nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01639-0

A mesocorticolimbic signature of pleasure in the human brain

Received: 27 August 2022

Accepted: 22 May 2023

Published online: 29 June 2023

Check for updates

Philip A. Kragel[1,2] ✉, Michael T. Treadway[1,2], Roee Admon[3,4], Diego A. Pizzagalli[3] & Evan C. Hahn[1]

Pleasure is a fundamental driver of human behaviour, yet its neural basis remains largely unknown. Rodent studies highlight opioidergic neural circuits connecting the nucleus accumbens, ventral pallidum, insula and orbitofrontal cortex as critical for the initiation and regulation of pleasure, and human neuroimaging studies exhibit some translational parity. However, whether activation in these regions conveys a generalizable representation of pleasure regulated by opioidergic mechanisms remains unclear. Here we use pattern recognition techniques to develop a human functional magnetic resonance imaging signature of mesocorticolimbic activity unique to states of pleasure. In independent validation tests, this signature is sensitive to pleasant tastes and affect evoked by humour. The signature is spatially co-extensive with mu-opioid receptor gene expression, and its response is attenuated by the opioid antagonist naloxone. These findings provide evidence for a basis of pleasure in humans that is distributed across brain systems.

Pleasure is central to human experience, and has served as a cornerstone for philosophical, socio-economic and psychological frameworks for understanding human behaviour for thousands of years[1]. Despite its centrality for daily life and philosophical systems alike, the neuroscientific understanding of pleasure in the human brain remains in its infancy[2,3]. This stands in stark contrast to the study of human 'reward', which has largely focused on identifying neural systems that mediate behavioural responses to reinforcing stimuli[4]. Such paradigms have yielded crucial insights into the circuitry underlying conditioning, learning and decision making, yet it is widely accepted that such behavioural manifestations of preference do not necessarily reflect the direct experience of pleasure[5], and that pleasure is not necessary for reinforcement[4,6]. As such, the neural basis for subjective pleasure remains elusive.

Indeed, much of the known functional neuroanatomy of pleasure in mammals[2,3] has been derived from studies in rodents, which have established a critical role for mu-opioid signalling within a network of regions including the nucleus accumbens (NAc) shell, ventral pallidum (VeP), orbitofrontal cortex (OFC) and insula[3,7,8]. Critically, microinjections of mu-opioid agonists into specific zones in these areas, referred to as hedonic 'hotspots', enhance putatively pleasure-related behaviours involving relaxed facial expressions and rhythmic tongue and mouth movements[3,7–13]. These same injections also suppress pleasure-related behaviours in neighbouring hedonic 'coldspots'[9,14]. Importantly, the role of the mu-opioid system has been proposed as selective to pleasure-related behaviours and neurobiologically dissociable from putatively dopaminergic aspects of behavioural reinforcement, such as conditioning, craving and invigoration[4,14].

Attempts to translate this pre-clinical literature to humans have been mixed. On the one hand, human neuroimaging studies have shown that the same set of regions are commonly activated by diverse rewards[15–17], with positron emission tomography further highlighting the involvement of endogenous opioids[18]. However, the functional homology of this network in humans and circuits identified in rodents is contested—particularly in prefrontal cortex and insula[19,20]. Moreover, studies of opioid antagonism on pleasure responses have revealed inconsistent effects, with some evidence suggesting that opioid antagonism may exhibit more effects on motivation than pleasure in

[1]Department of Psychology, Emory University, Atlanta, GA, USA. [2]Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA, USA. [3]Department of Psychiatry, Harvard Medical School and McLean Hospital, Belmont, MA, USA. [4]School of Psychological Sciences, University of Haifa, Haifa, Israel. ✉e-mail: pkragel@emory.edu

humans[21]. Finally, the nature of pleasurable stimuli accessible for study in animal models does not extend to many important modalities of human pleasure, such as music and humour. Consequently, the generalizability of rodent models for understanding the full complement of human pleasures is uncertain.

Uncertainty about hedonic brain systems in humans is also due to the size and spatial configuration of affective circuitry in subcortical structures, as well as the limits of conventional imaging approaches. In rodents, hedonic hotspots in the NAc form a spatial gradient in which anterior areas are involved in appetitive and posterior areas in aversive behaviours[22]. A mirrored gradient is present in the VeP, in which activation of caudal areas enhances appetitive behaviours and rostral areas enhance avoidant behaviours[11]. These hotspots comprise only a small portion (~10%) of the subcortical structures in which they are situated, which contain functionally and neurochemically heterogeneous neural populations[23,24]. Accordingly, when assessed with conventional functional magnetic resonance imaging (fMRI), signals from different neural populations are blurred together, obscuring which affective variables (for example, autonomic arousal and reward value) are encoded in each region. This has limited efforts to characterize hedonic brain systems in humans, making it difficult to isolate neural substrates involved in different components of reward.

