Reactivation of hedonic but not sensory representations in human emotional learning

Ehlers, M. R.[1,2], Kryklywy, J. H.[2], Beukers A. O.[3], Moore S. R.[4,5], Forys, B. J.[2], Anderson A.K.[4], Todd, R. M.[2,6]

[1]Department of Systems Neuroscience, University Hospital Hamburg-Eppendorf, Germany

[2]Department of Psychology, University of British Columbia, Canada

[3]Department of Psychology, Princeton University, United States

[4]Department of Psychology, Cornell University, United States

[5]Centre for Molecular Medicine and Therapeutics, University of British Columbia, Canada

[6]Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Canada

*Address correspondence to:

Mana Ehlers
Department of Systems Neuroscience
University Hospital Hamburg-Eppendorf
20246 Hamburg, Germany
Email: ma.ehlers@uke.de
Phone: +49 40 7410 55857

Abstract

Learning which stimuli in our environment co-occur with painful or pleasurable events is critical for survival. Previous research has established the basic neural and behavioural mechanisms of aversive and appetitive conditioning; however, it is unclear what precisely is learned. Here we examined what aspects of the unconditioned stimulus (US) - sensory and hedonic - are transferred to the conditioned stimulus (CS). To decode the content of brain activation patterns elicited during appetitive (soft touch) and aversive (painful touch) conditioning of faces, a novel variation of representational similarity analysis (RSA) based on theoretically driven representational patterns of interest (POIs) was applied to fMRI data. Once face associations were learned through conditioning, globally the CS reactivated US representational patterns showing conditioning-dependent reactivation. More specifically, in higher order brain regions, the CS only reactivated hedonic but not sensory aspects of the US – suggesting that affective conditioning primarily carries forward the valence of the experience rather than its sensory origins.

Introduction

The capacity to develop painful or pleasurable associations with predictive cues is highly conserved across species (Clark et al., 2002; Kryklywy et al., 2020; Pessoa et al., 2019). It is so central to our survival and our ability to make sense of the world that, often, a single exposure to an aversive or appetitive stimulus and an associated neutral stimulus is enough for us to remember the critical information for the rest of our lives (VanElzakker et al., 2014; Yamamoto et al., 1994). Decades of research on emotional learning processes have established basic neural and behavioral mechanisms by which human and non-human animals learn what cues (conditioned stimulus, CS+) predict the occurrence of inherently positive or negative events (unconditioned stimuli, US) (Andreatta & Pauli, 2015; Garcia & Koelling, 1966; LeDoux, 2003; Maren, 2001; Martin-Soelch et al., 2007). Yet longstanding fundamental questions about the nature of the information that we learn to associate with salient events remain to be resolved: When a conditioned cue is encountered, does the brain directly recapitulate representations of the pleasant or unpleasant experience the cue predicts? If so, what content is represented? Do objects associated with emotionally relevant events evoke the visceral sensory properties of the original salient event, or unconditioned stimulus? Or do they rather evoke the hedonic response to the experience, that is the painful and pleasurable qualities that were experienced along with the event?

These questions go back to different theories of learning. The stimulus-response (S-R) school argued that conditioning involves a representation of the association between a stimulus and its response (Holland, 2008), which can be a motor, physiological or hedonic response. In contrast, the stimulus-stimulus school (S-S) (Byrne, 2003) has proposed that, with learning, the initially neutral CS+ comes to elicit the same afferent activity initially elicited by the innately

46 arousing US (Hull, 1943), and an association between a sensory process and a motor response is

47 formed (Spence, 1950). These questions about the nature of emotional experience – about

48 overlap and interaction between sensation and cognition – are rooted in antiquity. Classical

49 philosophers have already debated whether emotion existed as a unitary experience, or as a set of

50 dissociable constructs, with unique components of cognitive appraisal and biological drives

51 (*Epicurus | Internet Encyclopedia of Philosophy*, n.d.; Sihvola & Engberg-Pedersen, 1998).

52 Interpreted in the context of the current study, they argued whether being exposed to an

53 appetitive or aversive event evokes a unitary experience involving both the activation of sensory

54 organs and a hedonic appraisal of the situation, or whether sensation and cognition are two

55 distinct entities independently contributing to an emotional experience that, in turn, can be

56 independently transferred to a neutral stimulus.

57 Today, behavioral evidence supporting either argument can be found. For example,

58 *autoshaping* or *sign-tracking* describes a phenomenon in classical conditioning in which animals

59 behaviorally engage with the CS in the same way as they would with the US (Brown & Jenkins,

60 1968; Flagel et al., 2009). Critically, this effect is so far reaching that some animals completely

61 neglect the US, thereby suggesting that S-S conditioning has occurred (Morrison et al., 2015). By

62 contrast, *outcome devaluation* experiments related to habit formation suggest that once a certain

63 action is established, a set response to a stimulus is initiated irrespective of the associated

64 outcome, thereby indicating S-R learning has occurred (Schwabe et al., 2007; Smith & Graybiel,

65 2016). Yet, a lack of appropriate techniques has hampered investigations into neurobiological

66 representations that could support our understanding of what is learned and transferred during

67 affective conditioning.

68   Recent advancements in functional neuroimaging, however, are providing exciting new

69 avenues towards addressing these questions. For example, in a series of experiments, Visser and

70 colleagues (Visser et al., 2011, 2013, 2015) employed trial-by-trial representational similarity

71 analysis (RSA), a multivariate approach to examining neural instantiations of the degree to

72 which content is categorized as more or less similar, to examine how associative learning

73 changes the representation of the initially neutral CS+. The authors showed that, before fear

74 conditioning, neutral stimuli were represented as most similar to other stimuli with similar visual

75 characteristics. However, after conditioning, stimuli paired with a US (the CS+), and those

76 without pairing (the CS-) formed separate representational categories, irrespective of their visual

77 similarity (Visser et al., 2011). The initial classification based on visual categories was

78 overwritten by the emotional association. Another study compared the neural pattern activation

79 in response to US and CS+ during a fear conditioning paradigm and tracked how the reactivation

80 of the US pattern by the CS+ develops over the course of learning in the insula (Onat & Büchel,

81 2015). However, these studies did not probe the *content* of the reactivated patterns they

82 observed; that is, they could not demonstrate if the US pattern gets reactivated in its entirety as

83 opposed to only components of it. Thus, the goal of the present study was to build upon this

84 work by investigating not only how the CS+ changes in representation following conditioning,

85 but also whether a CS+ reactivates brain activation patterns elicited by the US both in its general

86 representational form and as discriminable non-hedonic and hedonic components.

87   To do this, we employ Representational Similarity Analysis (RSA) (Kriegeskorte et al.,

88 2008; Kriegeskorte & Kievit, 2013), a multivariate approach to functional neuroimaging

89 analyses, that allows us to investigate population code representations. RSA was combined with

90 a hypothesis-driven variation of Pattern Component Modeling (PCM) (Diedrichsen et al., 2018).

91   This is an approach we recently developed to use fMRI to decode and fit representational

92   patterns in response to appetitive and aversive US to predefined pattern components modeling

93   theoretical representations patterns of interest (POIs) present in the US evoked signal (Kryklywy,

94   Ehlers, et al., 2021). With this technique, we could investigate the extent to which hedonic and

95   non-hedonic aspects of sensation are reproduced by the CS as novel affective associations are

96   acquired.

