Title: Model-based Pavlovian control of attention to threat.

Abbreviated title: Instant attention capture by inferred threat

Authors: D Talmi[a], M Slapkova[a], MJ Wieser[b]

Affiliations:

[a]Division of Neuroscience and experimental psychology, University of Manchester, Oxford Road, M13 9PL, Manchester, UK

[b]Erasmus School of Social and Behavioural Sciences, Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam , Burgemeester Oudlaan 50, 3062 PA Rotterdam, Netherlands

Acknowledgements. We thank N. Chater and C. Frith for the discussions that inspired this paradigm, A. Wilkinson for help with data collection, J. Taylor for help with data analysis, and M. Bauer for helpful comments. DT acknowledges the support of ESRC First Grant ES/I010424/1.

Address for correspondence

Deborah Talmi, Division of neuroscience and experimental psychology, School of Biological Sciences, University of Manchester, Manchester, UK, M139PL.

Telephone: 0161 275 1968

Email: Deborah.Talmi@manchester.ac.uk

Abstract

Signals for reward or punishment attract attention preferentially, a principle termed 'value-modulated attention capture' (VMAC). The mechanisms that govern the allocation of attention resources can be productively described with a terminology that is more often applied to the control of overt behaviours, namely, the distinction between instrumental and Pavlovian control, and between model-free and model-based control. While instrumental control of VMAC can be either model-free or model-based, it is not known whether Pavlovian control of VMAC can be model-based. To decide whether this is possible we measured Steady-State Visual Evoked Potentials (SSVEPs) while 20 healthy adults took part in a novel task. During the learning stage participants underwent aversive threat conditioning with two CSs, one that predicted pain (CS+) and one that predicted safety (CS-). Instructions given prior to the test stage in the task allowed participants to infer whether novel, ambiguous CSs (new CS+/ new CS-) were threatening or safe. Correct inference required combining stored internal representations and new propositional information, the hallmark of model-based control. SSVEP amplitudes quantified the amount of attention allocated to novel CSs on their very first presentation, before they were ever reinforced. We found that SSVEPs were higher for new CS+ than new CS-. Because task design precluded model-free or instrumental control this result demonstrates a model-based Pavlovian control of VMAC. It confirms, in the domain of internal resource allocation, the model-based Pavlovian control of incentive behaviour and underlines the potential transformative role of information as an emotion regulation technique.

Introduction

Regulating established emotional responses upon receipt of novel information can be adaptive. For example, it would be advantageous if, when told that our new eye-drops sting less, we immediately down-regulated the feeling of anxiety about administering this medication, as well as all of the linked cognitive, physiology and behaviours integral to that emotion. If previously we couldn't help but look at that bottle standing ominously on the shelf, we might now be able to ignore it. Indeed, signals for reward or punishment, such as the bottle of eye-drops, attract attention preferentially, a principle termed Value-Modulated Attention Capture (VMAC, Le Pelley et al., 2016). The mechanisms that govern attention allocation can be productively described with a terminology more commonly applied to the control of overt behaviours – the distinction between instrumental/Pavlovian control, and model-free/model-based control (Dayan & Berridge, 2014).

The distinction between instrumental and Pavlovian control has to do with the dependencies between behaviour and outcome (Mackintosh, 1983). In an instrumental learning task outcomes depend on agents' behaviour, so agents act in order to increase utility – either increase the likelihood of reward, or decrease the likelihood of punishment. An example for an instrumental control of VMAC is increased attention to stimuli when told that doing so will be remunerated. By contrast, in Pavlovian conditioning tasks the outcomes are independent of the agent's behaviour. Pavlovian control refers to behaviour that is triggered by stimuli that predict reward or punishment, for example when an animal freezes in response to a tone that predicts a painful foot shock. While Pavlovian control of VMAC is established in the case of rewarding stimuli, there are also demonstrations of the same effect with aversive stimuli (Van Damme, Crombez, Hermans, Koster, & Eccleston, 2006; Wang, Yu, & Zhou, 2013; Wentura, Müller, & Rothermund, 2014). One of the surest ways to be convinced that a particular behaviour is controlled by a Pavlovian, rather than an instrumental, process is when it incurs a loss (Dayan, Niv, Seymour, & D. Daw, 2006). Pavlovian control of VMAC was elegantly demonstrated in an experiment that used an omission schedule, where attending distractors that signalled reward magnitude resulted in the omission of the reward (Le Pelley, Pearson, Griffiths, & Beesley, 2015). Because increased attention to distractors that predicted high (compared to low) reward was never itself rewarded, VMAC in that experiment could not be attributed to instrumental control. Instead, the findings revealed a Pavlovian control of VMAC, with effects that extended to separate tasks (Bucker & Theeuwes, 2017).

The distinction between instrumental and Pavlovian control is orthogonal to the distinction between model-based and model-free control (Daw, Niv, & Dayan, 2005; Dayan & Berridge, 2014). Model-based responses are executed when we infer, based on our model of the environment, that responding in a particular way would maximise our expected utility. Model-based control can be contrasted to model-free control, which depends on the accumulated experience agents have in particular environments. Model-based control allows us to respond flexibly to a volatile, changing environment; model-free control gives us the wisdom of the average experience. The example above for instrumental control, where participants attend stimuli when told they will be rewarded for doing so, is in fact an example for *model-based* instrumental control. This is because propositional information in instructions shapes our model of the environment; we can take up a suggestion or follow an instruction regardless of previous experience in a task (Olsson & Phelps, 2004). Model-based instrumental responses, such as those informed by instructions, can become model-free if they are repeatedly executed and reinforced (Yin & Knowlton, 2006). For example, with repeated pairing between attention to certain stimuli and reward attainment participants acquire a habit to attend to those stimuli. The model-free nature of this behaviour is demonstrated when participants continue to pay preferential attention to these stimuli even when further reinforcement is unlikely (Luque et al., 2017).

Here we ask whether model-based Pavlovian control of VMAC is possible. The opening example demonstrates what Model-based Pavlovian control looks like: the information on the medication revises our model of the environment, yielding new inferences that might instantaneously transform the value of an old Conditioned Stimulus (CS) and, consequently, our attention to it. Animal work has demonstrated model-based Pavlovian control of overt behaviour. For example, placing animals in entirely new states, such as a salt-deprived state, instantly transforms the learned aversive value of a lever that predicts a salty taste (Robinson & Berridge, 2013). But entirely new states are difficult to achieve in humans. Indeed, not much is currently known about Model-based Pavlovian control in humans, although a recent study suggested that a model-based algorithm fitted conditioned threat response in the amygdala better than model-free algorithms (Prévost, McNamee, Jessup, Bossaerts, & O'Doherty, 2013). In relation to VMAC, because Pavlovian VMAC was evident even when participants had plenty of opportunity to learn that their attention allocation was detrimental, and even when they were fully informed about the nature of the omission schedule (Pearson, Donkin, Tran, Most, & Le Pelley, 2015), the Pavlovian control of VMAC may always be model-free (Le Pelley et al., 2016). The same conclusion appears to be supported by findings that instructed extinction did not modulate the classically-conditioned potentiated startle responses (Sevenster, Beckers, & Kindt, 2012). Our aim was to test this contention by using an optimised task to reveal model-based, Pavlovian control of VMAC. Such evidence will confirm, in the domain of internal resource allocation, the distinction between model-free and model-based Pavlovian control of incentive behaviour.

We measured the Steady-State Visual Evoked Potential (SSVEP), a validated neural signal of visual attention (Müller, Teder-Sälejärvi, & Hillyard, 1998; Norcia, Appelbaum, Ales, Cottereau, & Rossion, 2015). In the *conditioning stage* participants passively viewed two Conditioned Stimuli (CSs), which fully predicted a painful electric shock (CS+) or shock omission (CS-). The SSVEP is known to be sensitive to VMAC (Miskovic & Keil, 2013; Wieser, Miskovic, & Keil, 2016), so we expected greater SSVEP amplitudes for the CS+ than the CS-. In the *test stage* two novel CSs were presented once. New_CSs were then constructed such that their value could not be predicted by previous experience. Before the test, participants received propositional information that, when combined with their learned internal representation of the CSs, allowed them to infer the prospective value of new_CSs. Differential attention to new_CSs therefore served as a neural index of model-based control of VMAC.

