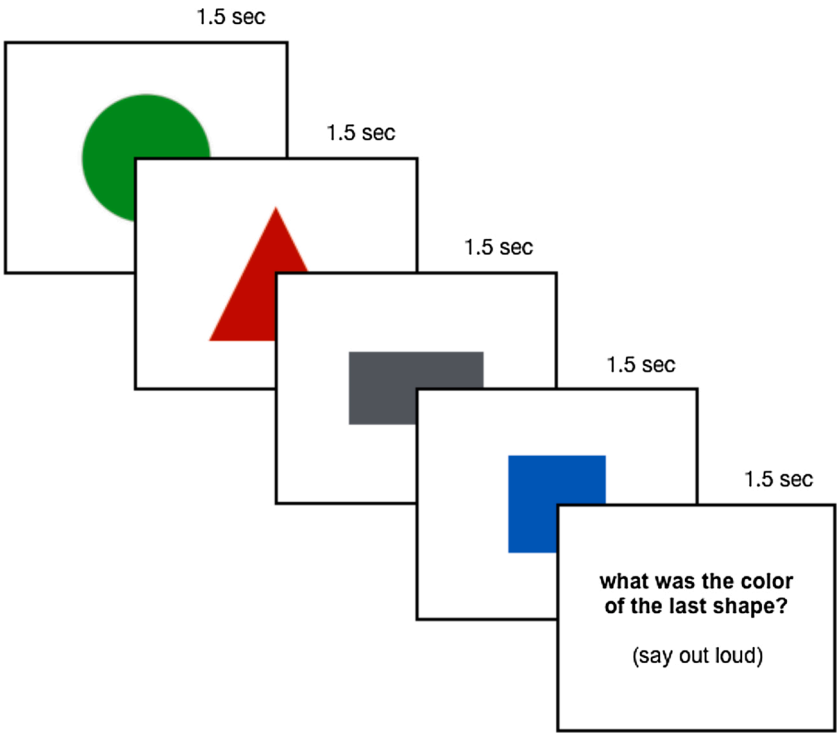

Fig. 1. Illustration of repetition-change trial as used in Experiments 1, 2, and 4. Each trial presents four shapes in different colors (randomized). During the repetition phase, participants report the color of the fourth shape. The critical trial presents a surprising stimulus instead of a question.

2.2. Results and discussion

2.2.1. Surprise and valence check

First, we examined the surprise and valence ratings. Participants rated the intensity of their surprise as $M = 5.72$ ($SD = 1.24$) and the valence of the stimulus as $M = 3.61$ ($SD = 1.09$). This shows that we successfully created a surprising event with a stimulus that was evaluated as relatively neutral (i.e., just below the midpoint of the scale). A total of six participants indicated not to recognize George Clooney as such. Excluding the data of these participants did not change the pattern of results, and therefore they were included in the analyses.

2.2.2. Heart rate (HR)

There were no statistical outliers in the current dataset. To test the prediction that HR would be lower during the surprise block than during baseline we used a repeated measures MANOVA. Results showed that the HR was lower during the surprise block ($M = 86.82$, $SD = 13.05$) than during baseline ($M = 93.68$, $SD = 14.02$), Wilks' Lambda = .43, $F(1,53) = 70.14$, $p < .001$, $\eta_p^2 = .57$ [.41, .67].

We further examined the temporal unfolding of HR changes after surprise, using the second-to-second approach described above using a repeated measures MANOVA with Time as a within-participant factor with 11 levels (HR 1 s before the surprise stimulus vs. HR during each of the subsequent 10 s after surprise). This showed an effect of Time (see Fig. 2a), Wilks' Lambda = .42, $F(10,44) = 5.99$, $p < .001$, $\eta_p^2 = .58$ [.29, .61]. Simple contrasts with HR during the second before the surprise as reference showed a pattern of HR deceleration: HR in the first second after the surprise did not differ, $F(1,53) = 0.01$, $p = .926$, $\eta_p^2 = .00$, while for all subsequent seconds HR was consistency lower, $Fs > 10.55$, $ps < .001$, η_p^2s > .16.

Note that this and subsequent second-to-second comparisons did not include a correction for multiple comparisons: Our main aim was to test the overall pattern of heart rate reduction, rather than to test the differences during specific seconds per se. We thus mostly rely on the overall effect of time, and subsequently explore how long possible effects remain. Moreover, we included these analyses in all studies, to test the robustness of this pattern.

Together, these results show that after surprise heart rate decelerates. This effect was not only found when comparing heart rate to baseline, but also when comparing heart rate relative to the last second before the surprise occurred. This latter analysis further showed that the effect not only occurred during the initial seconds after surprise, but during the entire surprise block. Moreover, by using the second just before the surprise as a comparison, these analyses rule out that the heart rate effects in the block-comparison were explained by a HR reduction over the course of the experiment.[3]

It is important to note, however, that the surprise stimulus was a headshot where Clooney was facing the camera. One possibility is that a person directly looking at you is (somewhat) socially threatening, especially when this picture is presented unexpectedly. This could have intensified the heart rate deceleration (see Noordewier et al., 2020; Roelofs et al., 2010). It should be noted however, that such a potential social threat effect was not so strong that it influenced the general evaluation of the stimulus, which was quite neutral. However, to exclude this social threat explanation more directly, we used a different picture of Clooney in Experiment 2. Moreover, in the next study we also

[3] To check whether the order of the different elements in our studies affected HR reduction, we ran extra analyses for all studies, where we compared HR during the repetition-trials to HR during the baseline and the surprise block (Supplemental Materials; Figures S1a-e and Table S1). Compared to baseline, HR increased in the first repetition-trial(s); and HR in the repetition-trial before the surprise was similar to (Experiments 1-3), or higher than (Experiments 2b and 4) baseline. Importantly, compared to the surprise block, HR is consistently higher during the repetition-trials—either for all repetition-trials (Experiments 1, 2b, and 4) or the preceding 8 or 11 repetition-trials before the surprise (Experiments 2 and 3). These results thus rule out that HR simply became lower throughout the course of the experiment, because surprise lowers HR compared to the repetition-trials that preceded it and these trials are either similar to baseline (Experiments 1-3) or even higher (Experiments 2b, 4).

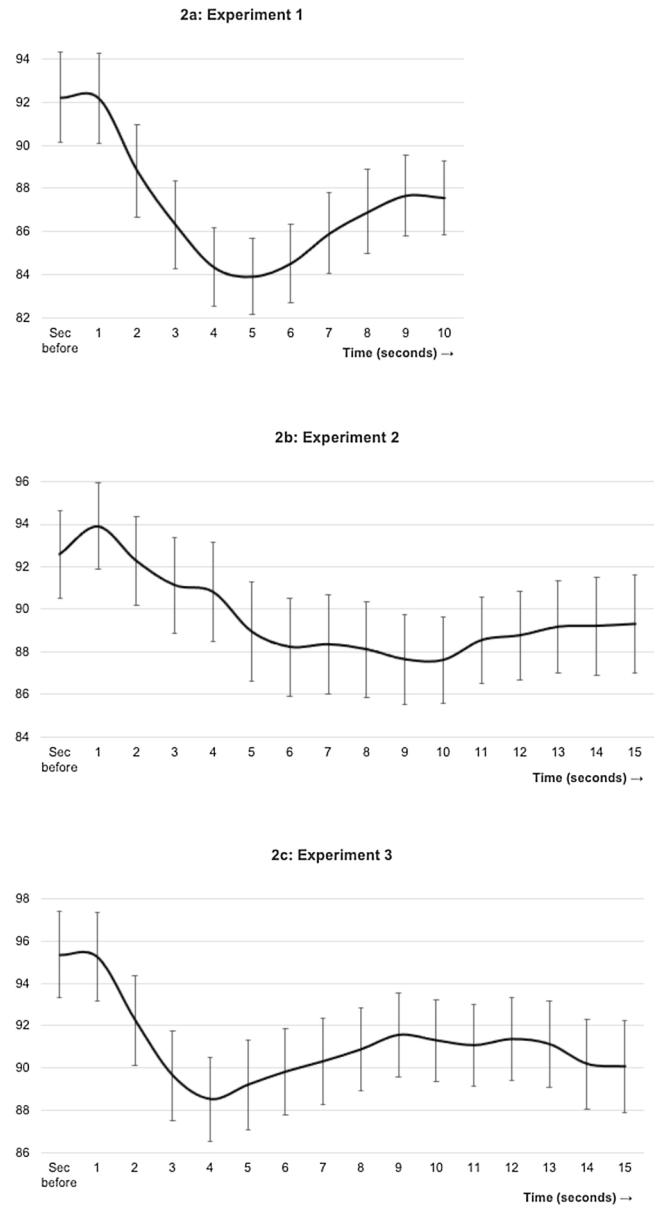

Fig. 2. Heart rate change after surprise. Experiment 1: Clooney facing the camera (2a). Experiment 2: Clooney facing sideways (2b). Experiment 3: Puppy (i.e., positive surprise; 2c). Error bars indicate +/− 1 SE.

included further self-report measures to differentiate the *evaluation* of Clooney from the *unexpectedness of the picture* of Clooney. Another aim with Experiment 2 was to examine behavioral freezing-like responses after unexpectedness by including a measure of body sway.

3. Experiment 2

In Experiment 2, we tested whether participants would show both reductions in heart rate and spontaneous body sway after a surprise. In addition, we extended the self-report measures to be able to differentiate between evaluations of the *unexpectedness* of the appearance of Clooney and the *valence* of Clooney himself.

3.1. Method

A total of 66 participants (28 females, 36 males, 1 other, 1 missing; $M_{age} = 22.12$, $SD_{age} = 4.92$) were recruited at the Faculty of Behavioural and Movement Sciences of the Vrije Universiteit Amsterdam[4] to participate in an experiment with the same within-subjects design and the same general procedure as Experiment 1, except for that we also measured body sway using a balance board (e.g., Roelofs et al., 2010). The educational background of participants was: 40.0 % movement science students, 21.5 % medicine students, 10.8 % psychology students, and 27.7 % other (e.g., law or international business). The data of one participant were excluded from the analyses because this person indicated to be colorblind and therefore could not name the colors of the shapes in the task. For two participants, the heart rate data was not recorded; for one participant the heart rate event marker was not recorded; for one participant, the background questions were not saved due to a software error. We included the remaining available data of these three last participants in the analyses. The sensitivity power analysis (with $N = 62$ for HR and $N = 45$ for body sway) showed that the minimum effect size to consider the observed effect as relevant is $\eta_p^2 = .11$ for the block-comparison of HR and $\eta_p^2 = .33$, for 1-second epoch comparisons of HR. For the body sway the value is $\eta_p^2 = .27$.