97   Participants completed two conditioning paradigms in the MR scanner in which either an

98   appetitive brush stroke to the forearm or painful pressure applied to the thumbnail were paired

99   with the presentation of specific facial identities. Representational similarity patterns from eight

100 regions of interest for CS and US were extracted on a trial-by-trial basis by means of RSA with

101 theoretically based pre-specified patterns of interest (POIs) which we identified via PCM. We

102 found that, once learned associations were established, the CS reactivates US representational

103 patterns in brain regions typically associated with conditioning (e.g. amygdala, ventromedial

104 prefrontal cortex and insula). By comparing patterns of representation with theoretically

105 grounded POIs, we demonstrate that primary sensory regions reactivate components of the US

106 that do not rely on learni ng. In contrast, higher order brain regions reactivate representations of

107 hedonic value of the US, supporting a model of stimulus-response learning in the human brain.

108

Results

110   In separate tasks for appetitive and aversive associative learning, two different faces (male

111 and female for CS+ in aversive and appetitive tasks) were paired with either pleasurable brush

112 strokes to the forearm or aversive pressure to the thumbnail (US). Two other faces (CS-) were

113   never paired with the US (see Figure 1a). All faces wore affectively neutral expressions and

114   individual identities were counterbalanced between participants.

115
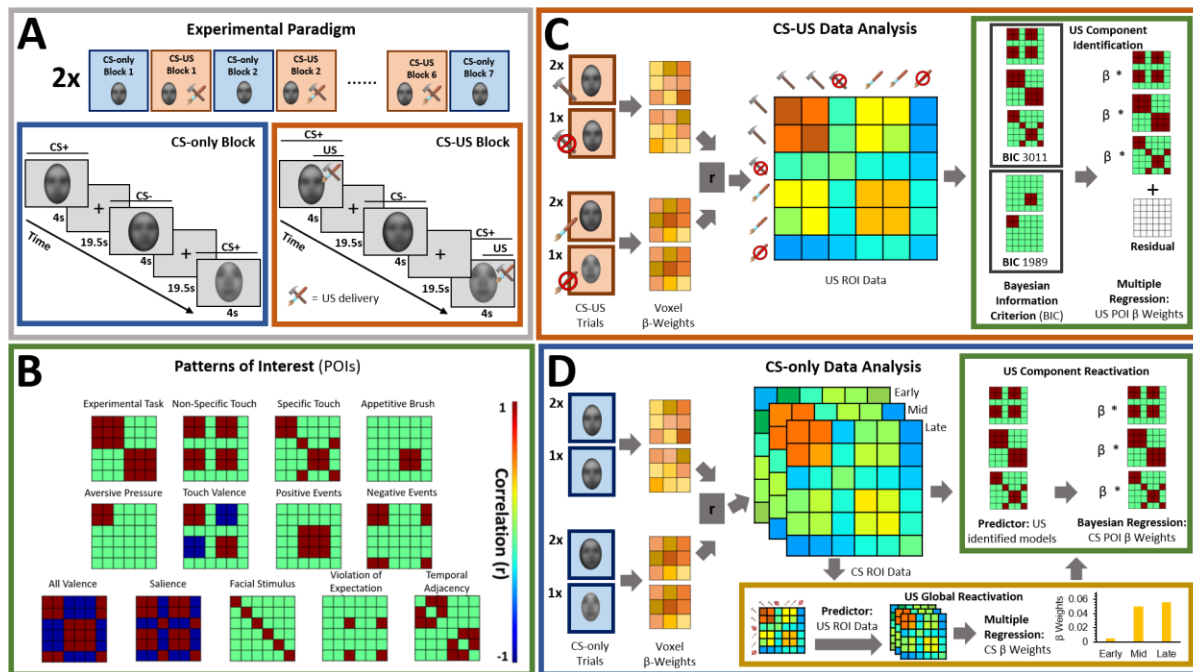
Figure 1. A. Both experimental tasks (appetitive and aversive conditioning) followed the same

general structure in which 7 CS-only blocks were interleaved with 6 CS-US paired blocks. In CS-

only blocks, faces (blurred on figure only) were presented by themselves, while the CS+ faces

were paired with appetitive brush or aversive pressure in the CS-US blocks. **B.** Patterns of interest

(POIs) demonstrate the representational pattern that would be observed in the experimental data if

a region of interest (ROI) perfectly represented the theoretically derived constructs. **C.** Voxel

activation patterns in each ROI, averaged over all CS-US trials for each task condition, were

correlated with those in each other task condition using a representational similarity analysis

(RSA) approach. Bayesian Information Criterion (BIC) was used to find the best combination of

POIs for each ROI. Multiple regression was used to obtain beta coefficients for each POI

(Kryklywy, Ehlers, et al., 2021). **D.** RSA was performed separately on early, mid, and late CS-

only trials to examine reactivation of US information over the course of conditioning. (Bayesian)

7

128   multiple regression was used in order to quantify the US model reactivation by CS-only data

129   globally and BIC identified components were used to quantify which US aspects were reactivated

130   by CS.

131

132   **Subjective ratings**

133   For each face, ratings of likeability and trustworthiness were acquired before and after

134   associative learning. CS discrimination, i.e. [CS+ minus CS-], in likeability and trustworthiness

135   ratings was calculated for ratings obtained before and after completion of the appetitive and

136   aversive tasks. Before appetitive conditioning, no CS discrimination was evident in likeability,

137    $t(68) = 0.29$, $p = .772$, 95% CI [-3.66, 4.90], or trustworthiness ratings, $t(68) = 1.21$, $p = .232$,

138   95% CI [-1.21, 6.98] as expected for two neutral facial expressions. Likewise no difference

139   between CS+ and CS- likeability, $t(68) = -0.92$, $p = .359$, 95% CI [-5.77, 2.12] and

140   trustworthiness, $t(68) = -1.52$, $p = .134$, 95% CI [-6.93, 0.94] ratings was measured before

141   aversive conditioning. Importantly however, CS+ and CS- likeability (appetitive conditioning:

142    $t(68) = -3.52$, $p < .001$, 95% CI [-13.59, -3.76]; aversive conditioning: $t(68) = 2.61$, $p = .011$,

143   95% CI [1.89, 14.06]) differed significantly after conditioning such that in the appetitive task, the

144   CS+ was more likeable, while in the aversive task, the CS- was more likeable indicating

145   successful conditioning in both tasks. CS discrimination in trustworthiness reached significance

146   only in the appetitive, $t(68) = -2.58$, $p < .012$, 95% CI [-11.47, -1.45], but not in the aversive

147   task, $t(68) = 1.45$, $p < .152$, 95% CI [-0.94, 5.92]. Overall stimulus ratings indicate successful

148   conditioning in both domains.

149

150

US pattern reactivation by CS

151

152   The goal of the current study was to determine the extent and content of US pattern

153   reactivation by the CS once associations were learned. This was achieved by comparing patterns

154   of information representation identified during CS-only conditions early, mid, and late in

155   conditioning to previously identified representational patterns observed in CS-US paired trials.

156   The extent was assessed as both a global reactivation of US representational patterns, and the

157   reactivation of specific informative patterns of interest (POIs, Figure 1b), which further provided

158   information about the content of reactivation.