Because the painful shock outcomes were independent of attention allocation, control of VMAC is Pavlovian by definition, although in the discussion section we consider the important possibility that attention control could have had hidden, internal utility. Importantly, of course, model-based Pavlovian responses can become model-free with repeated experience. A key feature of the paradigm that ensured that the control of VMAC towards the new_CS+ and the new_CS- was model-based was that we could quantify attention to the very first presentation of novel stimuli, before they were ever reinforced (while the entire task was repeated several times new stimuli were used in each repetition). This was possible because of the excellent signal-to-noise ratio of SSVEPs, which has to do with the precision of the signal – a modulation of a known driving frequency, which decreases measurement noise (Norcia et al., 2015). Another important feature was that the design equated the precision with which the value of CSs was represented. Because CSs were all fully predictive (100% of pain or no pain), increased attention to the new_CS+ could not be due to increased uncertainty (Daw et al., 2005). Increased attention to the new_CS+ compared to the new_CS- would have to involve "prospective cognition, formulating and pursuing explicit possible future scenarios based on internal representations of stimuli, situations and environmental circumstances… This knowledge jointly constitutes a model, and supports the computation of value transformations when relevant conditions change" (Dayan and Berridge, 2014, p. 5) – the hallmark of model-based Pavlovian control.

Materials and methods

Participants

Twenty seven undergraduate students from the University of Manchester participated in the study in exchange for course credits. None of the participants reported personal or family history of photic epilepsy, none were taking centrally-acting medication, none had a history of psychiatric or neurological disorders, and all had normal or corrected-to-normal vision. The experiment was approved by the University of Manchester ethics committee. Three participants did not complete the study and one participant did not exhibit an SSVEP signal. Three participants were excluded because they failed the contingency awareness criterion (see below). This resulted in a total of 20 participants (6 males, mean age 19.5, SD=1.15).

Materials

CSs. Stimuli resembled Navon figures (Navon, 1977), in that they were composed of global and local shapes where the outline of the large 'global' shape was created out of smaller 'local' shapes. To create these stimuli we first created 48 unique shapes using Adobe Illustrator, each with a black outline and white filling. These shapes were divided into 24 pairs so that the two shapes in each pair were visually dissimilar (e.g. an arrow and a star). Each pair was used to create a subset of 4 Navon figures, as shown in Figure 1. Two were congruent (e.g. a global arrow made of local arrows, a global star made of local stars), and two incongruent (e.g. a global arrow made of local stars, a global star made of local arrows). In total, the experiment used 24 such four-figure subsets (96 Navon figures). All figures were created and presented in grayscale to minimise differences in colours and luminance. Four-figure subsets were randomly allocated to experimental block. The two congruent figures were randomly allocated to the old_CS+ and old_CS- conditions. There were three types of task blocks, as described below, termed global, local, and control blocks. The new_CS+ in 'Global' blocks used the global attribute of the old_CS+ and the local attribute of the old_CS-. The new_CS+ in 'Local' blocks used the global attribute of the old_CS- and the local attribute of the old_CS+. Two additional four-figure subsets were used for the 2 practice blocks. The figures in practice blocks were created from 4 letters with one four-figure subset consisting of 'H' and 'O', and the other one 'Z' and 'I'.

US. The majority of studies of VMAC use rewarding USs, but there is also evidence for VMAC with aversive outcomes, including pain (e.g. Wang et al., 2013). Here the US was a painful electric stimulation delivered to the back of the right hand via a ring electrode built in-house (Medical Physics, Salford Royal Hospital) attached to a Digitimer DS5 Isolated Bipolar Constant Current Stimulator (Digitimer DS5 2000, Digitimer Ltd., Welwyn Garden City, UK). For reasons of participant safety this stimulator is limited to delivering a maximum of 5V/10mA. To ensure adequate conductance between the electrode and the skin, the back of each participant's hand was prepared with Nuprep Skin Preparation Gel and Ten20 Conductive Paste prior to attaching the electrode. The experiment was implemented using the Psych toolbox on a Matlab (The Mathworks Inc., Natick, MA, USA) platform. The inputs to the DS5 were sent from Matlab through a data acquisition interface (National Instruments, Austin, TX, USA). The behavioural ratings were taken on Microsoft Excel.

Procedure

On arriving at the laboratory, participants were given an information sheet informing them of the justification for the study and of the use of electrical stimulation. After they signed the consent form, participants were fitted with the electroencephalogram (EEG) cap, and sat in a dimly-lit and sound-attenuated room, 90 cm in front of the monitor screen, where an electrode was attached to the dorsum of their right hand. Once the electrode was attached the participants undertook a series of procedures, described below, in the following order: pre-experiment rating of materials, pain calibration, habituation, experimental task, and post-experiment rating of materials.

Pre- and post-experiment rating of likability and contingency . Participants were presented with all of the figures and rated how much they liked each one using a 5-point Likert scale (likability rating task). They then saw all figures again, and guessed, by entering a percentage, how likely each figure was to be followed by a painful stimulation (contingency rating task). The order of the figures in each rating task was randomised for each participant. The likability and contingency rating tasks were repeated at the end of the experiment.

Pain calibration. This procedure ensured that participant could tolerate the stimulation, and that the stimulations were psychologically equivalent across participants. During this procedure participants received a series of stimulations of varying voltage, starting from 0.2V, and incrementing by 0.2V at each step (as the current was constant, this varies the power of the stimulation). Participants rated each stimulation on a scale from 0 – 10 where a score of 0 reflected not being able to feel the stimulation, 3 reflected a stimulation level that was on the threshold of being painful, 7 related to a stimulation that was deemed 'painful but still tolerable' and 10 related to 'unbearable pain'. The scaling procedure was terminated once the participant reported the level of pain as being equivalent to '7' on the scale. This calibration procedure was performed twice to allow for initial habituation/sensitisation to the stimulation. The power levels that induced a rating of '7' on the second run of the calibration procedure were used as US.

Habituation and method of CS presentation*.* Participants passively viewed a randomised list of all of the CS figures. CS figures were displayed at the centre of the screen, a 17'' monitor with a resolution of 1024x768 pixels and a refresh rate of 60Hz. The duration of the presentation of each CS was 3,300ms. That time included 66 on-off cycles in which CS figures were displayed on a uniform white background for 33.3ms ('on') and the screen turned black for 16.6ms ('off'), resulting in a 20Hz flickering display. The inter-trial interval between CSs was 2,500ms, during which the screen was white.

Experimental task. We designed a novel task to reveal model-based Pavlovian control of VMAC. A schematic of the task is shown in Figure 1. The task progressed through two stages - a conditioning stage with 24 trials and a test stage with 4 trials, which are described in detail below. Each trial included the presentation of a CS; when this was a CS+, the trial always terminated with US delivery. Crucially, the logic of the task necessitated an extremely brief test stage that yielded only a single trial for the contrast of interest. This was necessary in order to ensure that VMAC could not be controlled through a model-free process; once new_CSs were reinforced, that reinforcement could inform the value assigned to new_CSs in their second presentation. Therefore, we needed to measure attention to new_CSs upon their first presentation, before they were reinforced, to prevent any possibility that threat value could be informed by the experience of reinforcement. This requirement led us to measure attention using SSVEPs (Müller et al., 1998). SSVEPs have excellent SNR because the driving frequency is known precisely, reducing measurement noise (Norcia et al., 2015; Wieser et al., 2016).

The task was repeated once in each of 24 task blocks. Each task block used novel stimuli, as described in the material section, preventing the transfer of learning across blocks. Each task block lasted 2.5 minutes, with a 5-second break between blocks. Participants practiced the experimental task before it commenced in two practice blocks, using the practice materials described above.

Before the experimental task began participants were given instructions for the experimental task. They were asked to fixate on the fixation cross throughout each block, to observe the figures, and to pay attention to the relationships between the figures and the pain. To encourage compliance, participants were told that their memory for these associations will be tested. This instruction does not privilege memory for the CS+ compared to the CS-, and therefore cannot be responsible for observed threat responses.

Conditioning stage. During the conditioning stage, participants learned that one figure (old_CS+) always predicted pain but another (old_CS-) was safe. CSs were fully predictive of their respective outcomes to reduce any effects of stimulus predictability and of uncertainty, which are tightly intertwined with the effect of value on attention control (Le Pelley et al., 2016). We used previous trial-by-trial dissection of threat effects on the SSVEP signal (Wieser, Miskovic, Rausch, & Keil, 2014) to decide how many conditioning trials were necessary in the conditioning stage. They observed a significant modulation of the SSVEP by aversive reinforcement was observed after 5-10 conditioning trials. Therefore, here we used 12 conditioning trials with each CS. The old_CS+ figure and the old_CS- figure were presented 12 times each, at a random order. The details of how each CS was presented was the same as during habituation, but here, when old_CS+ figures were presented, the US was delivered during the very last cycle, at the offset of the last 'on' screen.