3.1.1. Procedure and materials

Each participant was welcomed in a room with a custom-made balance board (1 × 1 m) in front of a TV-screen and a separate table with two computers (one for recording the heart rate data and for the participant to complete the questionnaires, and another one for recording movement data). After signing the informed consent, participants were connected to the ECG equipment. ECG was recorded continuously using a Biopac MP150 system, and a Bionomadix BN-RSPEC module. Two electrodes were attached right below the collarbones (with the electrode on the right side—from the experimenter's perspective—as the reference electrode) and one to the right lower rib. Note that—unlike Experiment 1—we used a wearable ECG module (the Biopac Bionomadix system), that transmits the signals via Bluetooth to the Biopac MP-system. This adjusted procedure was used to optimize body sway assessment. ECG signals were measured at a 1000 Hz sampling rate; a 150 Hz low-pass filter and a 1 Hz high-pass filter were applied offline.

Participants were asked to stand on a balance board in front of a 55-inch monitor. The balance board registers the center of pressure (COP), i.e., the point of application of the ground reaction force. During quiet upright standing there will always be a certain amount of spontaneous body sway in the left-right and fore-aft direction. This is reflected in the time-series of the COP, which fluctuates in the anterior-posterior (AP) direction and the mediolateral (ML) direction (see e.g., Pollock, Durward, Rowe, & Paul., 2000 for a biomechanical introduction to the concept of human balance). In the case of freezing, these fluctuations will be somewhat reduced, yielding lower values of the COP excursions in either direction. This equipment has repeatedly demonstrated its utility in detecting reductions in postural movements as part of a freezing response (e.g., Roelofs et al., 2010; Stins, Roerdink, & Beek, 2011). The COP was recorded at 100 Hz. As our outcome measure, we took the standard deviation of the time series (in both axes) over 1-sec bins (thus consisting of 100 samples). The *SD* is a relatively straightforward measure of postural activity; the lower the *SD*, the greater the postural immobility (note: a dead weight atop the balance board will have zero movement and thus an *SD* of 0).

[4] The data were collected in two waves. After the initial data collection ($N = 44$), we discovered an error in the balance board recordings, such that the movement data of only the first 25 participants were recorded. To make sure we had sufficient power to test our hypothesis, we ran an additional 21 participants in a second wave. This was done with the same equipment, but in a different lab. In the first wave, the experimenter was sitting behind a screen during data collection (without participant contact). In the second wave, the lab was smaller and the experimenter left the room at the start of the study.

Because participants could not use the computer keyboard or mouse to navigate through the study, we programmed the study such that, in addition to the experimental task, the instructions also proceeded automatically. We made sure that the various instructions could be read easily by people at different reading speeds. After instructions, the baseline phase, and the practice trials, the experimental task followed. In the current study a picture was used where Clooney is facing sideways rather than directly into the camera to address the possible social threat explanation as explained above.[5] The picture was shown for 15 s. The reason for taking a slightly longer exposure time in this study was to assess more precisely the temporal unfolding of heart rate, and the time it took to return to baseline.

After the surprise, participants were requested to step off the balance board and sit behind a laptop computer to answer various background questions. In this study, we wanted to measure the subjective evaluations of the surprise more thoroughly (see also Experiment 2b in the Supplemental Materials) to check whether and how participants would differentiate between the unexpectedness and the content of the stimulus. To this end, we included several items to test that it is indeed the unexpectedness and surprise that underlies the initial physiological and behavioral responses, rather than anything related to the image per se. Even though it is hard to examine the phenomenon of surprise using self-reports as it involves a retrospective evaluation (see Noordewier et al., 2016), we asked participants to reflect on their initial feelings. In order to get to the initial response in the most general way possible, we first asked an open-ended question: "Recall the moment you saw the image. Describe below in a couple of sentences what you experienced at the moment you saw the image." Second, to differentiate the discrete emotion of surprise from other possible emotional responses we asked the question: "Recall the moment you saw the image. How did you feel at the moment you saw the image? Indicate to what extent the word below describes the feeling you had when you saw the image." Participants then indicated the extent to which the following emotions applied to them: "surprised", "scared", "happy", "angry", and "sad". Participants indicated their responses on a scale that ran from 1 = *not at all* to 7 = *extremely*.

Finally, we checked the relative judgements of the *unexpectedness of the stimulus* vs. the actual *content of the stimulus*. Our prediction was that the unexpectedness of the appearance of the image is relatively threatening, intimidating, and confusing, while the content of the stimulus would be relatively interesting and attractive. Moreover, this measure allowed us to check the extent to which our setting may involve levels of social threat (see Discussion of Experiment 1). To this end, we asked: "You just saw an image of a person that was shown unexpectedly. Please indicate below how you felt about the unexpectedness of the image and how you felt about the person in the image." Then we asked participants to rate the unexpectedness and the person separately with the questions "How did you feel about the fact that the image was shown unexpectedly?" and "How did you feel about the person in the image?". Participants then indicated how the following feelings had applied to them: "threatening", "intimidating", "confusing", "interesting", and "attractive". Participants provided their responses on a scale that ran from 1 = *not at all* to 7 = *extremely*. Note that we thus kept the items similar for both unexpectedness and content-ratings, to allow comparison between the two and to only vary the target of the emotion assessment.

At the end of the experiment, we administered the TIPI and we asked participants whether they recognized George Clooney (yes/no). Two participants indicated they did not recognize George Clooney, but excluding their data did not change the overall pattern of results, so we left it in. Finally, we asked for age, gender, and educational background (open question). We included the latter question because we anticipated that participants in the current study would come from a more diverse educational background than the participants in Experiment 1.

[5] Surprising target of Experiment 2: https://images.app.goo.gl/C1YAdfqCoARyPD5h9.

3.2. Results and discussion

We first analyzed the surprise checks (i.e., the open question, emotion ratings); then we analyzed the heart rate data and finally the body sway data. Finally, we checked the unexpectedness vs. person ratings.

3.2.1. Surprise checks

We first checked the answers to the open question about what participants experienced at the moment they saw the image. Two coders rated whether answers referred to surprise, unexpectedness, or interruption or whether they referred to anything else (i.e., confusion, other negative feelings, enjoyment, or task-related comments). Inter-coder agreement was 86.2 % and disagreement was resolved through discussion. Results (see Table 1) showed that surprise-related answers were the most common ones and when other answers were provided, they often appeared in combination with something on surprise.

Second, to further verify that participants were indeed surprised after they saw the unexpected picture, we analyzed the emotion ratings. This showed that participants reported strong surprise ($M = 5.69$, $SD = 0.77$), while levels of happiness fell around the midpoint of the rating-scale ($M = 4.27$, $SD = 1.12$), and levels of fear ($M = 1.50$, $SD = 0.78$), anger ($M = 1.30$, $SD = 0.81$), and sadness ($M = 1.28$, $SD = 0.72$) were generally low. Taken together, this shows that participants were indeed surprised after unexpectedly seeing George Clooney.

3.2.2. Heart rate (HR)

We then analyzed heart rate and body sway. There were no statistical outliers on HR. HR differed between the surprise block and baseline, Wilks' Lambda = .80, $F(1,61) = 15.14$, $p < .001$, $\eta_p^2 = .20$ [.07, .33]. HR was lower in the surprise block ($M = 89.46$, $SD = 16.28$) than during baseline ($M = 93.00$, $SD = 16.90$).

Next, we performed the second-to-second analyses to examine the temporal unfolding of HR changes after the surprise (see Fig. 2b). Results showed an effect of Time on HR, Wilks' Lambda = .48, $F(15,47) = 3.46$, $p < .001$, $\eta_p^2 = .52$ [.18, .53]. Simple contrasts with the second before the surprise as reference showed that HR at Second 1 after surprise was *higher*, $F(1,61) = 11.52$, $p = .001$, $\eta_p^2 = .16$, while it did not differ at Seconds 2-4, $Fs < 1.82$, $ps > .18$, $\eta_p^2 s < .03$. From Second 5 onwards, HR was consistently lower than just before the surprising stimulus, $Fs > 5.68$, $ps < .021$, $\eta_p^2 s > .08$.

3.2.3. Body sway

Following the HR-analyses, we also analyzed body sway following a second-to-second approach (see Table 2). A total of 10 values were Winsorized (six in AP direction, four in ML direction). After this, we tested whether participants showed reduced body sway after surprise by comparing the AP and ML movement in the five seconds after surprise to the second just before the surprise. On the AP and the ML axis, we found no effect of Time, Wilks' Lambda = .87, $F(5,40) = 1.16$, $p = .345$, $\eta_p^2 = .13$ [.00, .20], and Wilks' Lambda = .85, $F(5,40) = 1.43$, $p = .236$, $\eta_p^2 = .15$ [.00, .23], respectively.

Table 1
Answers to open-ended question on initial experience after seeing the unexpected image (Experiment 2).

Category	Of total	With surprise	Example
Surprise	60 %	–	Surprised
Confusion	52 %	50 %	Lack of understanding
Negative feelings	14 %	67 %	Shock
Enjoyment	26 %	65 %	Funny, had to laugh
Task-related	57 %	57 %	Wanting to say the answer
Other	15 %	–	No feelings

Table 2
Mean body sway in mm (*SD*) in the second before vs. seconds (s1-5) after surprise (Experiment 2).

	AP	ML
sec before	1.36 (0.88)	0.65 (0.33)
s1	1.11 (0.56)	0.70 (0.38)
s2	1.18 (0.59)	0.72 (0.37)
s3	1.28 (0.68)	0.76 (0.51)
s4	1.45 (1.27)	0.84 (0.71)
s5	1.29 (0.75)	0.94 (0.79)

Note. Movement variability is expressed in standard deviation of participants' center of pressure time series in the anterior-posterior (AP) and mediolateral (ML) direction.

3.2.4. Unexpectedness vs. content of the stimulus

Finally, we compared the ratings of the *unexpectedness* of the appearance of the image vs. the *content* of the image (i.e., the person) with paired sample *t*-tests (see Table 3 for means, *SD*s, and test statistics). This showed that the unexpectedness of the stimulus was rated as marginally more threatening and intimidating than the person. It should be noted though, that in both cases the scores are rather low. Next, in line with predictions, the unexpectedness was rated as more confusing than the person. Contrary to predictions, the unexpectedness was rated as more interesting than the person. Finally, the unexpectedness was rated as less attractive than the person. These results show that the unexpectedness of the appearance of the image mainly induces confusion, while the person in the image is not rated as intimidating (in fact, he is rated as *less* intimidating than the unexpectedness of his appearance per se). This suggest that the confusing nature of the unexpected stimulus most likely drives the initial response, more than the content of the stimulus. This makes an alternative explanation of the heart rate finding in terms of social threat (see Discussion Experiment 1) seem unlikely.