159   After initial preprocessing and first-level analysis of fMRI data, representational

160   similarity analysis (RSA) was conducted on the patterns of BOLD activity within predefined

161   regions of interest (ROIs). This resulted in a set of similarity matrices, comprised of correlations

162   between all pairs of conditions (separate for CS-US paired, and CS-only trials; see Figure 1c/d).

163   Representational information in BOLD responses to the US was identified by using Bayesian

164   Information Criterion (BIC) to fit predefined POIs - idealized similarity matrices modelling

165   specific information content (e.g., the valence of the touch stimulus) - to the observed US

166   similarity data (Figure 1c). The representation of appetitive and aversive US, independent from

167   CS-only trials, are focus of a different study (Kryklywy, Ehlers, et al., 2021).

168   To identify the extent to which US patterns were reactivated by CS following learning, a

169   reconstructed measure of US activation similarity (rUS) was fit to CS similarity data from early-,

170   mid-, and late-conditioning trials. rUS were constructed by summing the scaled contribution of

171   each POI contributing to overall similarity in an ROI (identified though hypothesis-driven PCM;

172   see (Kryklywy, Ehlers, et al., 2021). Next, to examine the degree to which each specific US-

173   defined POI was comparable to the pattern of CS activation - the content driven reactivation -

9

174   separate Bayesian linear models with the combination of US–defined POIs as predictors, and

175   rUS and CS representational patterns as outcome variables were fitted in order to compare the

176   (beta) weight of each predictor for the CS representational pattern to that of the US (see Figure

177   1d).

178   A reactivation of the following US-defined POIs is interpreted as reactivation of hedonic

179   information: 'Aversive Pressure'/'Appetitive Brush' (representation of either US type

180   independently); 'Touch Valence' (representation of the positive relative to negative valence of

181   US stimuli on a single continuum); and 'Negative Events' (representation of both aversive

182   pressure and absence of pleasurable touch). Interpretation of these as hedonic information is

183   particularly strong if these POIs are observed *without* reactivation of general 'Non-Specific

184   Touch' representational patterns (representation of the tactile manipulation independent of

185   affective discrimination).

186

187   Sensory regions of interest

188   *Primary Somatosensory Cortex (S1):* Globally, the rUS pattern predicted the CS

189   representation in mid and late conditioning, while the association was at trend level for early

190   conditioning ($p = .069$) (see Table 1). Thus, there was some increase in the reliability of global

191   reactivation effects with learning. The analysis of POI reactivation (See Figure 2) showed that

192   the contribution of the POI 'Experimental Task', where there is similarity between CS+ and CS-

193   within each conditioning task, to CS representational patterns was consistent with that of the US

194   in mid and late conditioning, showing a conditioning effect. In contrast, 'Non-specific Touch'

195   was consistent with the US pattern in early and late conditioning, a pattern inconsistent with the

196   conditioning effect (see Figure 2).

197   *Primary/secondary visual cortex (V1):* Global CS representational patterns were

198   significantly predicted by V1 rUS during early, mid, and late conditioning, indicating no effect

199   of conditioning (see Table 1). Likewise, the reactivation of the POI 'Experimental Task' was

200   consistent with a contribution to US representational patterns during all conditioning phases (see

201   Figure 2), indicating that the representation of visual information does not depend on learned

202   associations.

203   *Ventral visual structures (VVS):* Similar to V1, the rUS pattern was a significant predictor

204   of representational patterns at all conditioning time points in ventral visual structures. Moreover,

205   the contribution of the POI 'Experimental Task' was consistent across US and CS

206   representational patterns (see Figure 2). Taken together these results indicate no effect of

207   affective pairings on representational patterns in these visual regions.

208

209   Integrative regions of interest

210   *Amygdala:* CS patterns predicted Amygdala rUS patterns in late conditioning only,

211   suggesting overall a slightly delayed response to learning (see Table 1). The analysis of

212   component reactivation, however, revealed that only 'Touch Valence' showed an effect

213   consistent with US data and only early in conditioning (see Figure 2).Thus, whereas the

214   amygdala was the only ROI to recapitulate the valence of the US with appetitive and aversive at

215   opposite poles, the effects were transient and inconsistent with the pattern of global

216   recapitulation.

217   *Ventromedial prefrontal cortex (vmPFC):* CS representational patterns in vmPFC showed

218   clear effects of conditioning: They were not significantly predicted by vmPFC rUS pattern

219   during early conditioning, but were during both mid and late conditioning. Examination of

11

220   individual POIs indicated distinct representations of appetitive and aversive stimuli, which

221   emerged at different time points: In mid conditioning 'Appetitive Brush' was reactivated,

222   whereas in late conditioning only 'Aversive Pressure' showed reactivation. The results suggest

223   that, in the vmPFC, aversive and appetitive information is carried forward in conditioning in a

224   nonlinear fashion and with different temporal patterns.

225   *Anterior cingulate cortex (ACC):* As for the vmPFC, ACC rUS could not predict ACC-

226   CS representational patterns in early conditioning, but was highly predictive for both mid and

227   late conditioning (see Table 1). These results indicate that ACC may represent affective

228   associations acquired though classical conditioning. US-consistent component reactivation for

229   'Experimental Task' (similarity between all conditions within a task) by CS was observed mid

230   and late conditioning. In addition, while 'Aversive Pressure' reactivation by CS data did not

231   reach the same levels as the US, it becomes apparent that an increase in reactivation from early

232   to mid and late conditioning is induced by conditioning (see Figure 2).

233   *Insula (anterior and posterior):* Here again, no significant relationship between the

234   anterior insula CS representational patterns and anterior insula rUS patterns was found during

235   early conditioning, but was observed during mid and late conditioning. Similarly, the posterior

236   insula rUS pattern was not predictive of the CS representational pattern during early

237   conditioning, but was during mid and late conditioning (see Table 1). This suggests that global

238   activation patterns in anterior and posterior insula represent conditioned affective associations.

239   The anterior insula showed clear conditioning effects, with US-consistent contribution of

240   'Experimental Task' and 'Aversive Pressure' in mid and late conditioning. The pattern of results

241   suggests a bias for the anterior insula to carry forward negative hedonic information during

242   conditioning. We also observed conditioning independent-reactivation of 'Negative Events'

12

243   (both aversive pressure and absence of pleasurable touch) in early conditioning. On the other

244   hand, the posterior insula only showed a US-consistent contribution of 'Experimental Task' in

245   mid and late conditioning. This might represent the differences between the specific tactile and

246   facial stimuli between the appetitive and aversive conditioning task (see Figure 2). Overall, the

247   current data indicates some distinct patterns of reactivation in anterior and posterior insula

248   supporting the notion that these subregions play functionally distinct roles in representing

249   information about conditioned cues.