Due to these limitations, a growing number of researchers have turned to multivariate approaches to evaluate how affective variables are represented in human brain activity[25,26]. Unlike standard univariate analysis, multivariate methods are capable of estimating a spatial profile of activity within and across regions that characterizes a variable of interest[27,28], even in cases where multiple neural populations overlap in a single region. Indeed, pattern-based methods have revealed responses in orbitofrontal cortex and adjacent ventromedial prefrontal cortex that discriminate states of pleasure from displeasure[26]. Thus far, however, there is surprisingly limited evidence that neural populations in subcortical structures represent diverse pleasures using a common code, or that they form part of a distributed network mediated by opioidergic mechanisms positioned to influence learning and decision making as predicted by contemporary accounts of reward learning[4]. Moreover, because brain areas consistently engaged by rewards also respond to a wide array of motivationally salient, aversive and painful stimuli[29], it is unclear whether these regions contain circuitry that regulates pleasure across contexts as opposed to other non-specific factors such as motivational salience or arousal[14,29,30].

Understanding hedonic systems in humans has implications for translational research since altered regional activity in the ventral striatum, basal forebrain and amygdala have emerged as candidate biomarkers for many neuropsychiatric disorders[31,32] and are common targets for clinical interventions[33–35]. Although specific regions of interest are often well motivated by pre-clinical research, recent work developing brain-based biomarkers of affective processes[36,37] has shown that univariate measures often produce smaller effects[38], and are less accurate and reliable than multivariate predictive models[39–41].

In this Article, we aim to more precisely model human brain responses to diverse pleasures, with a specific focus on regions known to contain hotspots in rodents. We combined a mega-analytic approach and pattern recognition techniques to characterize brain responses across 28 fMRI studies (total $n = 494$, Methods). We used patterns of brain activity from ten studies that manipulated pleasure using music, images of appetizing food, erotic images, cues of monetary rewards, and socially relevant stimuli (two studies of each type, total $n = 224$) to develop an fMRI-based model, or brain signature, that predicts the hedonic state of an individual. Brain responses during manipulations of positive affect were differentiated from those acquired during affectively salient, but not pleasurable experiences (18 studies, $n = 270$)[42].

Modelling brain activity across diverse experimental manipulations enabled us to tease apart signals that are consistent across instances of pleasure from those that are bound to a single stimulus or sensory modality[43]. By focusing on patterns of brain activity, we move beyond individual brain regions to characterize fine-grained topographies in distributed mesocorticolimbic circuitry that are defining features of hedonic systems. Training the model on data from 28 independent studies facilitated prospective tests of two key predictions from theories of hedonic function: (1) that a signature for pleasure should track the hedonic component of reward, as opposed to motivational or learning components, and further (2) that the response of such a signature should be sensitive to the distribution and manipulation of opioidergic circuits. Following training, we verified that the signature was sensitive and specific to pleasure in four fMRI studies (total $n = 89$), two that evoked states of pleasure using primary reinforcers and two that manipulated prospects for subsequent monetary reward but had relatively less hedonic impact. Further, we evaluated opioid involvement in the signature response during a pharmacological challenge using the antagonist naloxone ($n = 19$). These validation tests provide a strong assessment of the hypothesis that human pleasure is partially mediated by a distributed opioidergic network.

Results

Towards a brain signature for pleasure

We used a latent variable multivariate regression technique, partial least squares (PLS) regression[44], to predict states of pleasure from patterns of fMRI activity. This approach estimates the spatial layout and activation of multiple latent sources that explain both observed fMRI activity and affective variables of interest[42]. The model included signals from brain regions known to contain hedonic hotspots and interconnected areas involved in emotion, motivation and reward processing[45] (for details, see Methods). We identified a pattern across these areas that predicts pleasure independent of its sensory origin (whether music, images of food, erotic images, monetary reward or social stimuli) while accounting for spatially overlapping signals that do not generalize across instances of pleasure or signals that are shared with other affective states (findings from cross-validation in the training dataset are shown in Extended Data Fig. 1).

The signature contained positive coefficients, which estimate the spatial layout of population activity associated with pleasure, in multiple regions including the anterior ventral insula, anterior agranular insula, midcingulate cortex, dorsomedial prefrontal cortex, basolateral amygdala, extended amygdala, globus pallidus, ventral striatum and substantia nigra (voxel-wise $q_{FDR} < 0.05$, Fig. 1a and Supplementary Tables 1 and 2). Negative coefficients were present in posterior insula, dorsal mid-insula, midcingulate cortex and supplementary motor area. In line with evidence of interdigitated population activity during appetitive and aversive behaviours[46,47], most regions contained both positive and negative coefficients (Fig. 1b). Multiple regions differed in their balance of positive and negative coefficients. The habenula, anterior ventral insula, basolateral amygdala, internal globus pallidus and midbrain regions contained more positive coefficients, whereas middle and posterior insula contained primarily negative coefficients ($q_{FDR} < 0.05$, Supplementary Table 2).