In sum, results on the self-report measures showed that the initial subjective appraisal is characterized by a sense of surprise, confusion, and unexpectedness. In addition, and in line with Experiment 1, heart rate was lower after the surprise than during baseline. However, the unfolding of the heart rate changes developed somewhat slower compared Experiment 1, as we observed an initial *increase* in heart rate in the first second after surprise, followed by a subsequent decrease. This result might be due to the slightly different setting in which the current study took place, as in the current study the participant was standing on a balance board, and the experimenter was present in the room. Together, this may have resulted in a slightly more tense state of the participant, affecting the initial heart rate. Note that a replication of this study without the balance board (Experiment 2b, described in the Supplemental Materials) showed a similar pattern as Experiment 1.

Finally, there was no surprise effect on movement. A factor that might have played a role is the somewhat smaller sample for the movement data relative to the heart rate data (see Footnote 4), resulting in somewhat lower statistical power to detect possible effects—also in relation to the fact that we only have one critical surprise trial per participant, rather than multiple trials which is more common in body sway studies (e.g., Hagenaars et al., 2012; Roelofs et al., 2010). In addition, it could also be possible that reductions in body sway are specific for freezing after threat. However, before drawing more definite conclusions about this, the current study needs to be replicated. Future work may also include electromyographic (EMG) recordings of the lower legs (see e.g., Adkin & Carpenter, 2018), as this may reveal co-contraction, i.e., increased tension of agonist-antagonist muscle pairs, which results in rigidity of the legs and reduced movement.

In the first experiments we used a target that was relatively neutral. As outlined above, we argue that it is the unexpectedness rather than the valence of the target that drives surprise. Following our reasoning, decreased heart should also occur after clearly positive surprises. Therefore, in Experiment 3 we examined heart rate in response to an unexpected but clearly positive target.

Table 3
Mean emotion ratings (*SD*) as a function of target of rating (unexpectedness vs. person; Experiment 2).

	Unexpectedness	Person	*t*(63)	*p*	*d*[CI]
Threatening	1.94^{a} (1.30)	1.64^{b*} (1.01)	1.99	.050	0.26 [0.01, 0.51]
Intimidating	2.14^{a} (1.26)	1.92^{b*} (1.20)	1.67	.099	0.18 [−0.03, 0.39]
Confusing	5.56^{a} (1.18)	3.73^{b} (1.87)	8.03	<.001	1.17 [0.87, 1.50]
Interesting	4.70^{a} (1.16)	4.09^{b} (1.44)	3.77	<.001	0.47 [0.22, 0.72]
Attractive	3.09^{a} (1.44)	3.48^{b} (1.56)	−2.44	.017	−0.26 [−0.47, −0.05]

Note. Means with different superscripts in rows differ at $p < .02$ and * differ at $p < .10$ in paired sample *t*-tests.

4. Experiment 3

With Experiment 3 we aimed to conceptually replicate and extend the findings of Experiments 1 and 2 by testing whether a heart rate deceleration also occurs in response to a more unambiguously positive surprise. We used the same experimental set-up as Noordewier and Van Dijk (2019; Study 1), where, as a surprise induction, participants were unexpectedly presented with a picture of a puppy.

4.1. Method

A total of 68 participants were recruited at the Social Science Faculty of Leiden University (51 females, 16 males, 1 unknown; $M_{age} = 22.79$, $SD_{age} = 6.20$). A total of 12 participants were excluded from analyses, for one of the following reasons: ECG equipment error (a cable connection mistake; $n = 7$); the door of the cubicle opened during the study ($n = 1$), being unable to understand the instructions (because Dutch was not the first language of the participant; $n = 1$); a noisy and therefore unscorable ECG ($n = 1$); or having participated before in a similar study (i.e., Noordewier & Van Dijk, 2019, Study 1; $n = 2$). We report analyses on the data of the remaining 56 participants (44 females, 12 males; $M_{age} = 22.84$, $SD_{age} = 6.64$). A sensitivity power analysis showed that the minimum effect size to consider the observed effect as relevant is $\eta_p^2 = .13$ in the block-comparison and $\eta_p^2 = .38$ in 1-second epoch comparisons.

Similar to the previous studies, the experiment had a within-subjects design (baseline vs. surprise block), using pictures of buildings in the repetition-phase and a picture of a puppy as a positive surprise. Throughout the study, heart rate was measured with ECG in the same way as in Experiment 1.

4.1.1. Materials and procedure

After signing informed consent, participants were connected to the ECG equipment. As in Experiments 1 and 2, participants were asked to stand upright throughout the study and first watched a neutral movie during which we collected baseline ECG recordings. Next, participants were presented with the same repetition-change task as in Noordewier and Van Dijk (2019; Study 1). Each trial consisted of four different pictures of four buildings that were presented at 1.5-second intervals, after which the question "Was there any green in the last picture?" followed. Different from Noordewier and Van Dijk (2019) but similar to Experiments 1 and 2, participants provided their answers to this question verbally.

After four practice trials, 19 trials followed. Then, the critical surprise trial was presented, which was a puppy that moved its head and paw towards the camera. The puppy was presented for 15 s (for details, see Noordewier & Van Dijk, 2019; compared to the original study we used four more building-trials to further strengthen the expectancy and

presented the puppy 6 s longer to be able to track heart rate development after surprise). After the surprise, participants were asked "To what extent were you surprised by the dog?" from 1 = *not at all* to 7 = *extremely* and "What did you think of the dog?" from 1 = *negative* to 7 = p*ositive.* After filling out the TIPI, they reported their age, gender (male, female, other/rather not say), whether they participated before in a similar study (yes/no), and whether they were native-Dutch (yes/no; during that time, international students were recruited for other studies; five participants indicated that Dutch was not their first language, but they understood Dutch well enough to understand the instructions, and excluding their data did not change the results; therefore it remained part of the analyses).

4.2. Results and discussion

4.2.1. Surprise and valence check

First, we analyzed the surprise and valence ratings. Participants rated the intensity of their surprise as $M = 5.41$ ($SD = 1.53$) and the valence of the stimulus as $M = 5.75$ ($SD = 1.40$). This suggest that we successfully created a surprising event with a stimulus that was evaluated relatively positive.

4.2.2. Heart rate (HR)

Next, we compared HR during the surprise block to HR during baseline. There were no statistical outliers. Results showed an effect of block, Wilks' Lambda = .56, $F(1,55) = 43.57$, $p < .001$, $\eta_p^2 = .44$ [.27, .56], such that HR was lower during the surprise block ($M = 90.95$, $SD = 15.01$) than during baseline ($M = 95.53$, $SD = 16.02$).

Then, we analyzed the temporal unfolding of HR changes, using the same approach as in the previous studies (see Fig. 2c). The HR data of one participant showed six statistical outliers, which were Winsorized. Results showed an effect of Time, Wilks' Lambda = .25, $F(15,41) = 8.37$, $p < .001$, $\eta_p^2 = .75$ [.51, .75]. Simple contrasts with the second before surprise as a reference showed that from Second 2 onwards HR was consistently lower, $Fs > 15.99$, $ps < .001$, $\eta_p^2 s > .22$, while at Second 1 after the surprise effect was $F(1,55) = 0.08$, $p = .782$, $\eta_p^2 < .01$.

Taken together, these results showed that also after a positive surprise, heart rate decreases. Thus, we find consistent evidence for lowered heart rate after a rather neutral (Experiments 1-2) as well as a positive surprise (Experiment 3). However, to further substantiate the influence of the valence of the event on heart rate reduction after a surprising event, a more direct comparison between positive and negative surprises within a single study is needed. Testing this was the aim with the fourth and final study. Moreover, in this final study we included two additional physiological measures: finger temperature and blood pressure. A final goal with Experiment 4 was to test possible habituation after repeated exposure to (initially) unexpected stimuli.

5. Experiment 4

In Experiment 4, we directly compared positive and negative surprises and included blood pressure and finger temperature as additional physiological measures next to heart rate. Previous work with rats has shown that temperature at the extremities of the body lowers as part of the freezing response. This lowered skin temperature is the result of blood moving from the skin to the muscles to prepare for possible action (e.g., Vianna & Carrive, 2005). For this process to occur, arteries contract which should in turn increase blood pressure (Carrive, 2000; Vianna & Carrive, 2005). The current study showed no effect of surprise on temperature, which may be partially due to suboptimal measurement (see Footnote 6). For ease of presentation, we only report the heart rate and blood pressure results here. All information on temperature can be found in the Supplemental Materials.

Experiment 4 used the same repetition-change paradigm as in the previous experiments, but instead of one repetition-change cycle, we repeated this cycle five times. This allowed us to compare positive and negative stimuli that are unexpected and therefore surprising (Block 1) with positive and negative stimuli that are less unexpected and surprising (Blocks 2-5). On the basis of the previous experiments and our reasoning that the initial response after a surprising event is primarily driven by the unexpectedness of the event rather than the valence of the surprising stimulus, we predicted that in Block 1, positive and negative stimuli would result in similar physiological responses (i.e., lower heart rate, lower temperature, higher blood pressure compared to baseline). However, in Blocks 2-5 responses will probably start to differentiate depending on the valence of the stimulus: Negative stimuli are predicted to result in sustained freezing-like responses because of the negativity of the images, while responses to positive stimuli are predicted to go back to baseline in later blocks because the positive images are not threatening.

5.1. Method

A total of 114 participants were recruited at the Social Science Faculty of Leiden University (97 females, 17 males; $M_{age} = 21.02$, $SD_{age} = 4.49$).[6] We excluded the data of 11 participants from analyses: Six participated in a similar study before; two interrupted the task (one because she was colorblind and asked for help, and one because she felt uncomfortable due to the blood pressure measurement); one displayed an odd response-pattern on the self-report measures (he answered 1 in 87% of the cases), and two because their physiological data was not saved. We report analyses on the remaining 103 participants (89 females, 14 males; $M_{age} = 20.67$, $SD_{age} = 2.63$). A sensitivity power analysis showed that the minimum effect size to consider the observed effect as relevant is in the block-comparisons $\eta_p^2 = .12$, and in the 1-second epoch comparisons $\eta_p^2 = .16$.

Participants were randomly assigned to one of the two conditions of a mixed design with Valence (positive vs. negative) as between-subjects factor and Block (6) as within-subjects factor. Dependent measures were heart rate, finger temperature, and blood pressure. In addition, we measured self-reported feelings and ratings of the pictures of the (initially) unexpected stimuli to check the valence of the stimuli and the extent to which it was surprising.

5.1.1. Materials and procedure

After signing the informed consent, participants were led to an individual room where they were connected to the measurement equipment. As in the other experiments, physiological signals were sampled using a Biopac MP150 system. As in Experiments 1, we measured ECG using an ECG100c module, and a Lead I electrode configuration.