250

Tables and Figures

Table 1. Global US reactivation by CS representation patterns in early, mid and late conditioning

Global reactivation (df = 1, 1489)

ROI	Early Conditioning				Mid Conditioning				Late Conditioning			
	Adj.R^2	F	p	β	Adj.R^2	F	p	β	Adj.R^2	F	p	β
S1	.002	3.311	.069	.185	.004	6.671	**.010**	.287	.003	6.087	**.014**	.281
V1	.003	6.082	**.014**	.734	.004	7.241	**.007**	.842	.007	10.55	**.001**	1.086
VVS	.005	8.843	**.003**	.812	.005	8.766	**.003**	.848	.003	5.188	**.023**	.713
Amy	.-5.266e	0.922	.337	-.175	.0003	1.518	.218	.232	.003	5.562	**.018**	.471
vmPFC	-.0003	0.487	.486	-.153	.006	9.091	**.003**	0.743	.005	8.537	**.004**	.723
ACC	.0006	1.937	.164	-.177	.004	6.704	**.010**	.357	.011	17.20	**<.001**	.609
aIns	-.0001	0.792	.374	.101	.011	17.18	**<.001**	.515	.015	23.13	**<.001**	.630
pIns	-.0006	0.018	.893	.010	.006	10.43	**.001**	.248	.002	4.221	**.040**	.173

S1 = primary somatosensory cortex; V1 = primary/secondary visual cortex; VVS = ventral visual structures; Amy = amygdala; vmPFC = ventromedial prefrontal cortex; ACC = anterior cingulate cortex; aIns = anterior insula; pIns = posterior insula; bolded p-value indicates significance at α = 0.05
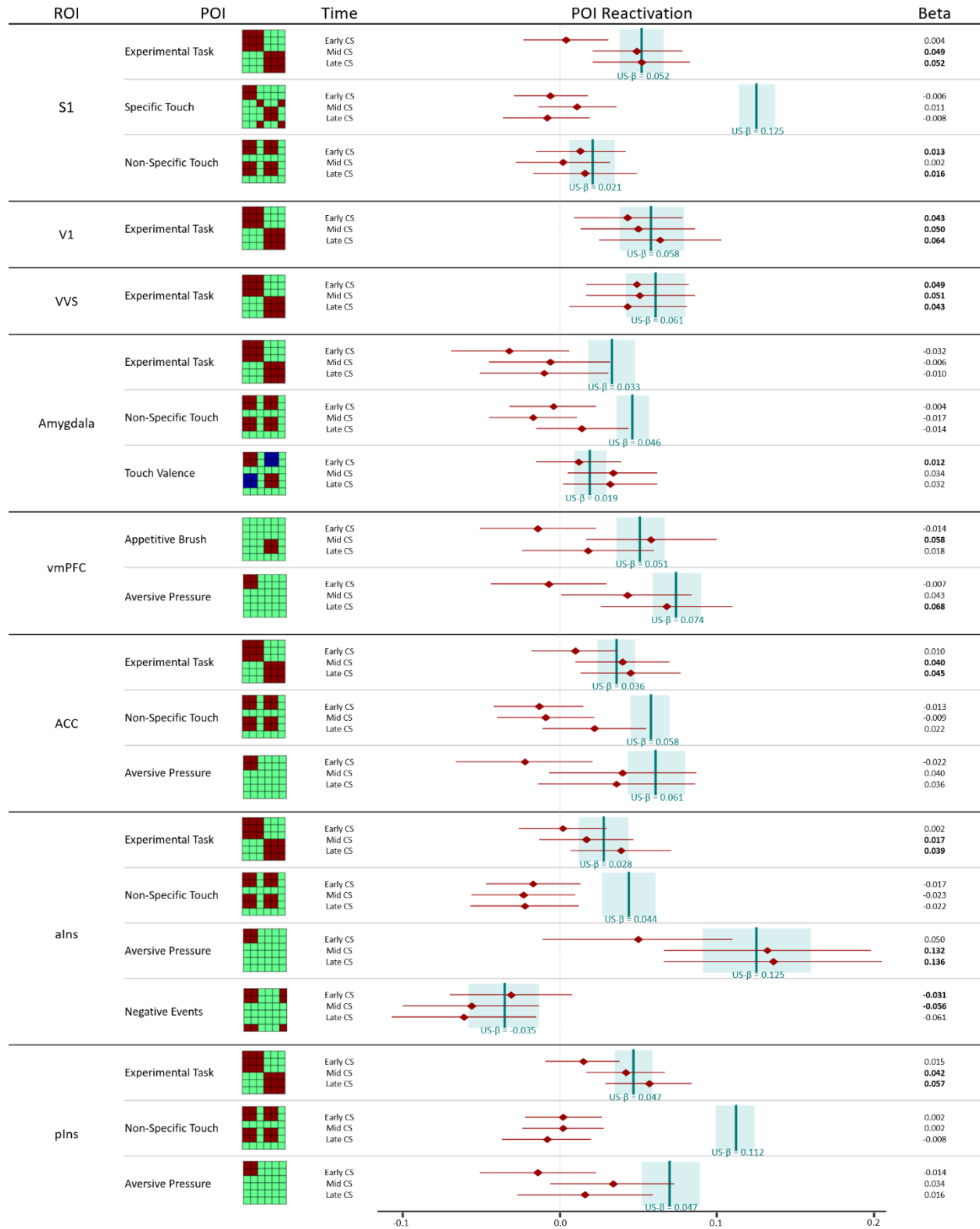
Figure 2. Pattern of Interest (POI) reactivation by reconstructed US (rUS) as well as early, mid and late CS data in different regions of interest (ROIs). Displayed are beta weights and 95% credible intervals for each POI reactivation obtained from separate Bayesian linear models performed on rUS and CS data with the illustrated POIs as predictors. The contribution of any given POI at any given time point of CS conditioning is considered consistent with that POIs contribution to rUS data if the point estimate of the CS data (in red) falls within the credible interval of the rUS data (in blue). Consistent contribution is further indicated by printing the obtained beta weight in bold.

S1 = primary somatosensory cortex; V1 = primary/secondary visual cortex; VVS = ventral visual structures; vmPFC = ventromedial prefrontal cortex; ACC = anterior cingulate cortex; aIns = anterior insula; pIns = posterior insula

Discussion

In this study we examined whether, with emotional learning, the initially neutral conditioned stimulus (CS) reactivates patterns of activation elicited by the unconditioned stimulus. We further probed whether neural population codes for the CS represent non-hedonic sensory and/or hedonic aspects of the appetitive or aversive unconditioned stimulus. In other words, we asked if what is carried forward in conditioning is the entire sensory and hedonic construct that a US encompasses, or whether only hedonic aspects of the appetitive or aversive stimulus become associated with the cue. We used representational similarity analysis (RSA) with a theoretically based version of pattern component modeling (PCM) on fMRI data to decode the content of neural representations observed over the course of aversive and appetitive classical conditioning. The data revealed that, for higher-order regions of interest, a significant

16

amount of variance in the CS data could be explained by patterns elicited by the US, indicating US pattern reactivation by CS. Critically, this effect was only found after affective pairing had been encountered, and not prior to the learning process. By employing pattern component modeling (PCM) and determining the best combination of patterns of interest (POIs) for different ROIs, we were further able to model the distinct content of information that is represented by each brain region for the US and the degree to which it was reactivated by the CS. In primary sensory regions, the CS primarily reactivated components of the US that do not rely on learning. In contrast, several brain regions implicated in emotional learning showed reactivation of those components representing hedonic value of the US – providing evidence for a dominance of stimulus-response learning.