Given evidence of both positive and negative signature coefficients in most regions, we next examined whether coefficients exhibited gradients similar to those identified in nonhuman animals. A mediolateral gradient was present in the NAc with larger coefficients in more medial portions that decreased laterally ($\hat{\beta} = -0.0135$, $z = -2.34$, $P = 0.0195$, 95% confidence interval −0.0249 to −0.0021), linear regression between coefficients and their distance from midline; Fig. 1c). This gradient is consistent with human neuroimaging evidence that rewards activate medial portions of the NAc whereas aversive stimuli activate more lateral areas[48,49]. Fine-grained topography was also present within the midcingulate cortex. Coefficients exhibited a clear peak near the bank of the callosal sulcus ($z = 4.06$, Montreal Neurological Institute coordinate ($MNI_{x,y,z}$) = [6, 8, 26], $P < 0.0001$, $q_{FDR} < 0.05$), whereas

https://doi.org/10.1038/s41562-023-01639-0

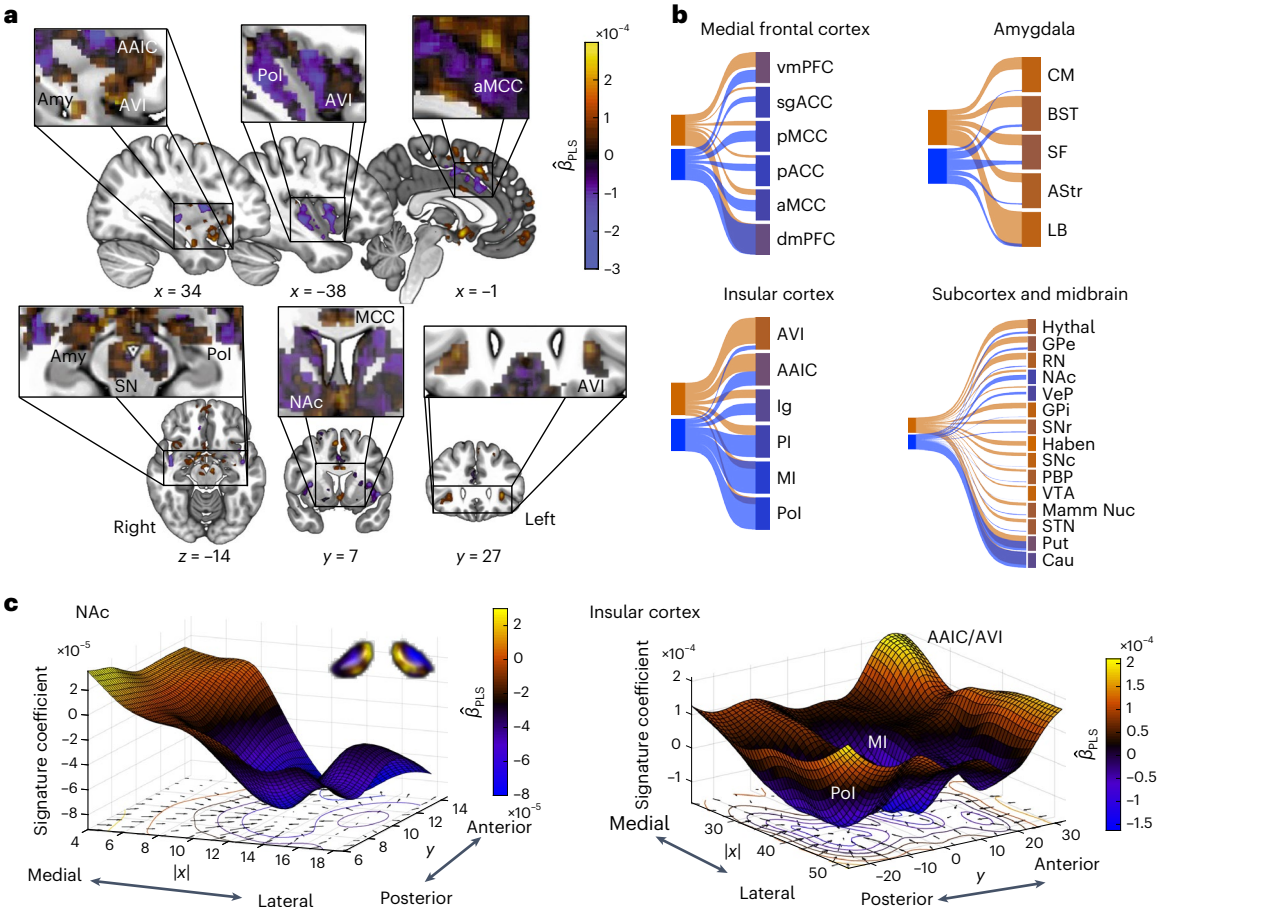


Fig. 1 | An fMRI-based signature for pleasure. a, PLS regression coefficients that define the signature. Warm colours indicate regions in which increases in brain activity contribute to predictions of pleasure, whereas cool colours indicate regions in which brain activity decreases classifications of pleasure. Extreme coefficients (magnitude >0.0001) are rendered in volumetric space, and heat map overlays show unthresholded regression coefficients. Amy, amygdala; AAIC, anterior agranular insula complex; AVI, anterior ventral insular area; PoI, posterior insular area; aMCC, anterior midcingulate cortex; SN, substantia nigra; NAc, nucleus accumbens. **b**, Alluvial flow plots depict the similarity of coefficients and anatomically defined regions of interest. Positive coefficients are depicted in orange and negative coefficients in blue. vmPFC, ventromedial prefrontal cortex; sgACC, subgenual anterior cingulate cortex; pMCC, posterior midcingulate cortex; pACC, perigenual anterior cingulate cortex; aMCC, anterior midcingulate cortex; dmPFC, dorsomedial prefrontal cortex; CM, centromedial amygdala; BST, bed nuclei of the stria terminalis; AStr, amygdalostriatal transition area; LB, basolateral amygdala; Ig, insular granular complex; PI, para-insular area; MI, middle insular area; Hythal, hypothalamus; GPe, external globus pallidus; RN, red nucleus; VeP, ventral pallidum; GPi, internal globus pallidus; SNr, substantia nigra pars reticulata; Haben, habenular nuclei; SNc, substantia nigra pars compacta; PBP, parabrachial pigmented nucleus; VTA, ventral tegmental area; Mamm Nuc, mammillary nucleus; STN, subthalamic nucleus; Put, putamen; Cau, caudate. **c**, Spatial topography of coefficients in the NAc (left) and insular cortex (right). Surfaces depict topographies estimated with fitting thin-plate smoothing splines using the *x* and *y* coordinates of signature coefficients. Contours and vector fields of the surface gradient in *x* and *y* dimensions are depicted below each surface.