Blood pressure was measured using a Nexfin HD system (Bmeye B.V., Amsterdam, The Netherlands). The Nexfin HD comprises an inflatable finger cuff that is attached around the middle phalanx of the ring finger of the participant's non-dominant hand. The blood pressure signal was measured at a sample rate of 200 Hz; a 20 Hz low-pass filter was applied offline. Like the ECG, the blood pressure signal was automatically scored (after visual inspection) in the physio data toolbox, which yielded measures of systolic and diastolic blood pressure as output.

Participants were seated during the study. They were presented with the baseline video, which was followed by the repetition-change task comprising geometrical shapes as used in Experiments 1-2. The repetition-change cycle was repeated five times, creating the within-subjects factor "Block". To prevent that participants could anticipate

[6] The data were collected in two waves (Spring [lab1]: $N = 32$ and Fall [lab2]: $N = 82$). Design and procedures were exactly the same, except for a different lab-room and temperature sensor. HR and temperature were lower, and systolic BP was higher in lab 1 relative to lab 2 (see Supplemental Materials, Table S4), but there were no interactions with Block, except for a Block x Temperature interaction: Participants' fingers got cooler during the experiment in Lab 1 and not in Lab 2, probably due to air conditioning in Lab 1.

the change simply by counting the trials, we varied the number of shape-trials across blocks (i.e., Blocks 1-5 had 16, 12, 13, 15, 11 trials, respectively).

We manipulated the valence of the change trials between subjects in two levels (positive vs. negative), such that after each of the five repetition-change cycles, one of five different pictures of positive or negative dogs were shown (in random order). For this we used pictures of dogs that had either a cute or aggressive visual appearance. The pictures were derived from the internet and Photoshopped on a light grey background. The cute vs. aggressively-looking dogs were matched on breed (i.e., Rottweiler, Pitbull, Dalmatian, German Shepherd, Labrador) and as much as possible on position and color. Each dog was presented on the screen for 10 s. Between each block, there was a 15-second break.

After the fifth block, participants were informed that the task was completed and they were then asked to answer several manipulation-check questions. We measured both the *general* feelings participants had when the pictures with the dogs appeared, as well as the specific *initial feeling* that the participants had when the first dog had appeared. We first asked "You have seen different pictures of dogs. How did you feel when you saw the pictures of the dogs?" Participants then rated the extent to which they had felt "surprised", "scared", and "happy" on a scale from 1 = *not at all* to 7 = *extremely*. After this, we measured how they had felt when the *first* dog had appeared: "Think back to the moment you saw the first picture of a dog. How did you feel when you saw a picture of a dog the first time?". Participants then rated the same emotions as for the general feelings. Note that this self-report is delayed, in the sense that participants were asked to report their feelings retrospectively. This is the case for all studies reported here, but more so in the current study involving repeated blocks. Previous research using recall procedures, however, showed that participants are able to recall their initial feelings after unexpectedness (Noordewier & Breugelmans, 2013).

Next, to check the valence of the stimuli, we asked participants to evaluate the pictures of the dogs they had seen on a scale from 1 = *negative* to 7 = *positive*. Participants in the negative surprise condition then saw the pictures of the positive dogs, to ease any discomfort they could have experienced due to the negative pictures.

Then, we asked participants to indicate their age, gender (male, female, other/rather not say), and whether they participated in a study with a similar set-up before (i.e., with repeated geometrical shapes; yes/no). Finally, to be able to check whether liking of dogs would affect the results (e.g., making the positive surprise less positive), we asked participants to indicate the extent to which they agreed with the statement "I like dogs" on a scale from 1 = *not at all* to 7 = *extremely*, where we emphasized that we were interested in their opinion in daily life, independent of the task they just completed. Results with this measure included can be found in the Supplemental Materials (i.e., dog liking did not affect the pattern of results for HR and BP, while for temperature it seemed that stronger dog liking related to a larger decrease in temperature between the baseline and Block 1). Participants were then fully debriefed and rewarded for their participation.

5.2. Results and discussion

We analyzed the data in three steps. First, we analyzed the manipulation checks. Second, we checked for block- and valence effects, by comparing the averaged physiological responses in each of the five blocks vs. baseline for the positive vs. negative condition. Third, we zoomed-in on Block 1 and checked for valence effects in second-to-second analyses of the physiological responses after surprise as compared to the last second before the surprise. Predictions on *differences* after surprise were tested with repeated measures MANOVAs and simple contrasts; predictions on (initial) *similarity* between valence conditions were tested with equivalence tests (more details below).

5.2.1. Step 1: manipulations checks

First, we analyzed the self-reported impact of the manipulation. Analyses of the dog evaluations showed some variation in the evaluations of the different dogs, but overall, the positive dogs were rated as more positive than the negative dogs (see Table S3 of Supplemental Materials).

5.2.1.1. Feelings. We checked participants' feelings when they saw the pictures, using repeated measures MANOVAs with focus (general feelings vs. feelings after first dog) as within-subjects factor and valence (positive vs. negative) as between-subjects factor (see Table 4 for all statistics).

On "surprise", there was an effect of focus, no effect of valence, nor an interaction. The pictures of the dogs were rated as surprising, particularly when they were encountered for the first time. On "happy", there was an effect of valence, no effect of focus, nor an interaction. Participants were happier in the positive than in the negative condition, both in general as when seeing a dog for the first time. Finally, on "scared", there was an effect of valence, an effect of focus, and an interaction. Participants were less scared in the positive than in the negative condition, both when they saw a dog for the first time as in general, and participants in the negative condition were particularly scared when they saw the first dog (vs. seeing the dogs in general), all *ps* $< .03$.

Taken together, participants were surprised by both the positive and negative dogs, while the positive dogs were more positive and less negative than the negative dogs. As such, we conclude that we successfully created a positive vs. negative surprise.

5.2.2. Step 2: block comparison of physiological responses

Next, we analyzed the physiological data in the blocks. The heart rate data of four participants were excluded due to poor signal quality. On systolic/diastolic blood pressure, one participant had missing data for all blocks, and another participant had missing data in Blocks 1 and 2; the blood pressure data of two further participants were excluded because of poor signal quality in one of the blocks (Block 1 for one participant, Block 2 for the other). All these cases concerned different participants.

We conducted our analyses in two ways. First, to test our hypothesis that the surprising stimulus in Block 1 decreased heart rate and increased systolic/diastolic blood pressure, we conducted a series of repeated measures MANOVAs with Block (6 levels: baseline and 5 change-blocks) as within-participant variable and Valence (positive vs. negative) as between-participants variable on heart rate, and systolic/diastolic blood pressure (see Table 5a). Second, to test our hypotheses that the means in Block 1 were initially *similar* for the positive and negative valence conditions we used equivalence tests (see Table 6afor all results). That is, where MANOVAs test for differences, equivalence tests establish whether means are equivalent—defined as when a difference between conditions is zero or smaller than is deemed meaningful (Lakens, McLatchie, Isager, Scheel, & Dienes, 2020; Lakens, Scheel, & Isager, 2018). The equivalence tests were conducted according to guidelines by Lakens et al. (2018, 2020). The smallest effect size of interest (SESOI) was used as a reference point in the analyses, which was determined using sensitivity power analyses for each dependent measure ($\alpha = .05$ and 80 % power; we used sensitivity power analyses, because we did not have a clear benchmark or prior studies to go on). This gave a lower equivalence bound ΔL and an upper equivalence bound ΔU of $d = -0.399$ and $d = 0.399$ (unless specified differently). Next, we performed two one-sided Welch's *t*-tests (Delacre, Lakens, & Leys, 2017) against each of these equivalence bounds to examine whether we can reject the presence of a meaningful effect. Means are equivalent when both the one-sided tests are significant. Note that even though our prediction was that responses would be similar during the first seconds, we tested the equivalence of all seconds after the surprise, because we did not know exactly when responses would start to

Table 4
Mean emotion-ratings (*SD*) as a function of Valence (positive vs. negative) and Focus (general vs. initial emotions; Experiment 4).

	Valence							
	Positive		Negative					
	General	Initial	General	Initial		Wilks' Lambda	$F(1,101)$	p	η_p^2 [CI]
Surprise	5.13[a] (1.33)	6.38[b] (0.93)	4.92[a] (1.26)	6.04[b] (1.20)	Valence	–	1.88	.173	.02 [.00, .08]
					Focus	.49	103.36	<.001	.51 [.39, .59]
					Valence × Focus	> .99	0.32	.571	.00 [.00, .04]
Happy	5.12[a] (1.41)	4.96[a] (1.70)	2.18[b] (1.40)	2.16[b] (1.58)	Valence	–	97.59	<.001	.49 [.37, .58]
					Focus	.99	1.23	.271	.01 [.00, .06]
					Valence × Focus	.99	0.74	.393	.01 [.00, .06]
Scared	1.42[a] (0.98)	1.25[a] (0.65)	3.80[b] (1.96)	3.02[c] (1.70)	Valence	–	59.35	<.001	.37 [.25, .47]
					Focus	.72	39.68	<.001	.28 [.16, .39]
					Valence × Focus	.86	16.17	<.001	.14 [.05, .24]

Note. Means with different subscripts in rows differ at $p < .05$ in a repeated measures analyses, with simple contrast in case of a focus-valence interaction.

Table 5a
Repeated measures analysis with Block (baseline and blocks 1-5), Valence (positive vs. negative) and their interaction, on heart rate (HR), systolic blood pressure (SBP), and diastolic blood pressure (DBP; Experiment 4).

		Wilks' Lambda	F	df	p	η_p^2 [CI]
HR	Block	.73	7.07	(5,93)	<.001	.28 [.12, .36]
	Valence	–	2.60	(1,97)	.110	.03 [.00, .10]
	Block × Valence	.99	0.28	(5,93)	.923	.02 [.00, .02]
SBP	Block	.56	14.72	(5,93)	<.001	.44 [.29, .52]
	Valence	–	0.01	(1,97)	.914	.00 [.00, .00]
	Block × Valence	.97	0.61	(5,93)	.696	.03 [.00, .05]
DBP	Block	.72	7.25	(5,93)	<.001	.28 [.13, .36]
	Valence	–	0.01	(1,97)	.922	.00 [.00, .03]
	Block × Valence	.98	0.42	(5,93)	.837	.02 [.00, .03]

Table 5b
Repeated measures analysis with Time (second before surprise vs. 10 seconds after surprise) and Valence (positive vs. negative) and their interaction, on heart rate (HR), systolic blood pressure (SBP), and diastolic blood pressure (DBP; Experiment 4).

		Wilks' Lambda	F	df	p	η_p^2 [CI]
HR	Time	.46	10.35	(10,88)	<.001	.54 [.37, .59]
	Valence	–	2.20	(1,97)	.141	.02 [.00, .09]
	Time × Valence	.92	0.81	(10,88)	.624	.08 [.00, .09]
SBP	Time	.53	7.71	(10,87)	<.001	.47 [.28, .52]
	Valence	–	0.13	(1,96)	.717	.00 [.00, .04]
	Time × Valence	.83	1.80	(10,87)	.073	.17 [.00, .20]
DBP	Time	.57	6.49	(10,87)	<.001	.43 [.23, .48]
	Valence	–	0.07	(1,96)	.789	.00 [.00, .03]
	Time × Valence	.91	0.90	(10,87)	.539	.09 [.00, .10]

differentiate, if at all.