Global US representation pattern reactivation by CS. We first wanted to establish whether, through conditioning, the representation of an initially neutral stimulus changes to resemble that of an inherently positive or negative one. The results from the current study suggest that, indeed, the CS reactivates patterns of US representation in all regions of interest included in the analysis. Closer examination of individual ROIs showed a more nuanced pattern of results. Visual cortex and ventral visual structures showed significant reactivation of US representational patterns across *all* experimental time points (early, mid and late conditioning). This pattern indicates that reactivation is not a result of conditioned associations but rather of experimental phase overlapping features inherent to the task (e.g., the stable visual properties of the stimuli). In contrast, amygdala, vmPFC, ACC as well as anterior and posterior insula predicted CS representational patterns from US data only *after* conditioned associations had developed. While one previous study (Onat & Büchel, 2015) has found preliminary evidence for such an effect for fear conditioning in the insula, the present study shows that this finding can be

17

generalized to both appetitive and aversive associative learning, and to other brain regions. Thus, in this study we provide strong evidence suggesting that the basic mechanisms observed behaviorally in conditioning - that is, that the CS elicits the same response as the US (Maren, 2001; Martin-Soelch et al., 2007; Pavlov, 1927) - are also represented in the brain.

US component reactivation by CS. After having established that, globally, the CS reactivates US representational patterns after associative learning, we next focused on the *content* of those representations. Rather than interpreting stimulus representation as a homogenous construct, we deconstructed the representation into theoretically driven patterns of interest (POIs) that allowed us to examine the aspects of the US that become attached to the CS as conditioned associations emerge. Most regions of interest included in the current analysis showed reactivation of the US representational pattern by CS. Primary visual cortex and ventral visual structures showed an experimental task-based component (POI: 'Experimental Task' indicating pain task vs. brush task) reactivation that did not depend on conditioning, consistent with the fact that, the visual input associated with each task did not change over the course of learning.

In contrast, component reactivation in the vmPFC and, to some degree, in the amygdala and the anterior insula, provide answers to the question of what aspects of the US become attached to the CS. A substantial body of literature has delineated complementary roles for the amygdala and vmPFC in establishing conditioning (Schoenbaum & Roesch, 2005). In the current study, in the vmPFC, representations of both appetitive and aversive touch (POIs: 'Aversive Pressure', 'Appetitive Brush') were reactivated. Of note, reactivation of appetitive touch was observed mid conditioning, while aversive touch was observed during late conditioning only. This might indicate that the association with appetitive touch developed more quickly than with

18

aversive pressure but that at the same time habituation to the appetitive brush is faster than to aversive pressure (Triscoli et al., 2014). First and foremost, however, the pattern of the results suggests that hedonic properties of the US are represented in the vmPFC and are reactivated during CS presentation. This finding is in line with previous studies showing that stimulus value (but not sensory properties) are represented in the vmPFC/OFC in conditioning (Chikazoe et al., 2019; Lim et al., 2013; McNamee et al., 2013). The amygdala was the only region where the representation of the hedonic experience of touch valence (POI. 'Touch Valence') was reactivated, rather than non-specific sensory input or aversive pressure. Yet the contribution of 'Touch Valence' was consistent with the US data early in conditioning only. This pattern of results shows that the amygdala represents appetitive and aversive associations as polar ends on a valence spectrum. This finding is consistent with recent evidence in mice and humans for spatially segregated antagonistic valenced neurons and, in humans, coding of a spectrum of valence in the amygdala (Jin et al., 2015; Kim et al., 2016). It should be highlighted, however, that in the current task this effect did not seem to emerge with conditioning but was rather inherent to the processing of the facial stimulus alone.

Interestingly, while the US representations in anterior and posterior insula showed large overlap, a distinctive role for each of the two regions emerged for the CS during conditioning. While both anterior and posterior insula activation showed a contribution of the task-based model (POI: 'Experimental Task') that contrasts appetitive and aversive conditioning tasks in general, the anterior insula further showed reactivation of components representing negatively valenced information. Much research has been done on the role of the insula and its subregions with somewhat inconsistent results. A recent meta-analysis (Kurth et al., 2010) revealed that most functions investigated have been associated with activation in the anterior portion,

especially emotional processing. Only sensorimotor processing has been exclusively mapped onto the posterior insula while pain processing has been shown to involve the entire insula (Craig, 2002). In the light of these meta-analysis findings, the current results of reactivation of negative information representation in the anterior insula, especially without reactivation of non-specific sensory input (reactivation POI: 'Non-specific Touch') strongly suggest reactivation of hedonic but not sensory US components. In contrast, the task-based (reactivation POI: 'Experimental Task') in the posterior insula could be explained by the differences in sensory input between appetitive and aversive tasks, i.e. female faces paired with appetitive brush and male faces paired with aversive pressure respectively. In summary, after demonstrating US component reactivation by CS with conditioning, the current findings show that several brain regions that have been previously associated with conditioning are biased to reactivate components that carry hedonic information instead of purely tactile information. Thus, the pattern of results observed here suggests that what is carried forward in conditioning and becomes associated with the CS is primarily the affective attachment with the stimulus rather than the pure sensory experience. In other words, when we are exposed to a conditioned stimulus we re-experience the pleasant or unpleasant feelings elicited by the US rather than the sensory stimulation. Taken together, the current analysis provides support for the notion of stimulus-response conditioning in which a CS reproduces the unconditioned response rather than the perceptual experience of the US.

In conclusion, we demonstrated for the first time that, in conditioning, a conditioned stimulus reactivates the pattern of activation initially elicited by the unconditioned stimulus in several brain regions. The results further show that it is primarily the hedonic components of the initial experience, rather than discriminative sensation, that is reproduced when we encounter a

20

cue that predicts it. Thus, when we encounter a cue signaling pain or pleasure, what we carry forward is the emotional meaning we attach to it rather than the sensory experience.

Materials and Methods

Participants

Data from 71 young, healthy participants (age: 21.1 ± 2.8 years, 41 females) was included in the analysis. Initially, 107 participants were recruited from Cornell University to participate in a brain imaging study of appetitive and aversive classical conditioning tasks. A number of participants had to be excluded for the following reasons: 20 participants had missing data (imaging run, stimulus onset files, motion correction files) while multi-echo preprocessing described below failed for 16 participants. All participants gave written, informed consent and had normal or corrected-to-normal vision. Participants were pre-screened for a history of anxiety and depression as well as other psychopathology, epilepsy and brain surgery in addition to general suitability for fMRI data collection. Pre-screening was followed up in person by an additional interview to ensure inclusion criteria were met. Due to the fact that this study was conducted as part of a larger research program, all participants were genotyped.

The neural representation of the appetitive and aversive US as well as the development of the PCM derivative and other methodical details have been described in (Kryklywy, Ehlers, et al., 2021).

Materials

Stimulus and apparatus

Six faces were chosen from the Karolinska directed emotional faces, comprising three male and three female exemplars each with a neutral expression (Goeleven, De Raedt, Leyman, & Verschuere, 2008). These faces were used as the conditioned stimuli (CS) in a classical conditioning paradigm. The (US) consisted of either an aversive pressure delivered to the right thumb, or an appetitive brush stroke to the participant's forearm. Aversive pressure stimuli were delivered using a custom designed hydraulic device, similar to those used in previous studies (Giesecke et al., 2004; López-Solà et al., 2010), capable of transmitting controlled pressure to 1 cm^2 surface placed on the subjects' right thumbnail. In individual calibration sessions, it was ensured that the pressure intensity was aversive but not excessively painful. Appetitive brush strokes were manually applied to the left forearm lasting ~4s. Individual subjective responding to brush stimuli were recorded in a separate session prior to all experimental scanning, with only participant who responded positively to the manipulation invited to participant in the scanning session.