Test stage. After the conditioning stage was completed, participants viewed one of three possible instructions for 10s. In the experimental condition the instruction was the word 'global' or the word 'local'. These words indicated the terms under which the US was to be delivered in the test stage, namely, whether the global or local attribute of the old_CS+ would be reinforced. In the control condition the instruction was a meaningless alphanumeric string, which gave participants no information as to which attribute of the old_CS+ would be reinforced.

Four trials were presented after the instructions. The first two included the presentation of old_Css (their order was randomised), and the last two the presentation of new_CSs (their order was also randomised). New_CSs were the "other" two figures from the same four-figure subset from which the old_CSs were drawn. As can be appreciated from examining Figure 1, each of the new_CSs consisted of one previously-reinforced attribute and one previously-safe attribute. The old_CS+ and the new_CS+ were reinforced; the old_CS- and the new_CS- were not.

Participants did not see the new_CSs before, so without the instructions they could not predict which one would be reinforced. The only way for participants to predict whether the US will follow the new_CS+ or the new_CS- was to infer this from the instructions by drawing on their memory of old_CSs. For example, after the instruction 'local' participants who remembered the local attribute of the old_CS+ could infer that (1) the global attribute of new CSs did not determine whether the US will be delivered or not (2) the US will follow any new CS with the same local attribute as the old_CS+. Taken together, such participants would predict pain after the new_CS+ but not after the new_CS-. New CSs were reinforced in accordance with the instructions, confirming participants' expectations. Old_Css were reinforced in accordance both with the contingencies established during the conditioning stage and the instructions.

While in previous research a newly-acquired conditioned response could be observed within 5-10 trials with each CS (Wieser et al., 2014), the test stage in each task block here only yielded just one trial in each condition. To increase SNR the same structure described thus far – a conditioning stage followed by the test stage - was repeated in each task block. 16 task blocks were allocated to the experimental condition (8 with 'global' and 8 with 'local' instructions) and 8 were allocated to the control condition.

EEG recording and analysis

EEG recording. Continuous EEG recordings were obtained from a 64-channel cap with in-build electrodes (Biosemi Active Two) using the 10-20 configuration system. Data were digitised at a rate of 2048Hz and filtered online between 0.1 and 100 Hz. The recording was referenced online to the Common Mode Sense active electrode. The Driven Right Leg passive electrode was used as ground. The impedance was kept below 40kΩ. Eye movement and blinks were recorded from horizontal and vertical electro-oculogram.

Preprocessing of EEG data. Data were analysed using SPM12. They were converted from their native format and then filtered with three 2nd order Butterworth IIR zero-phase forward and reverse filters: a 1Hz highpass, a 80Hz lowpass, and a 49.50Hz-50.5Hz notch filter to remove mainline noise. Data were downsampled to 200Hz and re-referenced to the average of all electrodes. Eye blinks and saccades were marked on the VEOG and HEOG channels (or Fp1 in two participants) using an automatic algorithm that was thresholded separately for each participant.

Further pre-processing was conducted for the purpose of complex demodulation (see below). Data were segmented between -600ms before the onset of CSs to 3250ms after CS onset (Just before the offset of the CS/US onset, 3300ms after CS onset). Segments where the following artefacts were present on occipital channels (Oz, POz, O1, O2, O3, O4) were rejected: jumps greater than 150 μV; peak-to-peak differences greater than 250 μV; flat segments. Channels where more than 20% of the trials were rejected were excluded from analysis. This left, on average, 282.62 learning trials and 7.83 test trials with each CS in each condition. Artefacts associated with eye blinks and saccades were corrected using the singular value decomposition (SVD) technique implemented in SPM12 which captures eye artefacts and removes the associated component.

Complex demodulation. Threat effects were operationalised as an increased response to the CS+ compared to CS-. Previous work suggested that threat effect are more pronounced later in the presentation of the CS, because the threat response is greater when threat is imminent, and the perceptual processing of predictive sensory features of the CS is enhanced only when the US is imminent (Miskovic & Keil, 2012; Paterson & Neufeld, 1987). Complex demodulation was therefore carried out to determine exactly when threat effects were present.

Complex demodulation was conducted using SPM12 on the entire segment, with the multitaper method, a hanning window, and a resolution of 1Hz. Data in each condition were averaged using robust averaging, a method that down-weights outliers (Litvak et al., 2011). The signal from electrodes Oz and POz was extracted around the driving frequency (19-21Hz). These data were averaged across all 12 trials in each condition and the 24 blocks of the experimental task (288 trials for each CS). In agreement with Miskovic and Keil (2012), threat effects were greatest during the second half of the presentation of the CS (Figure 2). An examination of the topographies associated with threat supported our selection of electrodes of analysis. We used these results to constrain our spectral analysis.

Spectral analysis. Based on the results of the complex demodulation step, spectral analysis was conducted using the spatiotemporal window of 1500-3000ms from CS onset, at Oz and POz, between 19-21 Hz. Data from each trial was Fourier-transformed using the FFT function in Matlab. Data from the conditioning stage were binned (3 trials per bin) and the threat response was examined across that stage, averaging over the 24 task blocks. The main hypothesis concerned the test stage, where power was limited by the availability of only very few trials in each condition. Therefore, we averaged the 16 blocks where participants were presented with new_CS+ in the experimental condition (collapsing across the 8 blocks in the 'global' and the 8 blocks in the 'local' conditions), the 16 blocks where participants were presented with new_CS- (again collapsing across the 16 'global' and 'local' blocks), and the 16 blocks where participants were presented with control new_CSs (collapsing across all ambiguous shapes in the control condition). To test the main hypothesis, spectral data were averaged for each of these three conditions, new_CS+, new_CS- and control new_CS, and the average data around the driving frequency (19-21Hz) were extracted for each individual. Across participants, the data significantly diverged from normality, according to the Shapiro-Wilk test, and were therefore log-transformed.

Results

Behavioural results

Sample selection. Behavioural and EEG data were first checked to verify the presence of a conditioned threat response. We excluded participants who may have disengaged from the task based on their 'learning score', computed as the increased contingency ratings given to old_CS+ stimuli after the experiment compared to the pre-experiment rating given to the same stimuli. Increased ratings must be based on learning that occurred during the experiment, and therefore reflects at least a minimal level of engagement. The learning score purposefully ignores the magnitude of the conditioned response, which could be computed as differential ratings of the old_CS+ and old_CS-, in order not to bias the selection of the sample. To be included participants had to show a numerical (above 0) increase in ratings. Based on this criterion, three participants were excluded from analysis, leaving a sample of N=20.

Manipulation check. Contingency and liking ratings for the stimuli were entered into two separate 3-way repeated-measures ANOVAs with the factors time (pre-experiment, post-experiment), threat (CS+, CS-), and status (old, new). In both the analysis of contingency ratings and in the (separate) analysis of liking ratings the 3-way interaction between time (pre-experiment, post-experiment), threat (CS+, CS-), and status (old, new) was significant (contingency: F(1,19)=34.95, P<.001, partial eta=.65; liking F(1,19)=4.79, p=.04, partial eta=.20). We unpacked this interaction by examining the old and new CSs separately. The 2-way interaction was significant for old CSs (contingency: F(1,19)=59.98, p<.001, partial eta=.76; liking: F(1,19)=17.89, p<.001, partial eta=.48) as well as for new CSs (contingency: F(1,19)=4.65, p=.04, partial eta =.20; liking: F(1,19)=9.16, p=.007, partial eta=.32). The results showed that participants expected the US more frequently following the CS+ and liked the CS+ less than the CS-, and that these effects were stronger for old_CSs than for new_CSs, probably because old_CSs were repeated multiple times, but new_CSs were only presented once.

EEG results

Manipulation check. Threat effects on the SSVEP during the conditioning stage were analysed with a 2 (threat) x 4 (trial bins) repeated-measures ANOVA. The successful manipulation of threat in this experiment was evident in differential responses to the CS+ and CS- during the conditioning stage (Figure 2), F(1,19)=5.78, p=0.045, partial η^2=0.19. The magnitude of threat effects remained consistent across the 4 trial bins, F<1 (Figure 2, bottom right). The main effect of binned trials, F(3,57)=5.85, p=.001, partial η^2=0.24, was also significant, due to an overall decrease in overall SSVEP amplitude across the trials of the conditioning stage in each block. To verify that threat responses were also obtained in the test stage, responses to old_CSs in the control condition, where participants had no reason to make any model-based inferences, were analysed with a one-tailed paired t-test, contrasting old_CS+ and old_CS. We observed differential responses to the old_CS+ and old_CS-, t(19)=2.27, p=.016 (one-tailed), Cohen's d = 0.50. Because only the data from the conditioning stage were pre-selected, the data from the test stage provide a useful confirmation.