5.2.2.1. Heart rate (HR). On HR, one statistical outlier was Winsorized (in Block 3). Results showed (see Fig. 3a) an effect of Block, no effect of Valence, and no Block x Valence interaction. Simple contrasts showed that HR was lower than baseline in the first block, $F(1,97) = 23.98$, $p < .001$, $\eta_p^2 = .20$, but not in the other blocks, $Fs < 0.70$, $ps > .405$, $\eta_p^2 < .01$. Finally, we tested the equivalence of valence conditions on HR difference scores.[7] Results showed (see Table 6a) that HR was similar in Blocks 1 and 3; for Blocks 2/4/5, the *p*s ranged between .057 and .072.

Thus, HR became lower when participants encountered the first picture of a dog, irrespective of whether this was a positive or a negative dog. Moreover, after the surprise in the first block, HR was no longer different from baseline levels in the blocks that followed. Finally, note that the overall heart rate in the current experiment is lower than it was in Experiments 1-3. This can be explained by the fact that participants sat down during the current experiment (similar to Experiment 2b, described in the Supplemental Materials), while they were standing during Experiments 1-3.

5.2.2.2. Systolic blood pressure (SBP). On SBP, a total of seven statistical outlying values in the data of two participants were Winsorized (in Blocks 1-5). After this, results showed (see Fig. 3b) an effect of Block, no effect of Valence, and no Block x Valence interaction. Simple contrasts showed that SBP was higher than baseline in all blocks, $Fs > 35.07$, $ps < .001$ η_p^2s $> .26$. Finally, we tested the equivalence of the valence conditions. Results showed that SBP was similar in Blocks 2-4; in Blocks 1 and 5 the *p*s were .058 and .051, respectively.

5.2.2.3. Diastolic blood pressure (DBP). On DBP, a total of nine statistical outlying values of two participants were Winsorized (baseline and Blocks 1–5). Results showed (see Fig. 3c) an effect of Block, no effect of Valence, and no Block x Valence interaction. Simple contrasts showed that DBP was higher than baseline in all blocks, $Fs > 12.59$, $ps < .002$, η_p^2s $> .11$. Finally, we tested the equivalence of the valence conditions. Results showed that DBP was similar in Blocks 2–5; for Block 1 the *p*-value was .058.

Thus, both SBP and DBP became higher after the surprise—irrespective of whether the surprise was positive or negative—and remained high in all subsequent blocks.

[7] At baseline, there were no differences between the positive and the negative valence condition, $t(96.9) = 1.54$, $p = .127$, $d = .309$ (i.e., a *t*-test comparing the differences). However, based on equivalence bounds, we cannot reject effect sizes that we still consider meaningful at baseline, $t(96.9) = -0.45$, $p = .327$. This suggests that there may be differences between conditions at baseline. Therefore, we used difference scores in our analyses (i.e., block minus baseline). Analyses on the raw scores can be found in the Supplemental Materials (see Table S5a).

Table 6a
Equivalence tests for heart rate (HR), systolic blood pressure (SBP), and diastolic blood pressure (DBP) between the positive vs. negative valence condition for baseline (b) and Blocks (b1-5) after surprise (Experiment 4).

	HR			SBP			SBP		
	df	*t*	*p*	df	*t*	*p*	df	*t*	*p*
b	–	–	–	97.00	−1.92	.029	96.75	−1.92	.029
b1	90.85	1.93	.029	96.87	1.58	.058	96.85	1.59	.058
b2	84.06	−1.59	.057	95.70	1.90	.030	96.97	1.86	.033
b3	87.59	1.74	.043	95.38	−1.98	.025	96.97	1.97	.026
b4	82.81	1.48	.072	94.98	−1.88	.032	96.89	1.97	.026
b5	96.94	−1.55	.063	96.99	1.65	.051	96.85	1.91	.030

Note. HR analyses are conducted on difference scores (see Results for more details). *T*-tests for differences between valence conditions within blocks showed HR *ps* > .621; SBP *ps* > .689; DBP *ps* > .693. Equivalence bounds were set at $d = -0.399$ and $d = 0.399$.

Table 6b
Equivalence tests for heart rate (HR), systolic blood pressure (SBP), and diastolic blood pressure (DBP) between the positive vs. negative valence condition for baseline (b) and seconds (s1-10) after surprise (Experiment 4).

	HR			SBP			DBP		
	df	*t*	*p*	df	*t*	*p*	df	*t*	*p*
b	–	–	–	95.92	−1.80	.037	95.82	−1.78	.039
s1	96.23	1.79	.038	95.97	−1.85	.034	95.85	−1.85	.033
s2	96.96	1.50	.069	95.99	−1.87	.032	95.69	1.90	.030
s3	97.00	1.44	.076	95.96	−1.93	.028	95.84	1.79	.038
s4	95.99	−1.79	.038	96.00	1.55	.062	95.99	1.42	.079
s5	96.16	−1.61	.056	95.99	1.34	.091	95.91	1.43	.078
s6	96.08	−1.97	.026	95.76	1.49	.070	95.73	1.40	.082
s7	95.74	−1.31	.097	95.71	1.29	.101	95.57	1.44	.076
s8	96.60	−0.34	.367	95.72	1.06	.146	95.34	1.54	.063
s9	93.91	−0.77	.221	95.89	1.25	.107	95.70	1.82	.036
s10	91.27	−1.30	.099	95.91	1.58	.059	95.48	−1.98	.025

Note. Heart rate analyses are conducted on difference scores (see Results for more details). *T*-tests for differences between valence conditions within blocks (incl. baseline) showed: HR *ps* > .104; SBP *ps* > .358; DBP *ps* >.563. Equivalence bounds were set at $d = -0.399$ and $d = 0.399$ for heart rate and temperature and at $d = -0.401$ and $d = 0.401$ for S/DBP (due to differences in missing data).

5.2.3. Step 3: second-to-second comparisons of physiological responses in Block 1

Finally, we again conducted second-to-second analyses to examine possible valence effects in the temporal unfolding of the physiological responses in the first block, where the positive or negative stimulus was still totally unexpected. For each measure, we compared each second after the surprise to the last second just before surprise (see Table 5b), followed by equivalence tests of the valence conditions (see Table 6b).

5.2.3.1. Heart rate (HR). There were no statistical outliers on HR. Results showed (see Fig. 4a) a main effect of Time, no effect of Valence, and no Time x Valence interaction. HR was consistently lower on all seconds after the surprise than the second just before the surprise, *Fs* > 7.09, *ps* < .001, η_p^2s > .06. Finally, we tested the equivalence of the valence conditions—again on difference scores.[8] Results showed that HR is similar during Seconds 1, 4, and 6. For Seconds 2, 3, 5, 7 and 10, the *p*s ranged between .069 and .099. For Seconds 8 and 9, similarity cannot be concluded, with $p = .367$ and $p = .221$, respectively.

5.2.3.2. Systolic blood pressure (SBP). On SBP, a total of 13 statistical outlying values in the data of two participants were Winsorized (Seconds 6–10 for one participant; Seconds 1–7 and the second before the surprise for the other participant). After this, results showed (see Fig. 4b) a main effect of Time, no effect of Valence, and a trend for a Time x Valence interaction. Note that without Winsorizing this interaction had a *p*-value of .110. Contrasts comparing the seconds after the surprise to the second before the surprise showed that SBP was higher after the surprise on Seconds 2–6 (*ps* < .041), but there were no differences at Seconds 1 and 7–10 (*ps* > .111). Next, contrasts of the Time x Valence interaction showed that on Seconds 1–3 after surprise SBP did not differ between the positive and the negative condition, *Fs* < 0.23, *ps* > .63, η_p^2s < .003. On Seconds 4 and 6, SBP was marginally higher in the negative than in the positive condition, $F(1,96) = 3.83$, $p = .053$, $\eta_p^2 = .04$ and $F(1,96) = 2.84$, $p = .095$, $\eta_p^2 = .03$, and on Seconds 5 and 7-9, SBP was higher in the negative than in the positive condition, *Fs* > 4.59, *ps* < .036, η_p^2s > .04. On Second 10, the two conditions did not differ, $F(1,96) = 2.51$, $p = .116$, $\eta_p^2 = .03$. Finally, we tested the equivalence of the valence conditions. Results showed that SBP was similar during Seconds 1–3, with *p*s between .028 and .034; for Seconds 4–6/10, the *p*s ranged between .062 and .091.

5.2.3.3. Diastolic blood pressure (DBP). On DBP, a total of 17 statistical outlying values in the data of two participants were Winsorized (Seconds 1–5 and the second before for one participant; all values for the other participant). After this, results showed (see Fig. 4c) main effect of Time, no effect of Valence, and no Time x Valence interaction. Contrast analyses showed that—irrespective of valence—DBP was higher on Seconds 1–5 after the surprise than DBP just before the surprise, *Fs* > 3.99, *ps* < .049, η_p^2s > .03. On Seconds 8–10 after surprise, DBP was *lower* than just before the surprise, *Fs* > 5.13, *ps* < .027, η_p^2s > .05. On Seconds 6–7 after the surprise DBP did not differ from DBP just before surprise, *Fs* < 0.65, *ps* > .423, η_p^2s < .008. Finally, we tested the equivalence of the valence conditions. Results showed equivalence of DBP on seconds 1–3/9–10; for seconds 4–8, the *p*s ranged between .063 and .082.

In sum, the *block-comparisons* showed that for both the positive and negative stimuli HR was lower and BP higher in Block 1 than during baseline. In subsequent blocks, when the positive and negative stimuli

[8] That is, even though at baseline the valence conditions did not differ, $t(96.84) = 1.15$, $p = .252$, d = .23 [-0.17, 0.64], based on the equivalence test we cannot conclude that they are similar either, $t(96.84) = -0.83$, $p = .204$. Analyses on the raw scores can be found in the Supplemental Materials (see Table S5b).