Procedure

Stimulus ratings

As a measure of subjective stimulus assessment and conditioning, participants were asked to rate the likeability and trustworthiness of the faces used as CS+ and CS- stimuli on a scale from 1-100 (1) before and (2) after conditioning as a measure of conditioning. CS Discrimination scores ([CS+ minus CS-]) before and after conditioning were calculated and

22

compared using t-tests. Due to technical difficulties, stimulus ratings were only available for 69 of the 71 participants included in the analysis.

Experimental tasks

While undergoing functional MR scanning, participants completed two unique conditioning tasks with nearly identical structure modeled after Visser and colleagues (2015). These tasks differed from each other only in the nature of the tactile unconditioned stimulus (US; see above), and the gender of the face stimuli. In each task, participants completed seven CS-only blocks interleaved with six CS-US paired blocks (see Figure 2a). Single blocks of either the CS-only or the CS-US pairing entailed one presentation of each of the three male or female face stimuli used in that task. The order of the CS+ and the CS- faces was randomized within each CS-US block. Individual trials started with an initial fixation period (19500 ms) followed by the presentation of a face stimulus (4000 ms). The fixed and long interstimulus interval was included in the experimental design to reduce intrinsic noise correlations (Visser et al., 2013). During CS-only trials, all faces were presented without tactile stimulation (see Figure 1d). During CS-US paired trials, two of three facial stimuli were paired with tactile stimulation, thus creating two CS+ and one CS- face stimuli (see Figure 2b). The US was delivered from the midpoint after the visual stimulus presentation (2000 ms post-onset), and remained for the duration of the visual

presentation (2000 ms). Face pairings were randomly assigned for each participant but held constant across the duration of the experiment.

MRI acquisition and preprocessing

Acquisition

Scanning was conducted on a 3 Tesla GE Discovery magnetic resonance scanner using a 32-channel head coil at Cornell University. For each subject, a T1-weighted MPRAGE sequence was used to obtain high-resolution anatomical images (repetition time (TR) = 7 ms, echo time (TE) = 3.42 ms, field of view (FOV) 256 x 256 mm slice thickness 1 mm, 176 slices). The functional tasks were acquired with the following multi-echo (ME) EPI sequence: TR = 2000 ms, TE1 = 11.7 ms, TE2 = 24.2 ms and TE3 = 37.1 ms, flip angle 77°; FOV 240 x 240 mm. A total of 102 slices was acquired with a voxel size of 3 x 3 x 3 mm. Pulse and respiration data were acquired with scanner-integrated devices.

Preprocessing

Multi-echo independent component analysis (ME-ICA, meica.py version 3.2 beta1) was used to denoise the multi-echo fMRI data. An optimally combined (OC) dataset was generated from the functional multi-echo data by taking a weighted summation of the three echoes, using an exponential T2* weighting approach (Posse et al., 1999). Multi-echo principal components analysis was first applied to the OC dataset to reduce the data dimensionality. Spatial independent component analysis (ICA) was then applied and the independent component time-series were fit to the pre-processed time-series from each of the three echoes to generate ICA weights for each echo. These weights were subsequently fitted to the linear TE-dependence

24

and TE-independence models to generate F-statistics and component-level κ and ρ values, which respectively indicate blood-oxygen-level-dependent (BOLD) and non-BOLD weightings (Kundu et al., 2012). The κ and ρ metrics were then used to identify non-BOLD-like components to be regressed out of the OC dataset as noise regressors (Kundu et al., 2013).

Functional imaging analyses

Regions of interest

To assess tactile (aversive pressure, appetitive brush) and hedonic representations in neural patterns, eight bilateral regions of interest (ROIs) were generated from the standard anatomical atlas (MNI_caez_ml_18) implemented in the Analysis of Functional NeuroImages (AFNI) software package (Cox, 1996): primary somatosensory cortex (S1), primary/secondary visual cortex (V1) were selected as the primary sites of tactile and visual information respectively. In addition, ventral visual structures (VVS) were chosen due to their role in visual classification (Kanwisher et al., 1997; Kravitz et al., 2013). Amygdala, ventromedial prefrontal cortex (vmPFC), anterior cingulate cortex (ACC) and insula were further selected for their hypothesized roles in affect processing (Anderson & Phelps, 2002; Craig, 2002; Winecoff et al., 2013) and pain representations (Kragel et al., 2018; Orenius et al., 2017). The insula was further divided into an anterior and posterior portion due to its functional and anatomical subdivisions (Nieuwenhuys, 2012) for a total of eight ROIs.

Representational similarity analysis

Data analysis of the fMRI data was conducted using Analysis of Functional NeuroImages (AFNI) software (Cox, 1996). Regressor files of interest were generated for all individual trials

25

across the experiment, modelling the time course of each stimulus presentation during each run (36 total events). The relevant hemodynamic response function was fit to each regressor to perform linear regression modeling. This resulted in a β coefficient and t value for each voxel and regressor. To facilitate group analysis, each individual's data were transformed into the standard brain space of Montreal Neurological Institute (MNI).

In order to identify the representational pattern elicited by the experimental stimuli, representational similarity analysis (RSA) was performed using the Python package PyMVPA (Hanke et al., 2009). For each participant, in each ROI the spatial pattern of β weights in response to each experimental condition or event was correlated with the pattern of activation in response to all other events. This step was performed separately for each ROI. Thus, pairwise Pearson coefficients for all experimental events of a single ROI resulted in a similarity matrix containing correlations for all CS-US combinations for all trials for each participant. Fischer transformations were performed on all similarity matrices to allow comparisons between participants.

Pattern Component Modeling

In order to characterize the content of CS (and US) representation in key regions of interest, we developed a theory-guided implementation of Pattern Component Modeling (PCM) (Diedrichsen et al., 2018; Kriegeskorte & Kievit, 2013; Kryklywy, Ehlers, et al., 2021; Kryklywy, Forys, et al., 2021). The details are described in Kryklywy, Ehlers, et al., 2021. In brief, we created 13 patterns of interest (POIs) to represent dissociable correlation patterns that would be observed in the experimental data if it would contain perfect representation of distinct theoretically-derived constructs (Fig 1C). POIs were constructed for 1) Experimental Task, 2) Non-Specific Touch, 3)

Specific Touch, 4) Appetitive Brush, 5) Aversive Pressure, 6) Touch Valence, 7) Positive Events, 8) Negative Events, 9) All Valence, 10) Salience, 11) Face Stimulus, 12) Violation of Expectation and 13) Temporal Adjacency (see Figure 1c). In order to determine the POI combinations that best explained the observed correlation in the US data in each ROI, a Bayesian Information Criterion (BIC) analysis and multiple regression implemented in our R package 'PCMforR' (Kryklywy, Forys, et al., 2021) were conducted. A reconstructed US (rUS) pattern was built from identified POIs in order to reduce noise. The rUS pattern was first used as predictor for the CS pattern in early, mid and later conditioning to determine the global reactivation of US patterns by CS-only data. Subsequently, the US identified POIs were used as predictors in Bayesian linear models for both rUS and CS data in order to compare the contribution (beta weight) of each POI between rUS and CS data. For that purpose the R package 'BayesFactor' (Morey et al., 2018) was used. The Bayesian linear model was estimated with 1,000,000 iterations allowing us to extract mean beta weights for each POI and their 95 % credible intervals (CrIs). In order to determine whether the contribution of each POI to the CS data is comparable to that of the US data, we adapted an approach developed to assess the robustness of replications (LeBel et al., 2018) that has recently also been employed in a Bayesian framework (Kuhn et al., 2021). While we are not comparing replication attempts, we have adapted the measure of consistency described previously (LeBel et al., 2018) in such a way that consistency between US and CS data is assumed when the beta weight point estimate obtained from CS data of any given POI is included in the credible interval of the beta weight obtained from US data for the same POI.