predominantly negative coefficients were present throughout the remainder of the midcingulate gyrus. A posterior-to-anterior gradient was present in the insular cortex ($\hat{\beta} = 0.0053$, $P < 0.0001$, $z = 4.63$, 95% confidence interval 0.0031 to 0.0075), with a peak in ventral anterior insula ($z = 4.80$, $MNI_{x,y,z} = [-30, 26, -2]$, $P < 0.0001$, $q_{FDR} < 0.05$; Fig. 1c). Rostro-caudal gradients similar to those observed in rodent studies[3] were not apparent in the VeP ($\hat{\beta} = -0.0100$, $P = 0.4815$, $z = -0.704$, 95% confidence interval −0.0335 to 0.0135), although the sign and organization of coefficients were roughly consistent with this layout (Extended Data Fig. 2). This is possibly due to the small size of this region, spanning roughly 6 mm along its rostro-caudal axis relative to the smoothness of model coefficients, estimated at ~4.5 mm full-width half-maximum (Extended Data Fig. 3)[50]. These findings demonstrate that, although the NAc, midcingulate and anterior insula are consistently activated by aversive, rewarding and salient stimuli[29], states of pleasure are characterized by unique patterns of activity within each of these regions, consistent with pre-clinical findings[3,12].

Validating the signature in independent studies

Accounts of hedonic function propose that certain stimuli, situations and behaviours are rewarding because they evoke pleasure[6,51], which is thought to be mediated by a distributed mesocorticolimbic network[3]. Evolutionary theories of pleasure go further to suggest that there is a final common pathway in which diverse pleasures are expressed and represented similarly[52]. If the signature we identified captures such a common representation, then it should generalize across instances of pleasure whether they are produced by basic sensory stimuli or more complex cognitive processes. To test this prediction, we evaluated the signature response in multiple independent archival datasets. We first applied the signature to brain activity measured as participants consumed commercial beverages that ranged from hedonically neutral to mildly pleasant[53]. The signature response to the most pleasing beverage (self-report: 5.32 ± 0.186 standard error of the mean (s.e.m.), pattern response: 0.0327 ± 0.0113 s.e.m.) and a beverage rated as neutral (self-report: 3.93 ± 0.274 s.e.m., pattern response: −0.00265 ± 0.0146