Main hypothesis. Our main hypothesis was that responses to the new_CS+ would be greater than responses to the new_CS-. The hypothesis was evaluated with a one-tailed paired t-test comparing SSVEPs to the new_CS+ and new_CS-. As predicted, SSVEP amplitudes were higher during the presentation of the new_CS+ compared to the new_CS-, t(19)=2.45, p=.012, Cohen's d = 0.55. Additional 2-tailed t-tests, controlled for multiple comparisons with a p-value of 0.25, compared each of these SSVEPS to the SSVEP elicited by ambiguous new_CSs in the control condition. SSVEP amplitudes were equivalent during the presentation of the new_CS+ and the new_CSs in the control condition, t<1, suggesting that participants experienced ambiguous figures as threatening when they could not use the instructions to disambiguate them. This interpretation is supported by a significant difference between the new_CSs in the control condition and the new_CS-, t(19)=3.46, p=0.003, Cohen's d = 0.77, suggesting that participants experienced less threat when they knew that pain is unlikely.

Discussion

During the learning stage of our Pavlovian conditioning task, we observed an increase in the amplitude of the SSVEP signal towards stimuli with learned aversive value. These results are not surprising, given much evidence that the valuation system can control attention allocation (Le Pelley et al., 2016). But they are particularly important in confirming previous findings of Pavlovian control of VMAC towards stimuli with aversive value (Van Damme et al., 2006; Wang et al., 2013; Wentura et al., 2014), which is less established than Pavlovian control of VMAC towards reward, or the effect of value on instrumental control of VMAC. Uniquely, these results also suggest a way to observe the neural evolution of VMAC across the early part of the learning process, and even in the first trial, something that has not been possible using other neuroimaging investigation (Olsson & Phelps, 2007; Phelps et al., 2001) or in animal models (Balcarras, Ardid, Kaping, Everling, & Womelsdorf, 2016).

Our key result was that SSVEP amplitudes were larger when participants were presented with a new shape that they inferred predicted physical pain, the new_CS+, compared to the new_CS-, a new shape that they inferred predicted safety. Because the amplitude of SSVEPs is known to be higher for attended stimuli compared to unattended ones (Müller et al., 1998; Muller et al., 1998), our findings suggest that more attention was allocated to the new_CS+ compared to the new_CS-. New CSs were constructed such that previous experience would render them equally ‘threatening’ and ‘safe’. It should therefore be difficult for a model-free algorithm to implement differential responses to these CSs. Their threat value could only be disambiguated through an inference based on a combination of stored knowledge and propositional information, implicating model-based control. In the control condition the propositional information was omitted, so that model-based inferences was prevented. Therefore, the threat value of new_CSs was ambiguous, and participants could not predict which one would be followed by pain. These ambiguous new_CSs in the control condition attracted increased attention. This result, which suggests orienting towards ambiguous stimuli, accords with previous findings, where instructed, ambiguous, novel CSs gave rise to increased physiological arousal and increased activation in the amygdala and the insula (Phelps et al., 2001). Increased attention to the ambiguous new_CSs in the control condition could stem either from a model-based or a model-free process, such as habitual orienting towards stimuli with uncertain outcomes. Indeed, model-based and model-free control can operate in parallel during reinforcement learning (Luque et al., 2017). This distinction resembles the one between cognitive and emotional control of Pavlovian responses (Sevenster et al., 2012).

The paradigm ruled out instrumental control, because it was entirely passive, and participants could not benefit from allocating differential attention to specific CSs. Indeed, participants were told explicitly, and also knew through experience across the 24 blocks of the task, that the stimulation levels were pre-determined and that they could therefore not influence it; this was made particularly salient through the instructions at the beginning of the test stage. Participants had no reason to allocate more attention to the new_CS+ in order to increase success in the post-experiment rating task, because they have already completed it once before they started the experiment, and knew, therefore, that performance would benefit equally from attending all of the stimuli. Based on these considerations, increased SSVEPs to new_CS+ could only be due to model-based Pavlovian control of VMAC; other controllers were ruled out by design.

Although we defined the control of VMAC here as Pavlovian, because it was observed in a classical conditioning (Pavlovian) task, where pain outcomes were independent participants’ behaviour or how they allocated their attention, it is possible, in principle, that this seemingly Pavlovian response incurred some *internal* benefit. Specifically, paying extra attention to threatening CSs here may have decreased subjective pain. It is important to consider this possibility because the experience of pain, and its expected negative value, are not true reflections of the objective empirical reality. Instead, they are strongly modulated by pain expectations (Atlas & Wager, 2012; Berns et al., 2006; Morley, Vlaeyen, & Schrooten, 2012; Paterson & Neufeld, 1987; Tabor, Thacker, Moseley, & Körding, 2017; Vlaev, Seymour, Dolan, & Chater, 2009), an effect implemented through endogenous analgesic mechanisms (Anchisi & Zanon, 2015; Tracey, 2010; Wager et al., 2004).

Yet closer scrutiny suggests that it is unlikely that attending the new_CS+ triggered endogenous analgesia. While participants needed to attend experimental new_CSs to decipher exactly which one predicted pain and which one did not, this need not result in *differential* attention to the two new_CSs. Because the US was completely predictable, always of the same intensity, and presented at the same time, attending the new_CS+ is unlikely to have altered pain expectations meaningfully (although extra attention could increase the precision of expectations; Kok, Rahnev, Jehee, Lau, & de Lange, 2012). In fact, the opposite is the case: much evidence suggests that distraction is an effective pain-coping strategy (Buhle, Stevens, Friedman, & Wager, 2012; Eccleston, 1995; Sharar et al., 2016; Weiss, Dahlquist, & Wohlheiter, 2011). Finally, to establish that a response is controlled by a Pavlovian process researchers have utilised omission schedules, where the response incurs a cost (Le Pelley et al., 2016; Mackintosh, 1983). This is impossible in the case of model-based Pavlovian behaviours, because the eventual costs should, by definition, alter the world-model that inspired the responses in the first place. In summary, although the feeling of pain is malleable, we can presently think of no a-priori reason that attending the new_CS+ would be advantageous. The intuitive sense that we would all *want* to look – that we might not be able to attend anything else – is perhaps simply the reflection of Pavlovian misbehaviour (Dayan et al., 2006).

This is the first demonstration that propositional information – a form of instruction – influences Pavlovian control of attention allocation. This demonstration is particularly important because a previous experiment suggested that in Pavlovian tasks attention allocation obeys associative learning principles, and is immune to propositional knowledge. In Moratti and Keil's study (Moratti & Keil, 2009) the SSVEF during CS presentation (steady-state visual evoked field, measured with MEG) increased with increased number of sequentially reinforced CSs, not with increased US expectancy, which increased when previous CSs have consistently not been reinforced. In the terminology used here, SSVEFs were driven by model-free, not model-based, Pavlovian control. Indeed, other studies have also observed that similar 'gambler's fallacy'-like paradigms give rise to conditioned responses that are based on model-free, not model-based value (Clark, Manns, & Squire, 2001; Perruchet, 1985). In the domain of attention allocation this result is particularly intriguing, because it contradicts evidence that expectancy influences visual attention (Downing, 1988). Taken together, it appears that model-free control dominates attention allocation in the gambler's fallacy paradigm, at least when using a delayed conditioning procedure (Clark et al., 2001), while in other paradigms – including those using delayed conditioning, as we did here – model-based Pavlovian control of VMAC is possible. More generally, our result adds to other demonstrations where instructions about Pavlovian contingencies encourage responses that mimic the effect of associative learning through experience (reviewed in Mitchell, De Houwer, & Lovibond, 2009). It is important to note, however, that while attentional responses were affected here, it is possible that other classically-conditioned responses were not. In particular, because attention allocation was affected in the first trial it bears stronger resemblance to US-expectancy ratings and to classically-conditioned skin conductance responses, which were immediately influenced by instructed extinction, than to potentiated startle, which was not (Sevenster et al., 2012). Further research is required to examine this dissociation using our paradigm. Further research is also needed to verify whether instructed extinction, like instructed threat, also alters VMAC instantaneously.