Author Contributions:

M.R. Ehlers: conceptualization, methodology, formal analysis, data curation, writing – original draft, visualization, project administration
J.H. Krkylywy: conceptualization, methodology, software, formal analysis, data curation, writing – original draft, visualization
A.O. Beukers: methodology, software, formal analysis, writing – review & editing
S.R. Moore: conceptualization, software, project administration
B.J. Forys: formal analysis, data curation, writing – review & editing
A.K. Anderson: conceptualization, methodology, resources, writing – review & editing, supervision, funding acquisition
R.M. Todd: conceptualization, methodology, resources, writing – original draft, supervision, funding acquisition

Conflict of Interest. The authors declare no competing financial interests.

References

Anderson, A. K., & Phelps, E. A. (2002). Is the human amygdala critical for the subjective experience of emotion? Evidence of intact dispositional affect in patients with amygdala lesions. *Journal of Cognitive Neuroscience*, *14*(5), 709–720. https://doi.org/10.1162/08989290260138618

Andreatta, M., & Pauli, P. (2015). Appetitive vs. Aversive conditioning in humans. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00128

Brown, P. L., & Jenkins, H. M. (1968). Auto-shaping of the pigeon's key-peck. *Journal of the Experimental Analysis of Behavior*, *11*(1), 1–8. https://doi.org/10.1901/jeab.1968.11-1

Byrne, J. H. (2003). *Learning & memory* (Vol. 2). Macmillan Reference.

Chikazoe, J., Lee, D. H., Kriegeskorte, N., & Anderson, A. K. (2019). Distinct representations of basic taste qualities in human gustatory cortex. *Nature Communications*, *10*(1), 1048. https://doi.org/10.1038/s41467-019-08857-z

Clark, R. E., Manns, J. R., & Squire, L. R. (2002). Classical conditioning, awareness, and brain systems. *Trends in Cognitive Sciences*, *6*(12), 524–531. https://doi.org/10.1016/S1364-6613(02)02041-7

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, an International Journal*, *29*(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews. Neuroscience*, *3*(8), 655–666. https://doi.org/10.1038/nrn894

Diedrichsen, J., Yokoi, A., & Arbuckle, S. A. (2018). Pattern component modeling: A flexible approach for understanding the representational structure of brain activity patterns. *NeuroImage*, *180*, 119–133. https://doi.org/10.1016/j.neuroimage.2017.08.051

Epicurus | Internet Encyclopedia of Philosophy. (n.d.). Retrieved 24 October 2021, from https://iep.utm.edu/epicur/

Flagel, S. B., Akil, H., & Robinson, T. E. (2009). Individual differences in the attribution of incentive salience to reward-related cues: Implications for addiction. *Neuropharmacology*, *56 Suppl 1*, 139–148. https://doi.org/10.1016/j.neuropharm.2008.06.027

Garcia, J., & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science*, *4*(1), 123–124. https://doi.org/10.3758/BF03342209

Giesecke, T., Gracely, R. H., Grant, M. A. B., Nachemson, A., Petzke, F., Williams, D. A., & Clauw, D. J. (2004). Evidence of augmented central pain processing in idiopathic chronic low back pain. *Arthritis and Rheumatism*, *50*(2), 613–623. https://doi.org/10.1002/art.20063

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics*, *7*(1), 37–53. https://doi.org/10.1007/s12021-008-9041-y

Holland, P. C. (2008). Cognitive versus stimulus-response theories of learning. *Learning & Behavior*, *36*(3), 227–241.

Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. Appleton-Century.

Jin, J., Zelano, C., Gottfried, J. A., & Mohanty, A. (2015). Human Amygdala Represents the Complete Spectrum of Subjective Valence. *Journal of Neuroscience*, *35*(45), 15145–15156. https://doi.org/10.1523/JNEUROSCI.2450-15.2015

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(11), 4302–4311.

Kim, J., Pignatelli, M., Xu, S., Itohara, S., & Tonegawa, S. (2016). Antagonistic negative and positive neurons of the basolateral amygdala. *Nature Neuroscience*, *19*(12), 1636–1646. https://doi.org/10.1038/nn.4414

Kragel, P. A., Kano, M., Van Oudenhove, L., Ly, H. G., Dupont, P., Rubio, A., Delon-Martin, C., Bonaz, B. L.,

　　Manuck, S. B., Gianaros, P. J., Ceko, M., Reynolds Losin, E. A., Woo, C.-W., Nichols, T. E., &

　　Wager, T. D. (2018). Generalizable representations of pain, cognitive control, and negative

　　emotion in medial frontal cortex. *Nature Neuroscience*, *21*(2), 283–289.

　　https://doi.org/10.1038/s41593-017-0051-7

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual

　　pathway: An expanded neural framework for the processing of object quality. *Trends in*

　　Cognitive Sciences, *17*(1), 26–49. https://doi.org/10.1016/j.tics.2012.10.011

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation,

　　and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412.

　　https://doi.org/10.1016/j.tics.2013.06.007

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis—Connecting

　　the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*.

　　https://doi.org/10.3389/neuro.06.004.2008

Kryklywy, J. H., Ehlers, M. R., Anderson, A. K., & Todd, R. M. (2020). From Architecture to Evolution:

　　Multisensory Evidence of Decentralized Emotion. *Trends in Cognitive Sciences*, *24*(11), 916–929.

　　https://doi.org/10.1016/j.tics.2020.08.002

Kryklywy, J. H., Ehlers, M. R., Beukers, A. O., Moore, S. R., Todd, R. M., & Anderson, A. K. (2021).