https://doi.org/10.1038/s41562-023-01639-0

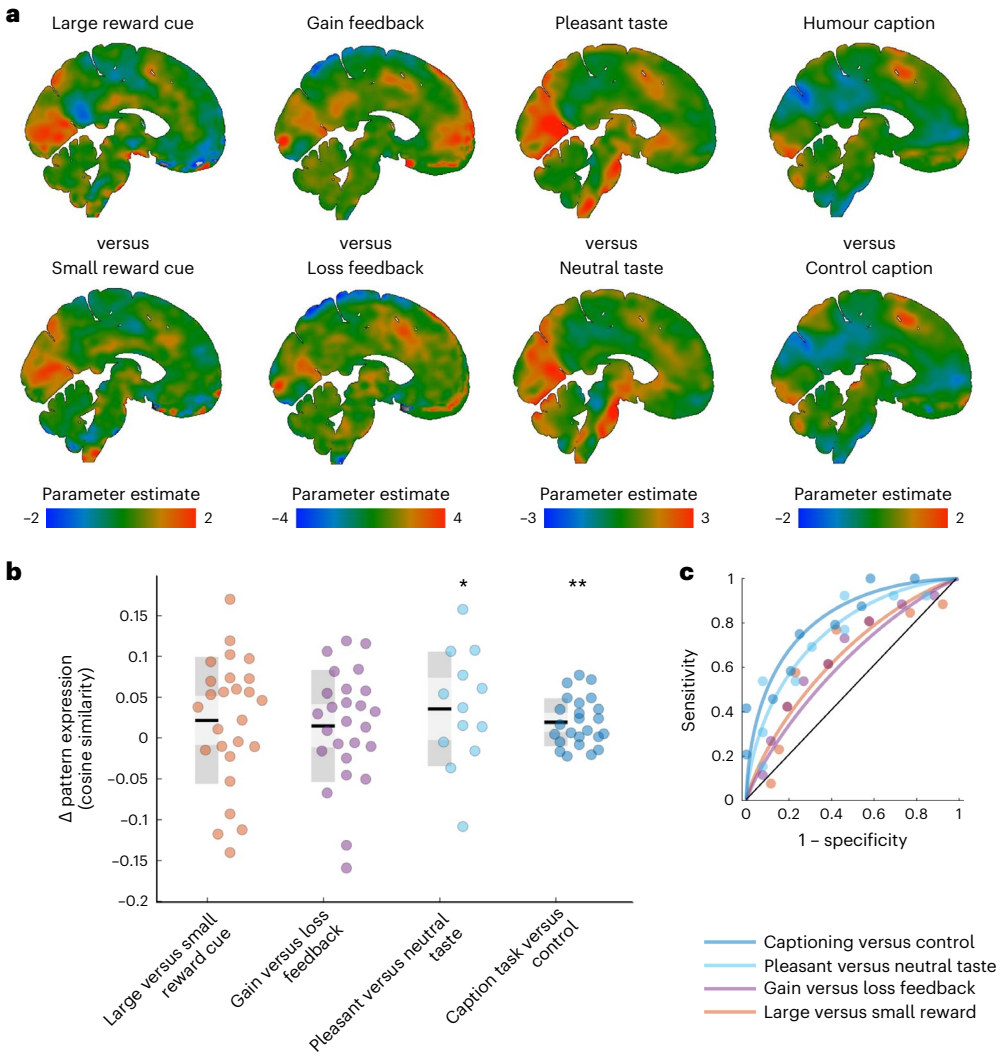


Fig. 2 | Validation of the pleasure signature. a, Group-average activation maps for four independent validation datasets (n = 26, 26, 13, 24 independent participants). Warm colours indicate increases and cool colours indicate decreases in brain activity during each condition compared with baseline. Pairs of experimental conditions are ordered by the predicted difference in signature response. **b**, Box-and-whisker plot shows differences in the signature response for each study. Black lines depict the mean response, light-shaded regions depict one standard deviation and darker-shaded regions depict two standard errors. Each point corresponds to the response of a single subject (n = 26, 26, 13, 24 independent participants; *mean 0.0214, t_{12} = 1.826, P = 0.0929, d = 0.506, 95% confidence interval −0.0085 to 0.0512, **mean 0.0191, t_{23} = 3.191, P = 0.0041, d = 0.651, 95% confidence interval 0.0074 to 0.0308, uncorrected two-sided paired t-tests). **c**, Receiver operating characteristic curves for each of the four studies.

s.e.m.) were discriminable with a medium effect size (area under the receiver operating characteristic curve (AUROC) = 0.80, d = 0.72, P = 0.0575, 95% confidence interval 0.624 to 0.973, Fig. 2), providing initial evidence of generalizability.

In a second generalization test, we examined the signature response during a positive mood induction that combines humour, decision making and social feedback[54]. In this task, participants viewed cartoons from the New Yorker Caption Contest and were instructed to select which caption was selected as the winner[55]. Regardless of their selection, participants received feedback that they had chosen the correct option in the majority of trials. We compared the signature response with humour captioning and a matched control task (descriptive captioning) that equated the perceptual, decision making and motor aspects of captioning but did not involve humour. The signature response to humour captioning (0.137 ± 0.0119 s.e.m.) and the control task (0.1180 ± 0.0111 s.e.m.) were discriminable from one another with a large effect size (AUROC 0.82, P = 0.0035, d = 0.92, 95% confidence interval 0.704 to 0.944), providing further evidence of generalizability.

The sensitivity of the signature to both pleasant tastes and humour suggests it may reflect common coding of pleasure. However, its response to these stimuli could be driven by variables correlated with hedonic impact in the training data, rather than pleasure per se. In particular, the signature may have capitalized on signals related to incentive salience and/or reward value[56] to classify brain states associated with pleasure. On the other hand, if the signature captures processing predominantly related to pleasure, then it should provide good discriminability of hedonic experiences, and worse discriminability of less hedonically impactful monetary stimuli. To examine this possibility, we tested whether the signature is sensitive to differences in brain activity as participants viewed reward-predictive cues and during reward receipt (that is, feedback about monetary gains and losses) that differed in terms of reward value but produced only minimal differences in subjective pleasure.

We evaluated the signature response as participants performed an effort-based decision-making task that used visual cues to indicate the magnitude of rewards and the physical effort required to obtain them.

We compared responses to cues on low-effort trials (<30% of maximal effort during a calibration procedure) that indicated participants could receive a large reward (US$5) to those indicative of small rewards (US$1). Focusing on these conditions ensured that differences in the signature response were related to reward magnitude, rather than the difficulty of the decision or negative value associated with effort. The signature response to large (0.0429 ± 0.0110 s.e.m.) and small reward cues (0.0215 ± 0.0127 s.e.m.) were only modestly discriminable from one another, with a small effect size (AUROC 0.68, $P = 0.0830$, $d = 0.39$, 95% confidence interval 0.535 to 0.827).