We now turn to the question of how Pavlovian model-based control occurred here. Pavlovian model-based control can be demonstrated in a number of ways, including, prominently, by placing animals in entirely new states, such as a salt-deprived state that instantly transforms the learned aversive value of a lever that predicts a salty taste (Robinson & Berridge, 2013). Yet doing so would be extremely challenging to achieve in an experiment with human participants. Here, following the experimental instructions, participants may have constructed a model of the new CSs and their predictive value by combining the new propositional information and stored internal representations. It is possible that while they viewed the instructions, participants recalled old_CSs and generalised their aversive value to imagined new_CSs. Dayan and Berridge (2014) discuss such recall and revaluation processes as mechanisms that allow model-based Pavlovian control. It is also possible that participants volitionally inhibited the representation of the global or local dimensions of memorised learned CSs as well as actual new_CSs, to support generalisation from the learning to the test stage.

By using neural measures to index model-based control we move a little closer to understanding how propositional information is implemented at the level of the neurobiological mechanism. Increased SSVEPs during the presentation of threatening stimuli is thought to be driven by re-entrant connections from the amygdala, ACC and OFC, which amplify the processing of adaptive information (Miskovic & Keil, 2012). Repeated pairing between a stimulus and a painful stimulation can change the neural representation of the pain-predicting stimulus. For example, repeated pairing between a tone and a painful shock change the tuning frequency of neurons that encode these tones, and stimulation of the amygdala is sufficient to produce this effect (Chavez, McGaugh, & Weinberger, 2013). Here, however, such a process could not occur because attention modulation was manifested before the reinforcement itself. A meta-analysis of studies of instructed fear found that the dorsomedial prefrontal cortex is uniquely associated with a conscious appraisal process (Mechias, Etkin, & Kalisch, 2010). Similarly, the same region has been shown to dynamically modulate model-free valuation in the OFC, striatum, and hippocampus (Li, Delgado, & Phelps, 2011). It is therefore likely that increased response to the new_CS+ was due to projections from the dorsomedial prefrontal cortex to the OFC and ACC, regions that are strongly connected to the amygdala and able to modulate its activity (Lee, Heller, van Reekum, Nelson, & Davidson, 2012; Schiller & Delgado, 2010), with downstream re-entrant effects in the visual cortex.

The constrained data yield of the paradigm should be acknowledged as a limitation of this study. While the effect sizes in all of the statistical tests were all of a ‘medium’ size, according to Cohen’s classification (Cohen, 1988), the study should be replicated in order to increase confidence in this novel result. For the same reason, we could not explore the influence of ‘dimension’ (global or local) in the results we obtained, because this would have halved the number of trials that we could analyse.

At the experiential level, increased attention to the new_CS+ suggests that the information given to participants worked as an emotion regulation technique – it rendered ambiguous stimuli instantly threatening. Drawing the connection between model-based and model-free control, on the one hand, and cognitive and emotional control, on the other (Sevenster et al., 2012) can help the quest to ground emotion regulation and behaviour change techniques more tightly in computational theories (Etkin, Büchel, & Gross, 2016).

In closing, we consider whether model-based, Pavlovian control of VMAC is adaptive. VMAC could enhance the encoding of CSs, strengthen their memory traces, and thus facilitate optimal decisions when the opportunity arises to act on the same stimuli (e.g. escape). More broadly, prediction error minimisation – something that is considered globally adaptive (Pezzulo, Rigoli, & Friston, 2015) - may be facilitated if the excellent encoding of valued stimuli increases the precision of the model we have of the world around us.

References

Anchisi, D., & Zanon, M. (2015). A Bayesian perspective on sensory and cognitive integration in pain perception and placebo analgesia. *PLoS ONE*. http://doi.org/10.1371/journal.pone.0117270

Atlas, L. Y., & Wager, T. D. (2012). How expectations shape pain. *Neuroscience Letters*. http://doi.org/10.1016/j.neulet.2012.03.039

Balcarras, M., Ardid, S., Kaping, D., Everling, S., & Womelsdorf, T. (2016). Attentional Selection Can Be Predicted by Reinforcement Learning of Task-relevant Stimulus Features Weighted by Value-independent Stickiness. *Journal of Cognitive Neuroscience*, *28*(2), 333–349. http://doi.org/10.1162/jocn_a_00894

Berns, G. S., Chappelow, J., Cekic, M., Zink, C. F., Pagnoni, G., & Martin-Skurski, M. E. (2006). Neurobiological substrates of dread. *Science (New York, N.Y.)*, *312*(5774), 754–8. http://doi.org/10.1126/science.1123721

Bucker, B., & Theeuwes, J. (2017). Pavlovian reward learning underlies value driven attentional capture. *Attention, Perception, & Psychophysics*, *79*(2), 415–428. http://doi.org/10.3758/s13414-016-1241-1

Buhle, J. T., Stevens, B. L., Friedman, J. J., & Wager, T. D. (2012). Distraction and Placebo. *Psychological Science*, *23*(3), 246–253. http://doi.org/10.1177/0956797611427919

Chavez, C. M., McGaugh, J. L., & Weinberger, N. M. (2013). Activation of the basolateral amygdala induces long-term enhancement of specific memory representations in the cerebral cortex. *Neurobiology of Learning and Memory*, *101*, 8–18. http://doi.org/10.1016/j.nlm.2012.12.013

Clark, R. E., Manns, J. R., & Squire, L. R. (2001). Trace and delay eyeblink conditioning: contrasting phenomena of declarative and nondeclarative memory. *Psychological Science : A Journal of the American Psychological Society / APS*, *12*(4), 304–308. http://doi.org/10.1111/1467-9280.00356

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates, Hillsdale, NJ. http://doi.org/10.1234/12345678

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711. http://doi.org/10.1038/nn1560

Dayan, P., & Berridge, K. C. (2014). Model-based and model-free Pavlovian reward learning: revaluation, revision, and revelation. *Cognitive, Affective & Behavioral Neuroscience*, *14*(2), 473–92. http://doi.org/10.3758/s13415-014-0277-8

Dayan, P., Niv, Y., Seymour, B., & D. Daw, N. (2006). The misbehavior of value and the discipline of the will. *Neural Networks*, *19*(8), 1153–1160. http://doi.org/10.1016/j.neunet.2006.03.002

Downing, C. J. (1988). Expectancy and visual-spatial attention: Effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(2), 188–202. http://doi.org/10.1037/0096-1523.14.2.188

Eccleston, C. (1995). The attentional control of pain: methodological and theoretical concerns. *Pain*, *63*(1), 3–10. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8577487

Etkin, A., Büchel, C., & Gross, J. J. (2016). Emotion regulation involves both model-based and model-free processes. *Nature Reviews Neuroscience*, *17*(8), 532–532. http://doi.org/10.1038/nrn.2016.79

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & de Lange, F. P. (2012). Attention Reverses the Effect of Prediction in Silencing Sensory Signals. *Cerebral Cortex*, *22*(9), 2197–2206. http://doi.org/10.1093/cercor/bhr310

Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N., Wills, A. J., & Le Pelley, M. (2016). Attention and associative learning in humans: An integrative review. *Psychological Bulletin*, *142*(10), 1111–1140. http://doi.org/10.1037/bul0000064

Le Pelley, M. E., Pearson, D., Griffiths, O., & Beesley, T. (2015). When Goals Conflict With Values: Counterproductive Attentional and Oculomotor Capture by Reward-Related Stimuli Predictiveness-Driven Attentional Capture. *Experimental Psychology*, *144*(1), 158–171. http://doi.org/10.1037/xge0000037

Lee, H., Heller, A. S., van Reekum, C. M., Nelson, B., & Davidson, R. J. (2012). Amygdala-prefrontal coupling underlies individual differences in emotion regulation. *NeuroImage*, *62*(3), 1575–1581. http://doi.org/10.1016/j.neuroimage.2012.05.044

Li, J., Delgado, M. R., & Phelps, E. A. (2011). How instructed knowledge modulates the neural systems of reward learning. *Proceedings of the National Academy of Sciences*, *108*(1), 55–60. http://doi.org/10.1073/pnas.1014938108

Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., … Friston, K. (2011). EEG and MEG data analysis in SPM8. *Computational Intelligence and Neuroscience*, *2011*, 852961. http://doi.org/10.1155/2011/852961

Luque, D., Beesley, T., Morris, R. W., Jack, B. N., Griffiths, O., Whitford, T. J., & Le Pelley, M. E. (2017). Goal-Directed and Habit-Like Modulations of Stimulus Processing during Reinforcement Learning. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *37*(11), 3009–3017. http://doi.org/10.1523/JNEUROSCI.3205-16.2017

Mackintosh, N. J. (1983). *Conditioning and associative learning*. Clarendon Press. Retrieved from https://books.google.co.uk/books/about/Conditioning_and_Associative_Learning.html?id=a8x9AAAAMAAJ&redir_esc=y&hl=en