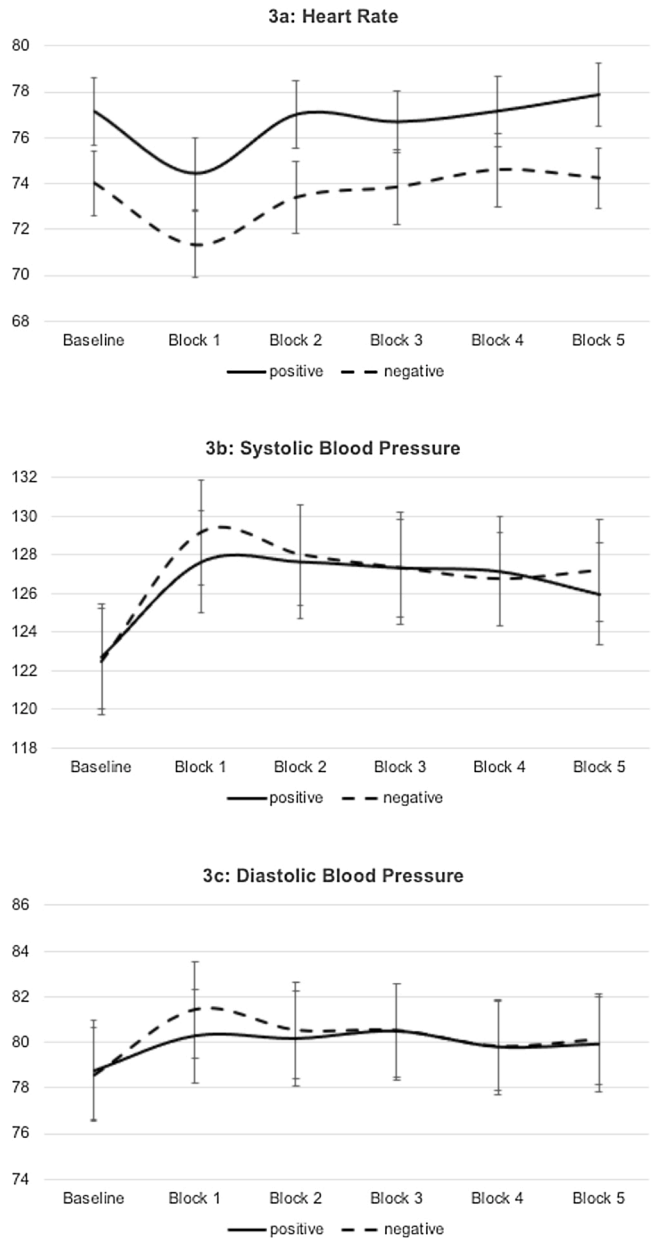

Fig. 3. Heart rate (3a), systolic blood pressure (3b), and diastolic blood pressure (3c) as a function of Valence of stimuli (positive vs. negative dogs) and Block (baseline vs. Block 1–5; Experiment 4). Error bars indicate +/− 1 SE.

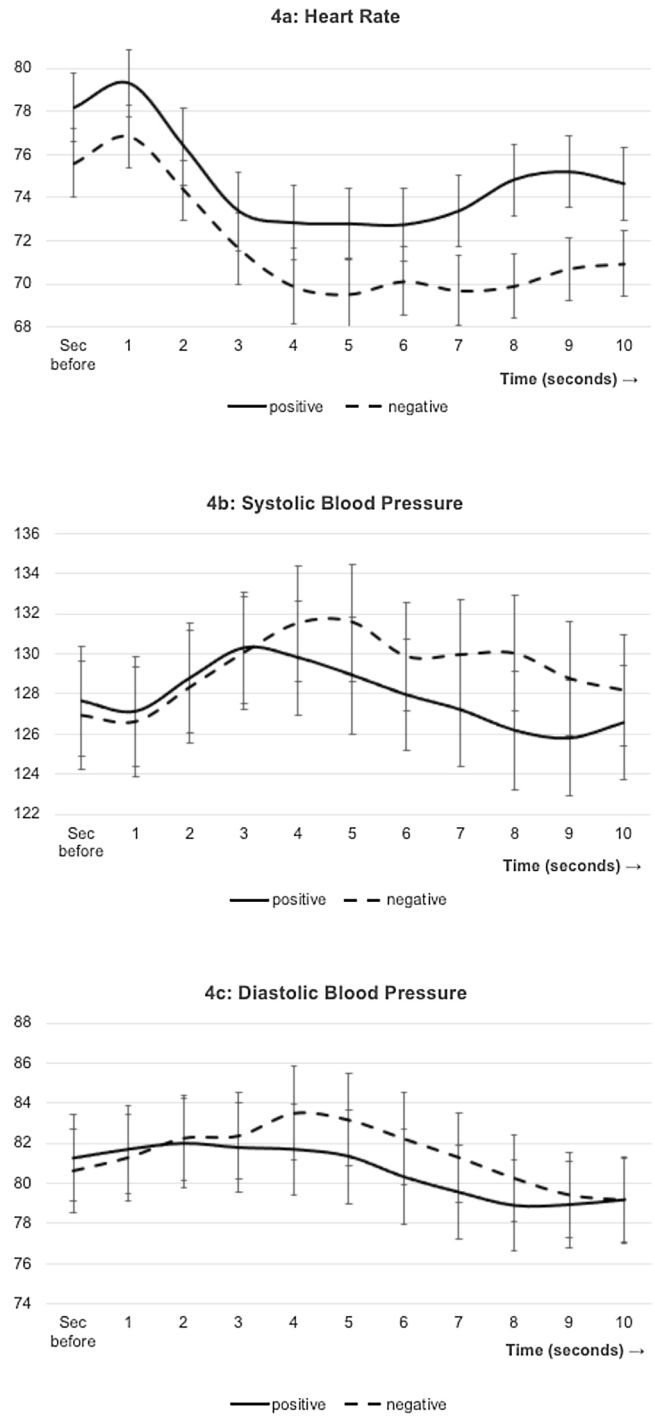

Fig. 4. Change in heart rate (4a), systolic blood pressure (4b), and diastolic blood pressure (4c) as a function of Valence of stimulus (positive vs. negative dog) and Time (the second before the surprise vs. seconds after surprise; Experiment 4). Error bars indicate +/− 1 SE.

were less/not surprising anymore, HR returned back to baseline, while BP remained somewhat higher than baseline for both the positive and the negative condition. The fact that HR returned to baseline in later blocks is an important finding as it also rules out that our effects are (partly) explained by a reduction in HR because participants became simply more relaxed during the experiment. The *second-to-second comparisons* within Block 1 showed that HR was consistently low, irrespective of the valence of surprise (in line with Experiments 1 and 3; note that in Experiment 2 this pattern also occurred, but somewhat delayed), while BP increased relative to the second before the surprise. For systolic BP there was a tendency for a moderation by valence: After negative vs. positive stimuli, systolic BP remained higher than baseline for a longer duration. Thus, BP increased generally after a surprise and especially systolic BP remained higher after a negative surprise; after the initial surprise in the first block, BP remained somewhat higher in subsequent blocks compared to baseline—possibly as part of the anticipation of further unexpected stimuli.

Taken together, Experiment 4 shows that surprise decreases HR and increases BP. These effects occur for both positive and negative surprise although systolic BP remained higher for a longer duration for a negative than for a positive surprise.

6. General discussion

The current research aimed to identify the physiological correlates of surprise, to better understand the psychology of initial interruption after

unexpectedness (Horstman, 2006, 2015; Meyer et al., 1997; Noordewier & Van Dijk, 2019; Noordewier et al., 2016; Reisenzein et al., 2017). We reasoned that physiological markers of surprise might correspond to those of orienting and freezing, as they have in common that they involve interruption and increased attention for effective action (Gladwin et al., 2016; Graham & Clifton, 1966; Hashemi et al., 2019; Horstman, 2006, 2015; Reisenzein et al., 2017; Roelofs, 2017; Scherer et al., 2004; Packer et al., 1989; Parmentier et al., 2018; Walker & Carrive, 2003; Wessel, 2017).

In four experiments, we presented participants with unexpected stimuli in a repetition-change paradigm. In all experiments, we focused on a primary physiological indicator of orienting and freezing: lowered heart rate. In Experiment 2, we also tested whether surprise would lower body movement and in Experiment 4, we tested whether surprise results in lower temperature and higher blood pressure. In addition, because people initially respond primarily to the unexpectedness of an event irrespective of its valence (Noordewier et al., 2016), these responses were predicted to occur irrespective of the surprising stimulus' valence.

We used neutral, positive, and negative stimuli to induce surprise, and self-report measures confirmed that these stimuli were indeed surprising (Experiments 1-3), particularly during the initial confrontation with unexpectedness (Experiments 2 and 4), which was also associated with confusion (Experiments 2; see also Experiment 2b in the Supplemental Materials). All experiments showed that the surprising stimuli indeed lowered heart rate, both compared to a baseline (Experiments 1-4) as well as in second-to-second comparisons relative to the second just before the surprising event (Experiments 1/3/4; in Experiment 2 this pattern also occurred, but somewhat delayed).[9] In Experiment 4, we presented stimuli repeatedly to test effects of surprise (initial exposure to the unexpected stimulus in the first block) vs. non-surprise (the reoccurrence of a similar stimulus in subsequent blocks). In the first block (when stimuli were surprising), heart rate was lower than baseline, while in the later blocks (when stimuli were not surprising anymore) it returned to baseline. These effects were not influenced by the valence of the stimulus.

We also found that surprise increases systolic and diastolic blood pressure (Experiment 4). Blood pressure was higher in the surprising first block than during the baseline, and it also remained somewhat higher than baseline in the later non-surprising blocks; again, valence did not moderate this pattern of results. However, second-to-second analyses within the first surprising block showed that the increase in blood pressure was especially pronounced (i.e., longer duration) for systolic blood pressure after a negative surprise. We did not find effects of surprise on temperature (Experiment 4, described in the Supplemental Materials) or movement (Experiment 2).

The blood pressure findings are noteworthy because blood pressure has thus far not been included in orienting or human freezing studies, nor in surprise research. Moreover, the combination of a lowered heart rate *with* an increased blood pressure is suggestive of both an inhibitory response and a relatively active state (Reyes del Paso et al., 1994; Sawada, 2003). This inhibitory response is functional for conserving energy when orienting before instigating an appropriate response and is marked by decreased heart rate; the relative active state (i.e., increased blood pressure) might be indicative of a state of vigilance, which is functional in considering what is happening next, what an appropriate response might be, and when to instigate it (Dorr, Brosschot, Sollers, & Thayer, 2007; see also Gladwin et al., 2016; Hashemi et al., 2019; Roelofs, 2017).

Notably, in the later, non-surprising, blocks in Experiment 4, heart rate returned to baseline while blood pressure remained somewhat elevated. This might be indicative of a somewhat vigilant state and part of the anticipation of other unexpected stimuli, since participants knew that other stimuli might appear than the ones they were expecting on the basis of the task instructions. Because the stimuli (pictures of dogs) that subsequently appeared in following blocks were not that surprising anymore, these stimuli did not result in further heart rate reductions (but see Niepel, 2001), though blood pressure remained somewhat elevated, now that it had appeared that the task was somewhat different than originally thought. That participants may have anticipated further changes after an initial surprise also fits with the perspective that unexpectedness is best operationalized by *one* deviant stimulus rather than with multiple deviant stimuli (i.e., as in a repetition-change paradigm rather than an oddball paradigm; see Horstmann, 2015, as also explained in the Introduction).