To further evaluate the possibility that the signature is sensitive to differences in value, we examined its response during a classic reinforcement learning task. In this task, participants learned which of several cues were associated with monetary gains and losses to maximize monetary reward[57]. Computational models of reinforcement learning have demonstrated that momentary changes in positive affect are associated with positive reward prediction errors[58], and are generally uncorrelated with objective reward magnitude. On the basis of this evidence, we predicted that if the brain signature responds to the hedonic impact of a stimulus, rather than its value or motivational salience, then it should weakly differentiate visual feedback about gains and losses. Consistent with this prediction, we found the signature response to gains (0.0435 ± 0.0152 s.e.m.) and losses (0.0288 ± 0.0129 s.e.m.) exhibited low levels of discriminability (AUROC 0.68, $P = 0.0746$, $d = 0.30$, 95% confidence interval 0.533 to 0.825).

To test the prediction that the signature should respond more strongly to hedonic experiences than information about monetary reward, we next examined the response of the signature across studies. Across the four validation studies, the signature responded more strongly to differences in pleasure compared with monetary rewards ($\hat{\beta} = 0.0248$, $t_{173} = 2.358$, $P = 0.0195$, 95% confidence interval 0.0042 to 0.0454; linear mixed-effects regression, Methods). Because functional gradients are a defining feature of hedonic systems, and multivariate predictive models can use information at multiple spatial scales, we additionally tested whether fine-grained patterns within regions that define the signature are necessary for accurate prediction. We constructed a model that replaced each coefficient in the signature with the average of all voxels in each region. Repeating the four validation tests revealed this constrained model did not respond more strongly to states of pleasure than other conditions during manipulations of reward ($\hat{\beta} = 0.0057$, $t_{173} = 0.4681$, $P = 0.6403$, 95% confidence interval −0.0188 to 0.0294, Extended Data Fig. 4), demonstrating that variation in fMRI response within regions is necessary for identifying states of pleasure.

To verify that the pleasure signature is functionally dissociable from evaluative and anticipatory components of reward, we compared the pleasure signature with a recently developed brain signature designed to discriminate between monetary rewards and losses[37]. Whereas the pleasure signature was most sensitive to pleasant taste and humour, the reward signature robustly discriminated gain and loss outcomes during reinforcement learning (AUROC 0.85, $P = 0.0010$, $d = 0.89$, 95% confidence interval 0.743 to 0.957) and failed to differentiate states of pleasure from other conditions ($\hat{\beta} = 0.004620$, $t_{173} = 0.6271$, $P = 0.5314$, 95% confidence interval −0.0082 to 0.0206, Extended Data Fig. 5). Further, signatures trained to classify brain states associated with other functional domains (pain, cognitive control and negative affect) did not accurately discriminate states of pleasure from matched control conditions (Supplementary Table 3).

Cortical and subcortical representations of pleasure

Although it is widely accepted that many sensory cortical areas exhibit a high degree of modularity (that is, functional specificity) that can be consistently detected with fMRI, the extent to which brain areas encoding affect exhibit a similar modular organization is debated[14]. Given that our training and validation studies included a range of reward modalities from primary taste/olfaction to more abstract monetary and humour rewards, we next sought to evaluate the extent to which cortical and subcortical areas contained representations of pleasure that generalized across stimulus types. First, we performed a representational similarity analysis[59] within the training dataset in regions hypothesized to contain hedonic modules to determine which areas showed modality-specific versus domain-general coding of pleasure. As predicted by the rodent literature[10,11,13] we observed the clearest evidence for generalizable pleasure coding in the NAc ($\hat{\beta} = 0.0095$, $P = 0.0088$, $z = 2.62$, 95% confidence interval 0.0022 to 0.0168), VeP ($\hat{\beta} = 0.0124$, $P = 0.0154$, $z = 2.42$, 95% confidence interval 0.0024 to 0.0224) and ventromedial prefrontal cortex ($\hat{\beta} = 0.006$, $P = 0.0338$, $z = 2.12$, 95% confidence interval 0.000512 to 0.0115), with evidence for subdomain-specific representational geometry within insula and ventromedial prefrontal cortex (Fig. 3 and Supplementary Table 4). A replication of this analysis including all features used to train the pleasure signature revealed similar findings, with evidence for domain-specific representation of pleasure, representations related to social, musical and erotic subdomains, and several constructs from other domains (Supplementary Table 5).

Next, we used our validation datasets to determine the extent to which synthetic 'lesions' that excluded signals in cortical and subcortical areas impacted the classification of reward-related brain activity. Here we found that constraining predictive models to exclusively use signals from NAc, VeP, insula and ventromedial prefrontal cortex had little impact on the discrimination of pleasant taste (ΔAUROC −0.0680, $P = 0.6501$) and substantially impaired humour classification (ΔAUROC −0.3325, $P = 0.0010$). Conversely, excluding signals from putative hotspot regions did not impair discriminability for pleasant tastes (ΔAUROC −0.0947, $P = 0.5515$) and resulted in a modest improvement in humour classification (ΔAUROC 0.1450, $P = 0.0470$).