Mechias, M.-L., Etkin, A., & Kalisch, R. (2010). A meta-analysis of instructed fear studies: Implications for conscious appraisal of threat. *NeuroImage*, *49*(2), 1760–1768. http://doi.org/10.1016/J.NEUROIMAGE.2009.09.040

Miskovic, V., & Keil, A. (2012). Acquired fears reflected in cortical sensory processing: a review of electrophysiological studies of human classical conditioning. *Psychophysiology*, *49*(9), 1230–41. http://doi.org/10.1111/j.1469-8986.2012.01398.x

Miskovic, V., & Keil, A. (2013). Perceiving threat in the face of safety: excitation and inhibition of conditioned fear in human visual cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(1), 72–8. http://doi.org/10.1523/JNEUROSCI.3692-12.2013

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*. http://doi.org/10.1017/S0140525X09000855

Moratti, S., & Keil, A. (2009). Not What You Expect: Experience but not Expectancy Predicts Conditioned Responses in Human Visual and Supplementary Cortex. *Cerebral Cortex December*, *19*, 2803–2809. http://doi.org/10.1093/cercor/bhp052

Morley, S., Vlaeyen, J. W., & Schrooten, M. G. S. (2012). Psychological interventions for chronic pain: reviewed within the context of goal pursuit. *Pain Management*, *2*(March), 1–10. http://doi.org/http://dx.doi.org/10.2217/pmt.12.2

Muller, M. M., Picton, T. W., Valdes-Sosa, P., Riera, P., Teder-Salejarvi, W., & Hillyard, S. A. (1998). Effects of spatial selective attention on the steady-state visual evoked potential in the 20-28 Hz range. *Cognitive Brain Research*, *6*, 249–261.

Müller, M. M., Teder-Sälejärvi, W., & Hillyard, S. A. (1998). The time course of cortical facilitation during cued shifts of spatial attention. *Nature Neuroscience*, *1*(7), 631–634. http://doi.org/10.1038/2865

Navon, D. (1977). Forest before trees: the precedence of global features in visual perception. *COGNITIVE PSYCHOLOGY*, *9*, 353–383.

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, *15*(6), 4. http://doi.org/10.1167/15.6.4

Olsson, A., & Phelps, E. A. (2004). Learned fear of "unseen" faces after pavlovian, observational, and instructed fear. *Psychological Science*, *15*(12), 822–828. http://doi.org/10.1111/j.0956-7976.2004.00762.x

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience*, *10*(9), 1095–1102. http://doi.org/10.1038/nn1968

Paterson, R. J., & Neufeld, R. W. J. (1987). Clear Danger: Situational Determinants of the Appraisal of Threat. *Psychological Bulletin*. http://doi.org/10.1037/0033-2909.101.3.404

Pearson, D., Donkin, C., Tran, S. C., Most, S. B., & Le Pelley, M. E. (2015). Cognitive control and counterproductive oculomotor capture by reward-related stimuli. *Visual Cognition*, *6285*(May 2015), 1–26. http://doi.org/10.1080/13506285.2014.994252

Perruchet, P. (1985). A pitfall for the expectancy theory of human eyelid conditioning. *The Pavlovian Journal of Biological Science : Official Journal of the Pavlovian*, *20*(4), 163–170. http://doi.org/10.1007/BF03003653

Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, *134*, 17–35. http://doi.org/10.1016/J.PNEUROBIO.2015.09.001

Phelps, E. A., O'Connor, K. J., Gatenby, J. C., Gore, J. C., Grillon, C., & Davis, M. (2001). Activation of the left amygdala to a cognitive representation of fear. *Nature Neuroscience*, *4*(4), 437–441. http://doi.org/10.1038/86110

Prévost, C., McNamee, D., Jessup, R. K., Bossaerts, P., & O'Doherty, J. P. (2013). Evidence for model-based computations in the human amygdala during Pavlovian conditioning. *PLoS Computational Biology*, *9*(2), e1002918. http://doi.org/10.1371/journal.pcbi.1002918

Robinson, M. J. F., & Berridge, K. C. (2013). Instant transformation of learned repulsion into motivational "wanting." *Current Biology*, *23*(4), 282–289. http://doi.org/10.1016/j.cub.2013.01.016

Schiller, D., & Delgado, M. R. (2010). Overlapping neural systems mediating extinction, reversal and regulation of fear. *Trends in Cognitive Sciences*. http://doi.org/10.1016/j.tics.2010.04.002

Sevenster, D., Beckers, T., & Kindt, M. (2012). Instructed extinction differentially affects the emotional and cognitive expression of associative fear memory. *Psychophysiology*, *49*(10), 1426–1435. http://doi.org/10.1111/j.1469-8986.2012.01450.x

Sharar, S. R., Alamdari, A., Hoffer, C., Hoffman, H. G., Jensen, M. P., & Patterson, D. R. (2016). Circumplex Model of Affect: A Measure of Pleasure and Arousal During Virtual Reality Distraction Analgesia. *Games for Health Journal*, *5*(3), 197–202. http://doi.org/10.1089/g4h.2015.0046

Tabor, A., Thacker, M. A., Moseley, G. L., & Körding, K. P. (2017). Pain: A Statistical Account. *PLoS Computational Biology*. http://doi.org/10.1371/journal.pcbi.1005142

Tracey, I. (2010). Getting the pain you expect: Mechanisms of placebo, nocebo and reappraisal effects in humans. *Nature Medicine*. http://doi.org/10.1038/nm.2229

Van Damme, S., Crombez, G., Hermans, D., Koster, E. H. W., & Eccleston, C. (2006). The role of extinction and reinstatement in attentional bias to threat: A conditioning approach. *Behaviour Research and Therapy*, *44*(11), 1555–1563. http://doi.org/10.1016/j.brat.2005.11.008

Vlaev, I., Seymour, B., Dolan, R. J., & Chater, N. (2009). The price of pain and the value of suffering. *Psychological Science*, *20*(3), 309–317. http://doi.org/10.1111/j.1467-9280.2009.02304.x

Wager, T. D., Rilling, J. K., Smith, E. E., Sokolik, A., Casey, K. L., Davidson, R. J., … Cohen, J. D. (2004). Placebo-induced changes in FMRI in the anticipation and experience of pain. *Science (New York, N.Y.)*, *303*(5661), 1162–1167.

http://doi.org/10.1126/science.1093065

Wang, L., Yu, H., & Zhou, X. (2013). Interaction between value and perceptual salience in value-driven attentional capture. *Journal of Vision*, *13*(2013), 1–13. http://doi.org/10.1167/13.3.5.doi

Weiss, K. E., Dahlquist, L. M., & Wohlheiter, K. (2011). The Effects of Interactive and Passive Distraction on Cold Pressor Pain in Preschool-aged Children. *Journal of Pediatric Psychology*, *36*(7), 816–826. http://doi.org/10.1093/jpepsy/jsq125

Wentura, D., Müller, P., & Rothermund, K. (2014). Attentional capture by evaluative stimuli: Gain- and loss-connoting colors boost the additional-singleton effect. *Psychonomic Bulletin and Review*, *21*(3), 701–707. http://doi.org/10.3758/s13423-013-0531-z

Wieser, M. J., Miskovic, V., & Keil, A. (2016). Steady-state visual evoked potentials as a research tool in social affective neuroscience. *Psychophysiology*. http://doi.org/10.1111/psyp.12768

Wieser, M. J., Miskovic, V., Rausch, S., & Keil, A. (2014). Different time course of visuocortical signal changes to fear-conditioned faces with direct or averted gaze: A ssVEP study with single-trial analysis. *Neuropsychologia*, *62*(1), 101–110. http://doi.org/10.1016/j.neuropsychologia.2014.07.009

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews. Neuroscience*, *7*(6), 464–76. http://doi.org/10.1038/nrn1919

FIGURE 1. EXPERIMENTAL TASK.

Left: timeline of a single block in the experimental task, including the conditioning, instructions, and test stages. **Right**: CSs in this block were drawn from the 4-figure subset crossing the local and global dimensions of the star and arrow shapes.