The heart rate and blood pressure effects in the later non-surprising blocks were not influenced by the valence of the stimulus, which means that we did not find sustained freezing-like responses for negative stimuli in the later non-surprising blocks. This is noteworthy because it has been proposed that a freezing response to threatening stimuli does not habituate (Campbell et al., 1997). One explanation for these blood pressure findings is that it remained somewhat elevated in the positive stimulus valence condition as well, which may explain the lack of difference with the negative valence condition. Another plausible explanation is that the negative stimuli used in this study were not intense enough to result in continuous threat. Future work could include more intensely negative unexpected stimuli (e.g., bodily mutilations) to further test the overlap between unexpectedness and valence with freezing. It should be noted though that this introduces ethical concerns, given that surprise can intensify responses (see e.g., Mellers, Fincher, Drummond, & Bigony, 2013).

Contrary to these block-effects, however, the second-to-second comparisons in the surprising first block did show some effect of valence: Systolic blood pressure seemed more pronounced for negative than for positive surprises. This finding fits our reasoning that after an initial increase in blood pressure (i.e., an effect of surprise in the first seconds), responses start to differentiate depending on the valence of the surprise: Systolic blood pressure remained high for negative stimuli, because of the threatening nature of these stimuli (i.e., a sustained freezing-like effect). This "unfolding" logic is in line with previous research showing similar temporal dynamics in facial expressions after surprise (Noordewier & Breugelmans, 2013; Noordewier & Van Dijk, 2019).

The unfolding logic also fits sequential appraisal theories on emotion, stating that initial novelty/expectedness appraisals are followed by appraisals of the pleasantness of a stimulus (e.g., Scherer, 1999, 2001; see also Delplanque et al., 2009; Hagenaars, Oitzl et al., 2014; Hagenaars, Roelofs et al., 2014). It is noteworthy then, that besides this (marginal) effect on systolic blood pressure, we see no other valence-effects in these second-to-second analyses. It is unlikely that the pleasantness of the stimulus is not appraised at some point during the 10- or 15-sec presentation of the stimulus—also given the differentiation in facial expressions after a couple of seconds, as found in previous work (Noordewier & Breugelmans, 2013; Noordewier & Van Dijk, 2019). A plausible explanation is that the effect of unexpectedness is strong—even to such an extent that it has potentially masked the impact of valence. Importantly, this logic also implies that initial effects of unexpectedness do not necessarily rule out an initial impact of valence. It is possible that unexpectedness and valence may have a so-called "horse-race relationship"—meaning that effects of unexpectedness are initially stronger, which may mask any valence effects.

What may also have contributed to the impact of our stimuli in our studies is that they were unexpected on multiple dimensions, with deviance in the features of the stimulus (e.g., color and contrast) as well as the category of the stimulus (e.g., initially shapes and then a face as

[9] Pooled surprise ratings (initial surprise ratings for Exp 4) and HR difference scores (first 10 seconds of the surprise block [block 1 for Exp 4] minus baseline, pooled across all five studies) were not related ($\rho = .04$, $p = .438$; $N = 346$). In addition, for the second-to-second data we correlated surprise ratings to the lowest heart rate within the epoch, relative to the second before the surprise. This showed a relatively weak positive correlation of $\rho = .13$, $p = .018$.

surprise stimulus), and deviance of the task context ("what does the experimenter expect me to do now?"). One may wonder whether similar effects would be obtained using stimuli that are unexpected on only one dimension and/or within one stimulus-category (e.g., only shapes or faces). We predict that effects would go in the same direction, as previous research also managed to induce surprise using relatively subtle changes in word-color or background patterns in a repetition-change paradigm (Meyer et al., 1991; Reisenzein et al., 2006; Schützwohl, 1998) and using categorically similar stimuli (e.g., a series of neutral faces that changed to a positive or negative face; Noordewier & Van Dijk, 2019). The strength and duration of the effects may be different though. Effects are likely to be weaker when the surprise is less intense. Moreover, the duration of effects may depend on the speed with which one is able to make sense of the unexpectedness: When a surprise is categorically similar to the preceding context, it is easier to categorize, which facilitates sense-making. As a result, surprise will dissipate faster, which increases the chance for valence-effects to occur (for a similar logic, see Noordewier & Van Dijk, 2019). Future studies may test these possible variations in intensity and duration.

Finally, a possible limitation of our approach is that our surprise stimuli were fixed—meaning that the surprising pictures were the same in each study or condition. While this is common in repetition-change studies (e.g., Camras et al., 2002; Meyer et al., 1991; Niepel, 2001; Noordewier & Van Dijk, 2019; Reisenzein et al., 2006; Schützwohl, 1998) and we replicated our findings using different stimuli in different studies (e.g., puppy, dogs, and two versions of George Clooney), a risk of this procedure is stimulus-specific effects. Therefore, future work could replicate the current studies, using random selection stimuli from a pool of, for example, pictures with comparable valence. In addition, while heart rate was consistently measured in all studies, other physiological measures (i.e., movement, blood pressure, temperature) varied between studies. Therefore, it is important to replicate these studies to establish whether the obtained results are robust.

Taken together, our findings thus show that physiological responses to surprise correspond to those that have been associated with orienting and freezing in response to threat. One may wonder whether we in fact showed orienting, freezing, or a combination of both. Orienting and freezing are difficult to disentangle empirically. It is not clear yet whether orienting and freezing are two distinct processes or whether they only differ in quantity and intensity—with orienting responses being shorter and weaker (Campbell et al., 1997; Cook & Turpin, 1997; Hagenaars, Oitzl et al., 2014; Hagenaars, Roelofs et al., 2014; cf. Barry, 1986; Barry & Maltzman, 1985; Germana & Klein, 1968; Turpin, 1986; Vossel & Zimmer, 1989). Orienting is, however, assumed to habituate faster than freezing (Hagenaars, Oitzl et al., 2014; Hagenaars, Roelofs et al., 2014; Roelofs, 2017). It is important to note that also based on the current data, we cannot empirically differentiate between orienting and freezing, as there is no clear reference point for the strength or duration of the effects. If we, however, follow the logic that orienting is the product of novelty, while freezing occurs under the (expectancy) of threat (Hagenaars, Oitzl et al., 2014; Hagenaars, Roelofs et al., 2014; Roelofs, 2017), it makes sense to conclude that unexpectedness results in orienting (see also Reisenzein et al., 2017). Yet, freezing is also plausible, given the threatening aspects of unexpectedness because unexpectedness violates meaning maintenance and cognitive consistency motives (e.g., Clark, 2016; Gawronski & Strack, 2012; Harmon-Jones et al., 2009; Noordewier et al., 2016; Proulx et al., 2012, 2017; Topolinski & Strack, 2015). The confusing nature of an unexpected event may thus be associated with freezing, as people first need to come to terms with the fact that they did not anticipate the event before they can evaluate a surprising stimulus as—for instance—disappointing, sad, enjoyable, or fun.

To conclude, surprise resulted in a lower heart rate and a higher blood pressure, which shows that the physiology of interruption after unexpectedness corresponds to the physiological markers of orienting and freezing in response to threat. Notably, systolic blood pressure was the only physiological parameter that differentiated positive and negative surprise. Future studies may follow up on this finding and explore its use as a distinctive marker for specific arousing states.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

Thanks to Louise Leitsch and Carlijn Clemens for the data collection and input for the study set-up of Experiment 2. Thanks to Kelly Boogerts and Sebastian Simon for their help in developing stimulus materials for Experiment 4 and data collection of Experiments 2b/4. Thanks to Lennart Reddmann and Joyce Snijdewint for their assistance in processing the physiological data. Thanks to Luuk Snijder and Tom Heyman for the assistance with data analyses.

References

Abelson, R. P., Aronson, E., McGuire, W., Newcomb, T., Rosenberg, M., & Tannenbaum, P. (1968). *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally.

Adkin, A. L., & Carpenter, M. G. (2018). New insights on emotional contributions to human postural control. *Frontiers in Neurology, 9*, Article 789.

Algina, J., & Keselman, H. J. (1997). Detecting repeated measures effects with univariate and multivariate statistics. *Psychological Methods, 2*, 208–218.

Azevedo, T. M., Volchan, E., Imbiriba, L. A., Rodrigues, E. C., Oliveira, J. M., Oliveira, L. F., … Vargas, C. D. (2005). A freezing-like posture to pictures of mutilation. *Psychophysiology, 42*, 255–260.

Barry, R. J. (1986). Heart rate deceleration to innocuous stimuli: An index of the orienting response or stimulus registration? *Physiological Psychology, 14*, 42–48.

Barry, R. J., & Maltzman, I. (1985). Heart rate deceleration is not an orienting reflex; heart rate acceleration is not a defensive reflex. *The Pavlovian Journal of Biological Science, 20*, 15–28.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624–652.

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion, 1*, 276–298.

Bradley, M. M., Lang, P. J., & Cuthbert, B. N. (1993). Emotion, novelty, and the startle reflex: Habituation in humans. *Behavioral Neuroscience, 107*, 970–980.

Brosch, T. (2009). Novelty (in appraisal). In D. Sander, & K. R. Scherer (Eds.), *The Oxford companion to emotion and the affective sciences* (p. 285). Oxford: Oxford University Press.

Campbell, B. A., Wood, G., & McBride, T. (1997). Origins of orienting and defensive responses: An evolutionary perspective. In P. J. Lang, R. F. Simons, & M. T. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 41–67). Hillsdale, NJ: Erlbaum.

Camras, L. A., Meng, Z., Ujiie, T., Dharamsi, S., Miyake, K., Oster, H., … Campos, J. (2002). Observing emotion in infants: Facial expression, body behavior, and rater judgments of responses to an expectancy violating event. *Emotion, 2*, 179–193.

Carrive, P. (2000). Conditioned fear to environmental context: Cardiovascular and behavioral components in the rat. *Brain Research, 858*, 440–445.

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. New York: Oxford University Press.

Cook, E., III, & Turpin, G. (1997). Differentiating orienting, startle, and defense responses: The role of affect and its implications for psychopathology. In P. J. Lang, R. F. Simons, & M. T. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 137–164). Hillsdale, NJ: Erlbaum.

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of Student's t-test. *International Review of Social Psychology, 30*, 92–101.

Delplanque, S., Grandjean, D., Chrea, C., Coppin, G., Aymard, L., Cayeux, I., & Scherer, K. R. (2009). Sequential unfolding of novelty and pleasantness appraisals of odors: Evidence from facial electromyography and autonomic reactions. *Emotion, 9*, 316–328.