　　Decoding representations of discriminatory and hedonic information during appetitive and

　　aversive touch. *BioRxiv*, 2020.09.24.310383. https://doi.org/10.1101/2020.09.24.310383

Kryklywy, J. H., Forys, B. J., & Todd, R. M. (2021). *Pattern Component Modeling for R*.

　　https://github.com/bf777/PCMforR

Kuhn, M., Gerlicher, A., & Lonsdorf, T. (2021). *Navigating the manifold of skin conductance response quantification approaches – a direct comparison of Trough-to-Peak, Baseline-correction and model-based approaches in Ledalab and PsPM*. PsyArXiv. https://doi.org/10.31234/osf.io/9h2kd

Kundu, P., Brenowitz, N. D., Voon, V., Worbe, Y., Vértes, P. E., Inati, S. J., Saad, Z. S., Bandettini, P. A., & Bullmore, E. T. (2013). Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(40), 16187–16192. https://doi.org/10.1073/pnas.1301725110

Kundu, P., Inati, S. J., Evans, J. W., Luh, W.-M., & Bandettini, P. A. (2012). Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *NeuroImage*, *60*(3), 1759–1770. https://doi.org/10.1016/j.neuroimage.2011.12.028

Kurth, F., Zilles, K., Fox, P. T., Laird, A. R., & Eickhoff, S. B. (2010). A link between the systems: Functional differentiation and integration within the human insula revealed by meta-analysis. *Brain Structure & Function*, *214*(5–6), 519–534. https://doi.org/10.1007/s00429-010-0255-z

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A Unified Framework to Quantify the Credibility of Scientific Findings. *Advances in Methods and Practices in Psychological Science*, *1*(3), 389–402. https://doi.org/10.1177/2515245918787489

LeDoux, J. (2003). The emotional brain, fear, and the amygdala. *Cellular and Molecular Neurobiology*, *23*(4–5), 727–738. https://doi.org/10.1023/a:1025048802629

Lim, S.-L., O'Doherty, J. P., & Rangel, A. (2013). Stimulus value signals in ventromedial PFC reflect the integration of attribute value signals computed in fusiform gyrus and posterior superior temporal gyrus. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(20), 8729–8741. https://doi.org/10.1523/JNEUROSCI.4809-12.2013

López-Solà, M., Pujol, J., Hernández-Ribas, R., Harrison, B. J., Ortiz, H., Soriano-Mas, C., Deus, J., Menchón, J. M., Vallejo, J., & Cardoner, N. (2010). Dynamic assessment of the right lateral

frontal cortex response to painful stimulation. *NeuroImage*, *50*(3), 1177–1187.

https://doi.org/10.1016/j.neuroimage.2010.01.031

Maren, S. (2001). Neurobiology of Pavlovian fear conditioning. *Annual Review of Neuroscience*, *24*, 897–

931. https://doi.org/10.1146/annurev.neuro.24.1.897

Martin-Soelch, C., Linthicum, J., & Ernst, M. (2007). Appetitive conditioning: Neural bases and

implications for psychopathology. *Neuroscience & Biobehavioral Reviews*, *31*(3), 426–440.

https://doi.org/10.1016/j.neubiorev.2006.11.002

McNamee, D., Rangel, A., & O'Doherty, J. P. (2013). Category-dependent and category-independent

goal-value codes in human ventromedial prefrontal cortex. *Nature Neuroscience*, *16*(4), 479–

485. https://doi.org/10.1038/nn.3337

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2018). *BayesFactor: Computation*

of Bayes Factors for Common Designs (0.9.12-4.2) [Computer software]. https://CRAN.R-

project.org/package=BayesFactor

Nieuwenhuys, R. (2012). The insular cortex: A review. *Progress in Brain Research*, *195*, 123–163.

https://doi.org/10.1016/B978-0-444-53860-4.00007-6

Onat, S., & Büchel, C. (2015). The neuronal basis of fear generalization in humans. *Nature*

NEUROSCIENCE, *18*(12), 11.

Orenius, T. I., Raij, T. T., Nuortimo, A., Näätänen, P., Lipsanen, J., & Karlsson, H. (2017). The interaction of

emotion and pain in the insula and secondary somatosensory cortex. *Neuroscience*, *349*, 185–

194. https://doi.org/10.1016/j.neuroscience.2017.02.047

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral*

cortex (pp. xv, 430). Oxford Univ. Press.

Pessoa, L., Medina, L., Hof, P. R., & Desfilis, E. (2019). Neural architecture of the vertebrate brain: Implications for the interaction between emotion and cognition. *Neuroscience & Biobehavioral Reviews*, *107*, 296–312. https://doi.org/10.1016/j.neubiorev.2019.09.021

Posse, S., Wiese, S., Gembris, D., Mathiak, K., Kessler, C., Grosse-Ruyken, M. L., Elghahwagi, B., Richards, T., Dager, S. R., & Kiselev, V. G. (1999). Enhancement of BOLD-contrast sensitivity by single-shot multi-echo functional MR imaging. *Magnetic Resonance in Medicine*, *42*(1), 87–97. https://doi.org/10.1002/(sici)1522-2594(199907)42:1<87::aid-mrm13>3.0.co;2-o

Schoenbaum, G., & Roesch, M. (2005). Orbitofrontal cortex, associative learning, and expectancies. *Neuron*, *47*(5), 633–636. https://doi.org/10.1016/j.neuron.2005.07.018

Schwabe, L., Oitzl, M. S., Philippsen, C., Richter, S., Bohringer, A., Wippich, W., & Schachinger, H. (2007). Stress modulates the use of spatial versus stimulus-response learning strategies in humans. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *14*(1), 109–116. https://doi.org/10.1101/lm.435807

Sihvola, J., & Engberg-Pedersen, T. (1998). *The Emotions in Hellenistic Philosophy*. Springer Science & Business Media.

Smith, K. S., & Graybiel, A. M. (2016). Habit formation. *Dialogues in Clinical Neuroscience*, *18*(1), 33–43.

Spence, K. W. (1950). Cognitive versus stimulus-response theories of learning. *Psychological Review*, *57*(3), 159–172. https://doi.org/10.1037/h0058250

Triscoli, C., Ackerley, R., & Sailer, U. (2014). Touch Satiety: Differential Effects of Stroking Velocity on Liking and Wanting Touch Over Repetitions. *PLoS ONE*, *9*(11). https://doi.org/10.1371/journal.pone.0113425

VanElzakker, M. B., Kathryn Dahlgren, M., Caroline Davis, F., Dubois, S., & Shin, L. M. (2014). From Pavlov to PTSD: The extinction of conditioned fear in rodents, humans, and anxiety disorders. *Neurobiology of Learning and Memory*, *113*, 3–18. https://doi.org/10.1016/j.nlm.2013.11.014

Visser, R. M., Kunze, A. E., Westhoff, B., Scholte, H. S., & Kindt, M. (2015). Representational similarity analysis offers a preview of the noradrenergic modulation of long-term fear memory at the time of encoding. *Psychoneuroendocrinology*, *55*, 8–20. https://doi.org/10.1016/j.psyneuen.2015.01.021

Visser, R. M., Scholte, H. S., Beemsterboer, T., & Kindt, M. (2013). Neural pattern similarity predicts long-term fear memory. *Nature Neuroscience*, *16*(4), 388–390. https://doi.org/10.1038/nn.3345

Visser, R. M., Scholte, H. S., & Kindt, M. (2011). Associative Learning Increases Trial-by-Trial Similarity of BOLD-MRI Patterns. *Journal of Neuroscience*, *31*(33), 12021–12028. https://doi.org/10.1523/JNEUROSCI.2178-11.2011

Winecoff, A., Clithero, J. A., Carter, R. M., Bergman, S. R., Wang, L., & Huettel, S. A. (2013). Ventromedial Prefrontal Cortex Encodes Emotional Value. *The Journal of Neuroscience*, *33*(27), 11032–11039. https://doi.org/10.1523/JNEUROSCI.4317-12.2013

Yamamoto, T., Shimura, T., Sako, N., Yasoshima, Y., & Sakai, N. (1994). Neural substrates for conditioned taste aversion in the rat. *Behavioural Brain Research*, *65*(2), 123–137. https://doi.org/10.1016/0166-4328(94)90097-3