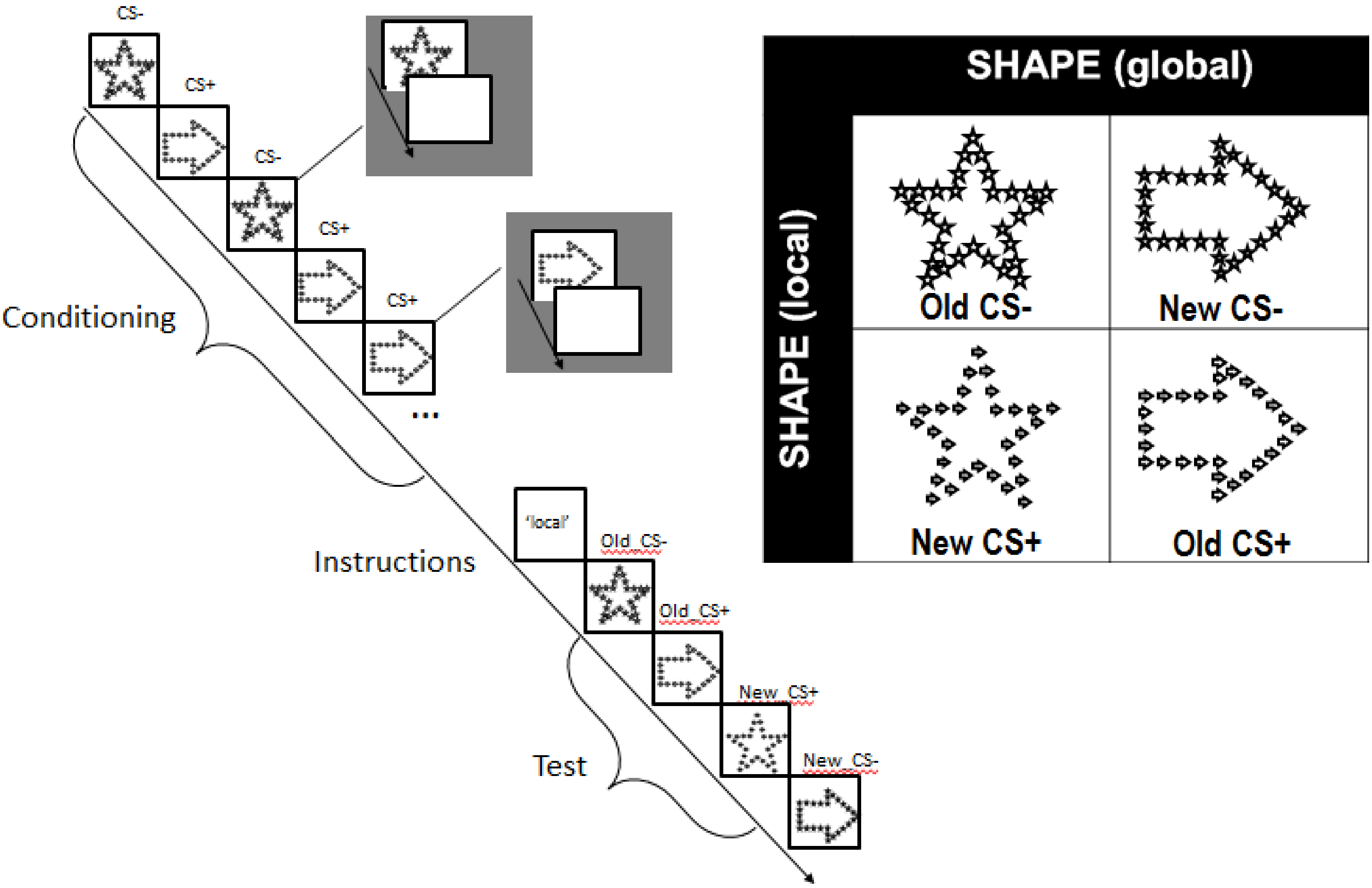


FIGURE 2. THREAT EFFECTS IN THE CONDITIONING STAGE

Top left. SSVEP signal amplitudes in the conditioning stage for CS+ and CS-, extracted from occipital electrodes Oz and POz, and averaged across 19-21Hz. Shaded areas plot the standard error.

Top right. Threat effects in occipital electrodes Oz and POz, operationalised as the difference between CS+ > CS-, are plotted as a function of time from CS onset, averaged across 19-21Hz. The time window 1500-300ms (shaded grey) was used in all analysis of threat effects. The topology inset shows that the threat effects across that time window.

Bottom left. The result of the spectral analysis of signal in the conditioning stage for CS+ and CS-, extracted from occipital electrodes Oz and POz at the 1500-3000ms time window, showing that threat modulated the 20Hz SSVEP signal. **Insert:** The magnitude of the 20Hz threat effect, operationalised as the difference between CS+ > CS-, collapsed across time, showing the variability of this effect across participants. The red line indicated the mean; the box indicates the inter-quartile range.

Bottom right. The evolution of threat effects across the conditioning stage. SSVEP amplitude is plotted as a function of binned trials (3 trials in each bin). Shaded areas represent standard error.

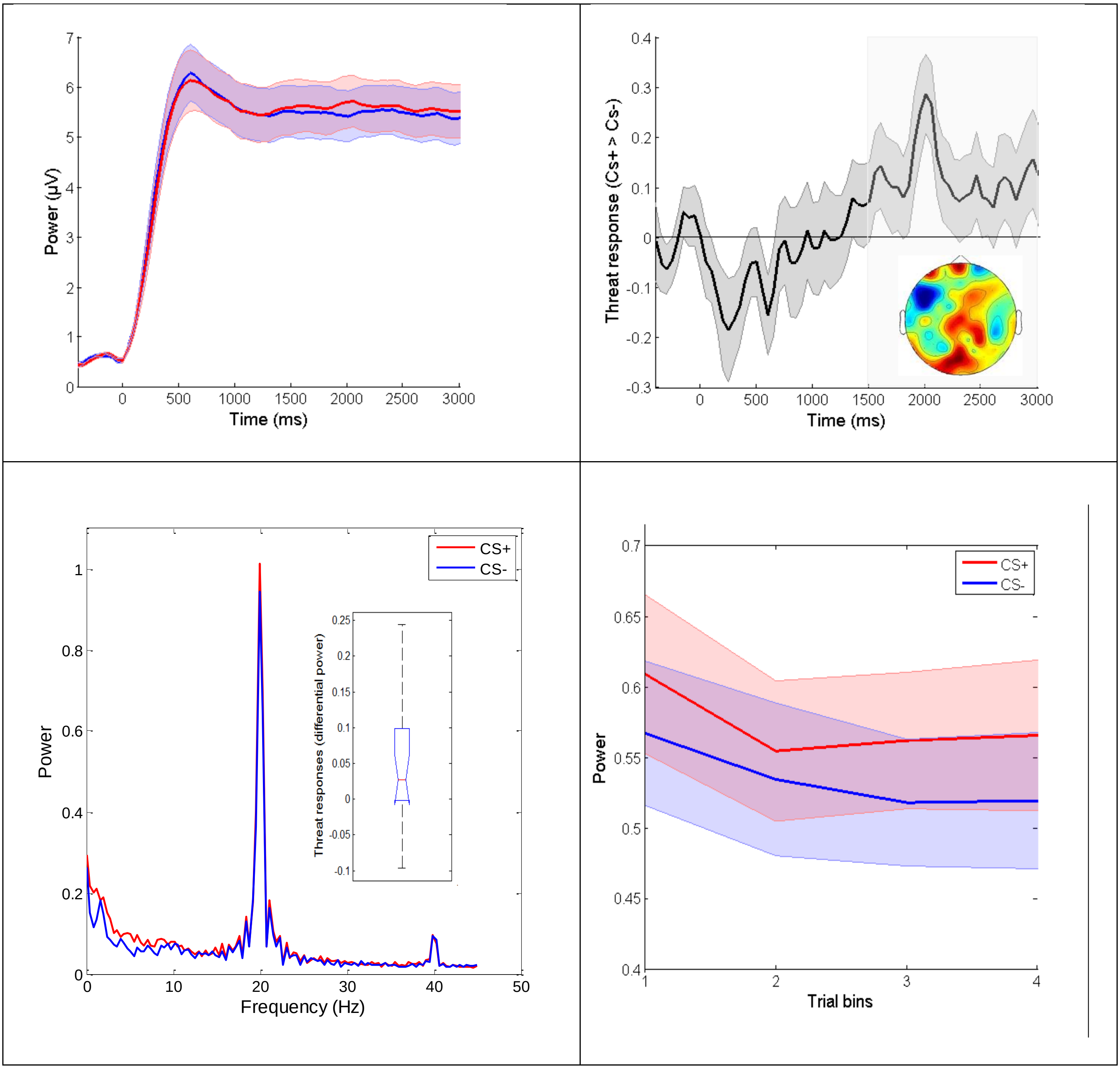

Power (µV)
0
1
2
3
4
5
6
7
0
500
1000
1500
2000
2500
3000
Time (ms)
Threat response (Cs+ > Cs-)
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
0
500
1000
1500
2000
2500
3000
Time (ms)
Power
0
0.2
0.4
0.6
0.8
1
0
10
20
30
40
50
Frequency (Hz)
CS+
CS-
Threat responses (differential power)
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
Power
0.7
0.65
0.6
0.55
0.5
0.45
0.4
1
2
3
4
Trial bins
CS+
CS-

FIGURE 3. CONTINGENCY AND LILKABILITY OF CONDITIONED STIMULI.