These results suggest that—as in rodents—putative hotspots in human NAc, VeP and OFC show generalizable coding of pleasure across a range of primary and secondary rewards. Indeed, models trained to detect states of pleasure generalized to predict pleasant tastes. However, distributed patterns in these regions were not sufficient to classify humour, which appeared to depend more strongly on the inclusion of medial prefrontal cortex—potentially reflecting the contribution of self-referential processing[2] or more broadly theory of mind[60]. Taken together, this implies a distributed architecture for pleasure encoding rather than a highly modular organization and highlights distinctions between cortical and subcortical areas for pleasure associated with primary and secondary rewards.

Evaluating opioid contributions to the signature

Opioidergic mechanisms in mesocorticolimbic structures are thought to play a central role in driving and regulating appetitive behaviour, with mu-opioids being particularly involved in hedonic components of reward[3,7–13]. If this is the case, and the signature captures the activity of opioidergic neural populations, there should be a correspondence between the magnitude of signature coefficients and the density of mu-opioid receptors. We tested this hypothesis by performing a spatial regression between the signature coefficients and neurotransmitter gene expression data from the Allen Human Brain Atlas[61]. Due to the considerable overlap of dopaminergic and opioidergic populations in striatum[24], amygdala[62] and midbrain[63], we included patterns of gene expression for dopamine receptors (*DRD1*, *DRD2* and *DRD3*) and opioid receptors (*OPRD1*, *OPRK1* and *OPRM1*) in a multiple regression to predict the signature coefficients (for spatial correlations of these maps, see Extended Data Fig. 6). Consistent with our hypothesis, this analysis revealed a positive relationship between the spatial profile of signature coefficients and *OPRM1* ($\hat{\beta} = 0.2529$, $z = 2.659$, $P = 0.0078$, 95% confidence interval 0.0665 to 0.4392). Follow-up comparisons revealed this relationship was greater than associations with the expression of other genes, on average ($\Delta\hat{\beta} = 0.2952$, $z = 2.802$,