Dorr, D., Brosschot, J. F., Sollers, J. J., III, & Thayer, J. F. (2007). Damned if you do, damned if you don't: The differential effect of expression and inhibition of anger on cardiovascular recovery in black and white males. *International Journal of Psychophysiology, 66*, 125–134.

Dutra, I. C., Waller, D. A., & Wessel, J. R. (2018). Perceptual surprise improves action stopping by nonselectively suppressing motor activity via a neural mechanism for motor inhibition. *The Journal of Neuroscience, 38*, 1482–1492.

Ekman, P. (2003). *Emotions revealed. Understanding faces and feelings*. London: Weidenfeld & Nicolson.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 572–595). New York: Oxford University Press.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.
Gawronski, B., & Strack, F. (Eds.). (2012). *Cognitive consistency: A fundamental principle in social cognition*. New York: Guilford Press.
Germana, J., & Klein, S. B. (1968). The cardiac component of the orienting response. *Psychophysiology, 4*, 324–328.
Gladwin, T. E., Hashemi, M. M., van Ast, V., & Roelofs, K. (2016). Ready and waiting: Freezing as active action preparation under threat. *Neuroscience Letters, 619*, 182–188.
Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the big five personality domains. *Journal of Research in Personality, 37*, 504–528.
Graham, F. K., & Clifton, R. K. (1966). Heart-rate change as a component of the orienting response. *Psychological Bulletin, 95*, 305–320.
Hagenaars, M. A., Stins, J. F., & Roelofs, K. (2012). Aversive life events enhance human freezing responses. *Journal of Experimental Psychology: General, 141*, 98–105.
Hagenaars, M. A., Oitzl, M., & Roelofs, K. (2014). Updating freeze: Aligning animal and human research. *Neuroscience & Biobehavioral Reviews, 47*, 165–176.
Hagenaars, M. A., Roelofs, K., & Stins, J. F. (2014). Human freezing in response to affective films. *Anxiety, Stress, & Coping, 27*, 27–37.
Harmon-Jones, E., & Mills, J. (1999). *Cognitive dissonance: Progress on a pivotal theory in social psychology*. Washington, DC: American Psychological Association.
Harmon-Jones, E., Amodio, D. M., & Harmon-Jones, C. (2009). Action-based model of dissonance: A review, integration, and expansion of conceptions of cognitive conflict. *Advances in Experimental Social Psychology, 41*, 119–166.
Hashemi, M. M., Gladwin, T. E., Valk, N. M., Zhang, W., Kaldewaij, R., & Ast, V. (2019). Neural dynamics of shooting decisions and the switch from freeze to fight. *Scientific Reports*, 1–10.
Horstman, G. (2006). Latency and duration of the action interruption in surprise. *Cognition and Emotion, 20*, 242–273.
Horstmann, G. (2015). The surprise-attention link: A review. *Annals of the New York Academy of Sciences, 1339*, 106–115.
Horstmann, G. U., & Ansorge, U. (2006). Attentional shifts to rare singletons. *Visual Cognition, 14*, 295–325.
Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: The MIT Press.
Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science, 4*, 1–14.
Lakens, D., McLatchie, N., Isager, P. M., Scheel, A. M., & Dienes, Z. (2020). Improving inferences about null effects with Bayes factors and equivalence tests. *The Journals of Gerontology: Series B, 75*, 45–57.
Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science, 1*, 259–269.
Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: Affect, activation and action. In P. J. Lang, R. F. Simons, & M. T. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 87–135). Hillsdale, NL: Erlbaum.
Lang, P. J., Simons, R. F., & Balaban, M. T. (1997). *Attention and orienting: Sensory and motivational processes*. Hillsdale, NL: Erlbaum.
Leiva, A., Andrés, P., & Parmentier, F. B. R. (2015). When aging does not increase distraction: Evidence from pure auditory and visual oddball tasks. *Journal of Experimental Psychology: Human Perception and Performance, 41*, 1612–1622.
Lojowska, M., Gladwin, T. E., Hermans, E. J., & Roelofs, K. (2015). Freezing promotes perception of coarse visual features. *Journal of Experimental Psychology: General, 144*, 1080–1088.
Mellers, B., Fincher, K., Drummond, C., & Bigony, M. (2013). Surprise: A belief or an emotion? *Progress in Brain Research, 202*, 3–19.
Meyer, W. U., Reisenzein, R., & Schützwohl, A. (1997). Towards a process analysis of emotions: The case of surprise. *Motivation and Emotion, 21*, 251–274.
Miceli, M., & Castelfranchi, C. (2015). *Expectancy and emotion*. Oxford: Oxford University Press.
Muller, K. E., & Peterson, B. L. (1984). Practical methods for computing power in testing the multivariate general linear hypothesis. *Computational Statistics and Data Analysis, 2*, 143–158.
Niepel, M. (2001). Independent manipulation of stimulus change and unexpectedness dissociates indices of the orienting response. *Psychophysiology, 38*, 84–91.
Noordewier, M. K., & Breugelmans, S. M. (2013). On the valence of surprise. *Cognition and Emotion, 27*, 1326–1334.
Noordewier, M. K., & Van Dijk, E. (2019). Surprise: Unfolding of facial expressions. *Cognition and Emotion, 33*, 915–930.
Noordewier, M. K., Topolinski, S., & Van Dijk, E. (2016). The temporal dynamics of surprise. *Social and Personality Psychology Compass, 10*, 136–149.
Noordewier, M. K., Scheepers, D. T., & Hilbert, L. P. (2020). Freezing in response to social threat: A replication. *Psychological Research, 84*, 1890–1896.
Notebaert, W., Houtman, F., Opstal, F. V., Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: An orienting account. *Cognition, 111*, 275–279.
Packer, J. S., Siddle, D. A. T., & Tipp, C. (1989). Restoration of stimulus associability, electrodermal activity, and processing resource allocation. *Biological Psychology, 28*, 105–121.
Parmentier, F. B. R., Vasilev, M. R., & Andrés, P. (2018). Surprise as an explanation to auditory novelty distraction and post-error slowing. *Journal of Experimental Psychology: General, 148*, 199–200.
Pavlov, I. P. (1927). *Conditioned reflexes*. New York: Oxford University Press.
Pollock, A. S., Durward, B. R., Rowe, P. J., & Paul, J. P. (2000). What is balance? *Clinical Rehabilitation, 14*, 402–406.
Proulx, T., Inzlicht, M., & Harmon-Jones, E. (2012). Understanding all inconsistency compensation as a palliative response to violated expectations. *Trends in Cognitive Sciences, 16*, 285–291.
Proulx, T., Sleegers, W., & Tritt, S. M. (2017). The expectancy bias: Expectancy violating faces evoke earlier pupillary dilation than neutral or negative faces. *Journal of Experimental Social Psychology, 70*, 69–79.
Reisenzein, R., Bördgen, S., Holtbernd, T., & Matz, D. (2006). Evidence for strong dissociation between emotion and facial displays: The case of surprise. *Journal of Personality and Social Psychology, 91*, 295–315.
Reisenzein, R., Meyer, W.-U., & Niepel, M. (2012). Surprise. In V. S. Ramachandran (Ed.), *Encyclopedia of human behavior* (2nd ed, pp. 564–570). Oxford, England: Academic Press.
Reisenzein, R., Horstmann, G., & Schützwohl, A. (2017). The cognitive-evolutionary model of surprise: A review of the evidence. *Topics in Cognitive Science, 11*, 50–74.
Reyes del Paso, G. A., Vila, J., & García, A. (1994). Physiological significance of the defense response to intense auditory stimulation: A pharmacological blockade study. *International Journal of Psychophysiology, 17*, 181–187.
Roelofs, K. (2017). Freeze for action: Neurobiological mechanisms in animal and human freezing. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*, Article 20160206.
Roelofs, K., Hagenaars, M. A., & Stins, J. (2010). Facing freeze: Social threat induces bodily freeze in humans. *Psychological Science, 21*, 1575–1581.
Sawada, Y. (2003). Blood pressure and heart rate responses to intrusion on personal space. *Japanese Psychological Research, 45*, 115–121.
Scherer, K. R. (1999). On the sequential nature of appraisal processes: Indirect evidence from a recognition task. *Cognition and Emotion, 13*, 763–793.
Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). New York: Oxford University Press.
Scherer, K. R., Zentner, M. R., & Stern, D. (2004). Beyond surprise: The puzzle of infants' expressive reactions to expectancy violation. *Emotion, 4*, 389–402.
Schröger, E. (1996). A neural mechanism for involuntary attention shifts to changes in auditory stimulation. *Journal of Cognitive Neuroscience, 8*, 527–539.
Schützwohl, A. (1998). Surprise and schema strength. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 1182–1199.
Seery, M. D., Leo, R. J., Lupien, S. P., Kondrak, C. L., & Almonte, J. L. (2013). An upside to adversity? *Psychological Science, 24*, 1181–1189.
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2013). Life after p-hacking. *Paper Presented at the Fourteenth Annual Meeting of the Society for Personality and Social Psychology*.
Sjak-Shie, E. E. (2018). *PhysioData Toolbox (Version 0.4) [Computer software]*. Retrieved from https://PhysioDataToolbox.leidenuniv.nl.
Soriano, C., Fontaine, J. R. J., & Scherer, K. R. (2015). *Review of Cognitive Linguistics, 13*, 436–460.
Steiger, J. H. (2004). Beyond the *F* test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods, 9*, 164–182.
Stins, J. F., Roerdink, M., & Beek, P. J. (2011). To freeze or not to freeze? Affective and cognitive perturbations have markedly different effects on postural control. *Human Movement Science, 30*, 190–202.
Topolinski, S., & Strack, F. (2015). Corrugator activity confirms immediate negative affect in surprise. *Frontiers in Psychology, 6*, 1–8.
Turpin, G. (1986). Effects of stimulus intensity on autonomic responding: The problem of differentiating orienting and defense reflexes. *Psychophysiology, 23*, 1–14.
Vasilev, M. R., Parmentier, F. B. R., Angele, B., & Kirkby, J. A. (2019). Distraction by deviant sounds during reading: An eye-movement study. *Quarterly Journal of Experimental Psychology, 72*, 1863–1875.
Vianna, D. M. L., & Carrive, P. (2005). Changes in cutaneous and body temperature during and after conditioned fear to context in the rat. *European Journal of Neuroscience, 21*, 2505–2512.
Vossel, G., & Zimmer, H. (1989). Heart rate deceleration as an index of the orienting response? *Journal of Psychophysiology, 3*, 111–124.
Walker, P., & Carrive, P. (2003). Role of ventrolateral periaqueductal gray neurons in the behavioral and cardiovascular responses to contextual conditioned fear and poststress recovery. *Neuroscience, 116*, 897–912.
Wessel, J. R. (2017). Perceptual surprise aides inhibitory motor control. *Journal of Experimental Psychology: Human Perception and Performance, 43*, 1585–1593.