The contingency and likability ratings of stimuli used as CSs before and after the experimental task. Error bars indicate the standard error of the mean.

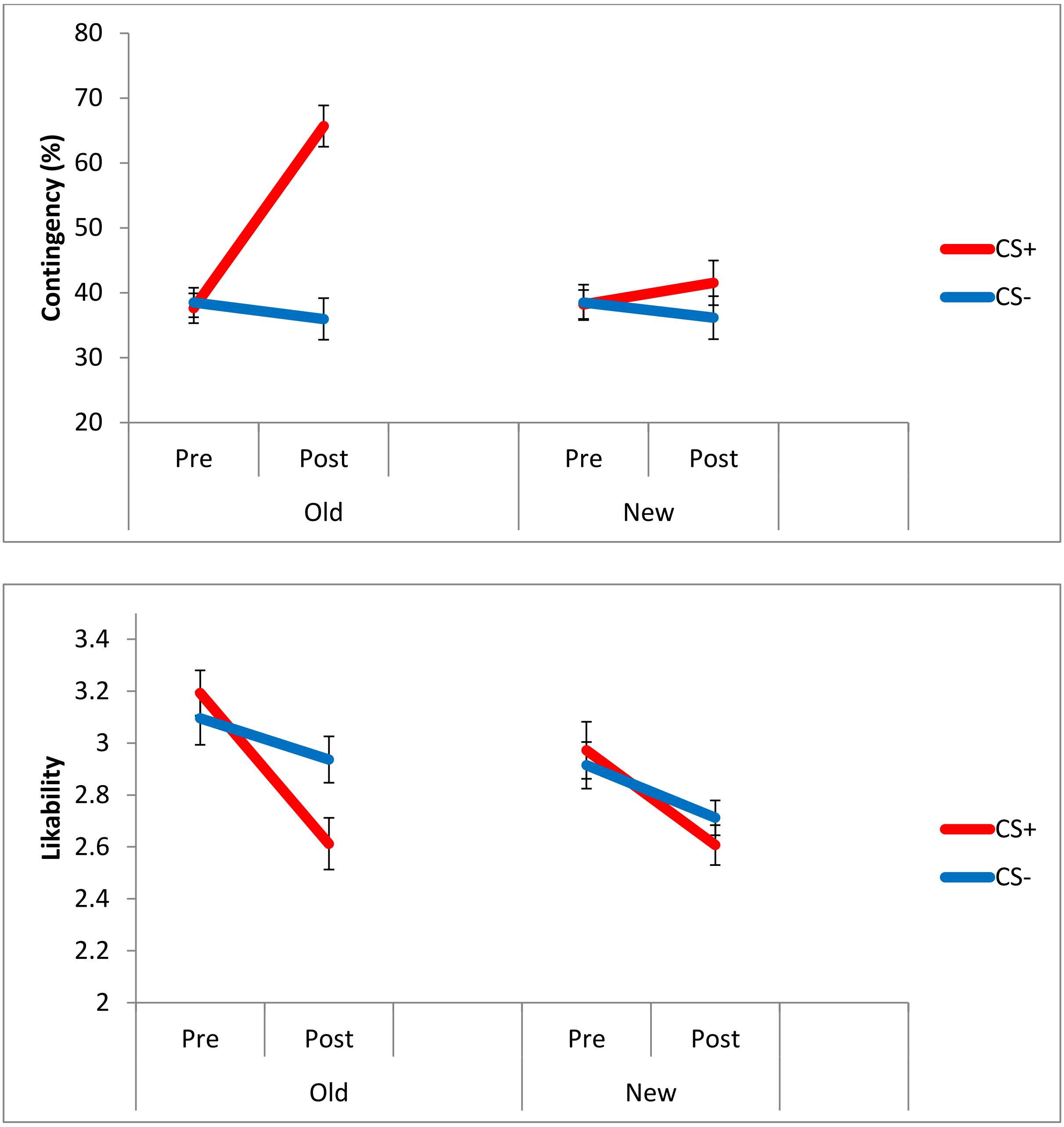
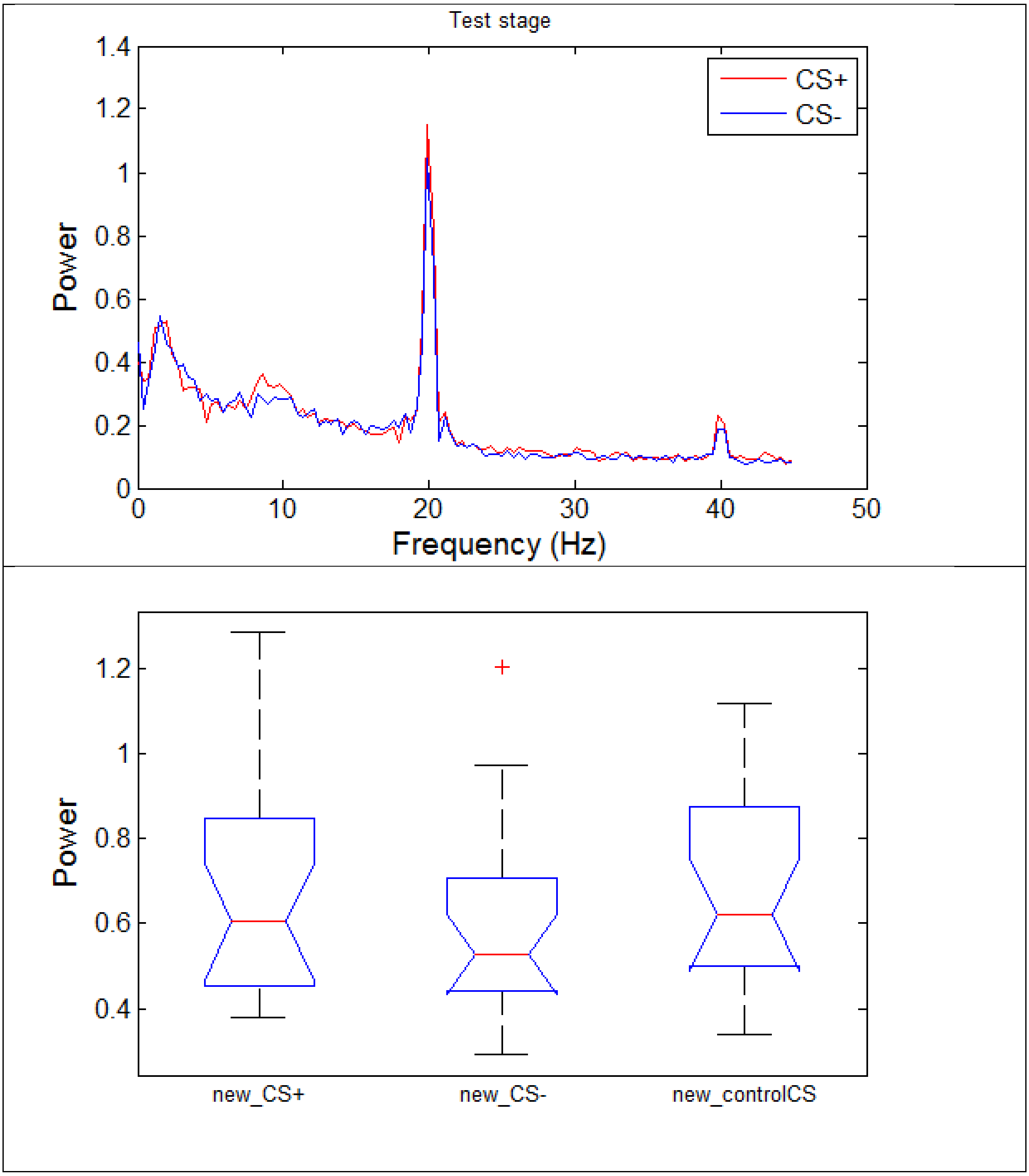


FIGURE 4. MODEL-BASED PAVLOVIAN MODULATION OF ATTENTION.

Top. SSVEP amplitudes in response to the new CSs in blocks that followed the experimental instructions.

Bottom. The magnitude of SSVEP amplitudes for the new CSs in the blocks that followed experimental and control instructions. The red line indicated the mean; the box indicates the inter-quartile range.


Test stage
Power
Frequency (Hz)
CS+
CS-
Power
new_CS+
new_CS-
new_controlCS