https://doi.org/10.1038/s41562-023-01639-0

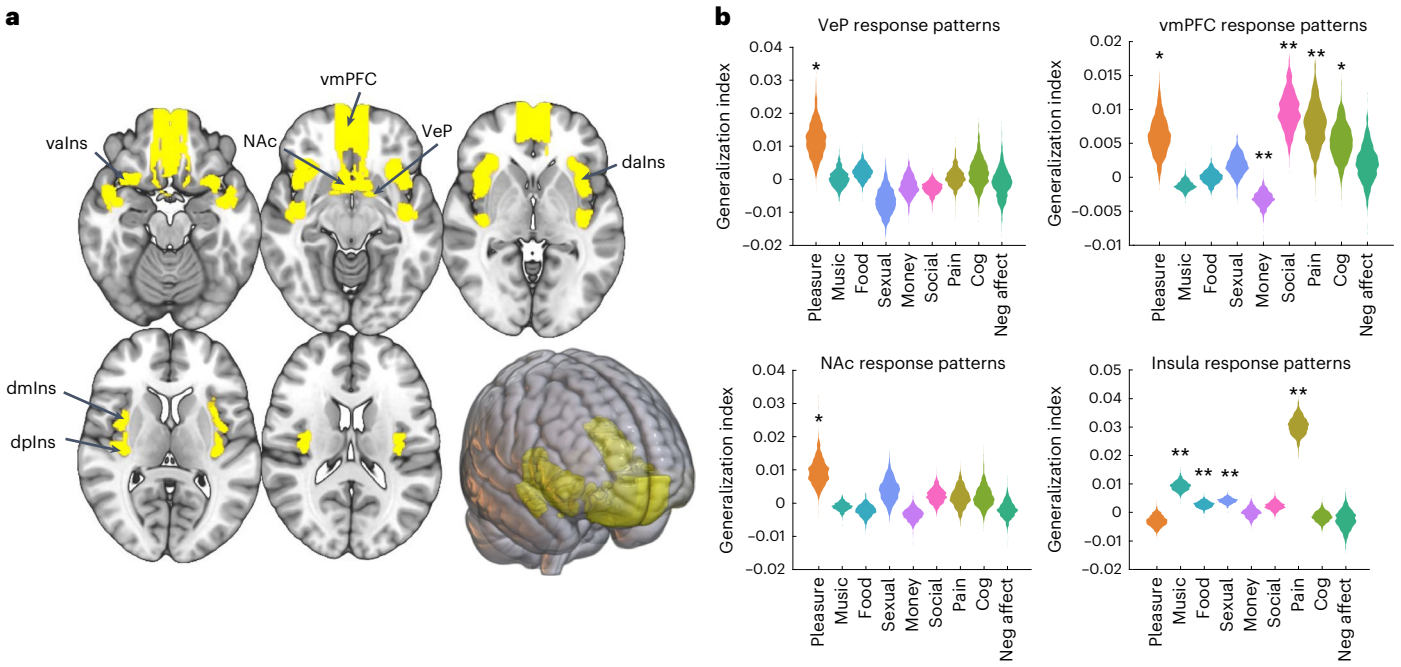


Fig. 3 | Cortical and subcortical representations of pleasure. a, Volumetric rendering of regions of interest known to contain hedonic hotspots (yellow) on the ICBM152 template. **b**, Hierarchical representational similarity analysis reveals generalizable patterning related to pleasure in the VeP, NAc and ventromedial prefrontal cortex (vmPFC). Representations specific to social pleasure were identified in vmPFC, and representations specific to pleasure induced by music, images of appetizing food and erotic images were identified in the insula. Bootstrap distributions for regression coefficients are shown for the four general domains (pleasure, pain, cognitive control (Cog) and negative (Neg) affect) and five pleasure subdomains (music, food, erotic, monetary and social pleasure). vaIns, ventral anterior insula; daIns, dorsal anterior insula; dmIns, dorsomedial insula; dpIns, dorsal posterior insula. *$P < 0.05$, **$q_{FDR} < 0.05$ (two-sided, bootstrap tests of coefficients from a spatial regression). Complete statistics are reported in Supplementary Table 5.

$P = 0.0051$, 95% confidence interval 0.0887 to 0.5016) although not larger than *DRD2* ($\Delta\hat{\beta} = 0.0915$, $z = 0.798$, $P = 0.425$, 95% confidence interval −0.1333 to 0.3162); for details, see Fig. 4 and Supplementary Table 6. Confirmatory analyses using simple correlations provided similar results, revealing a positive association with *OPRM1* ($r = 0.0659$, $z = 1.998$, $P = 0.0457$, 95% confidence interval 0.0013 to 0.1306) that was larger than the association with other gene expression maps ($\Delta r = 0.0803$, $z = 2.4529$, $P = 0.0142$, 95% confidence interval 0.0161 to 0.1445). Exploratory analyses examining spatial correlations between gene expression maps and coefficients for other domain- and subdomain-specific terms showed that coefficients for the erotic subdomain were positively correlated with *OPRM1* expression, whereas coefficients for the cognitive control domain and the working memory subdomain were associated with dopamine gene expression (Extended Data Fig. 7 and Supplementary Table 7).

The spatial correspondence between signature coefficients and gene expression is consistent with evidence that opioids mediate human pleasure[64,65]. To assess whether opioids influence the signature response, we tested it on a placebo-controlled cross-over study[66] using the opioid antagonist naloxone ($n = 19$). In this fMRI study, participants performed an incentive delay task that required participants to make speeded button presses to either obtain monetary rewards or view erotic images. The task was performed in two scanning sessions, concurrent with fMRI and the administration of intravenous naloxone or saline placebo (brain maps showing the effect of naloxone in each condition are shown in Extended Data Fig. 8). Compared with placebo, naloxone reduced self-reported pleasure from erotic images (mean difference −9.223, $z = -2.130$, $P = 0.0329$, $d = -0.4770$, 95% confidence interval −17.918 to −0.528), but not monetary rewards (mean difference −4.342, $z = -0.989$, $P = 0.3229$, $d = -0.2267$, 95% confidence interval −12.032 to 5.192). Naloxone attenuated the response of the pleasure signature with similar effect sizes, reducing its response to erotic images (mean difference −0.0376, $z = -2.037$, $P = 0.0416$, $d = -0.4927$, 95% confidence interval −0.075 to −0.00016), but not to monetary rewards (−0.0168, $z = -1.350$, $P = 0.1771$, $d = -0.3215$, 95% confidence interval −0.0081 to 0.0417), suggesting that opioids regulate the signature response and pleasure experience to a similar degree, particularly in response to primary rewards.

Discussion

Diverse forms of human pleasure are thought to be driven by brain systems that originally developed to support the attainment of basic rewards essential for survival—food, social interaction, sex and maternal care. Under such modular, pre-adaptation accounts[52], the same hedonic circuitry that mediates pleasure evoked by basic rewards has been co-opted for more abstract sources of pleasure, such as music, aesthetics and humour, which additionally involve cortical processing and are heavily influenced by learning. Our findings are broadly consistent with such accounts, as the signature we developed is sensitive to both basic sensory and abstract pleasures, and it does not respond robustly to salient, positive events that lack hedonic impact, such as a cue indicative of a potential monetary reward. Although this suggests the signature may capture activity from a common pleasure pathway, we found that the prediction of humour depended on activity in prefrontal cortex and insula, demonstrating that subcortical modules are insufficient to characterize affective experience in humans.

Following a rich history of attempts to identify neural sources of pleasure[67,68], our label of 'pleasure signature' is situated in the context of current theories of hedonic function and neuroimaging data available for training. Eight of the ten studies used to develop the signature verified that participants experienced pleasure using self-reported valence (five studies) or ratings of the appetitive nature of stimuli (three studies). The remaining training data comprised brain responses to reward cues (reflecting the magnitude and probability of gains and

https://doi.org/10.1038/s41562-023-01639-0

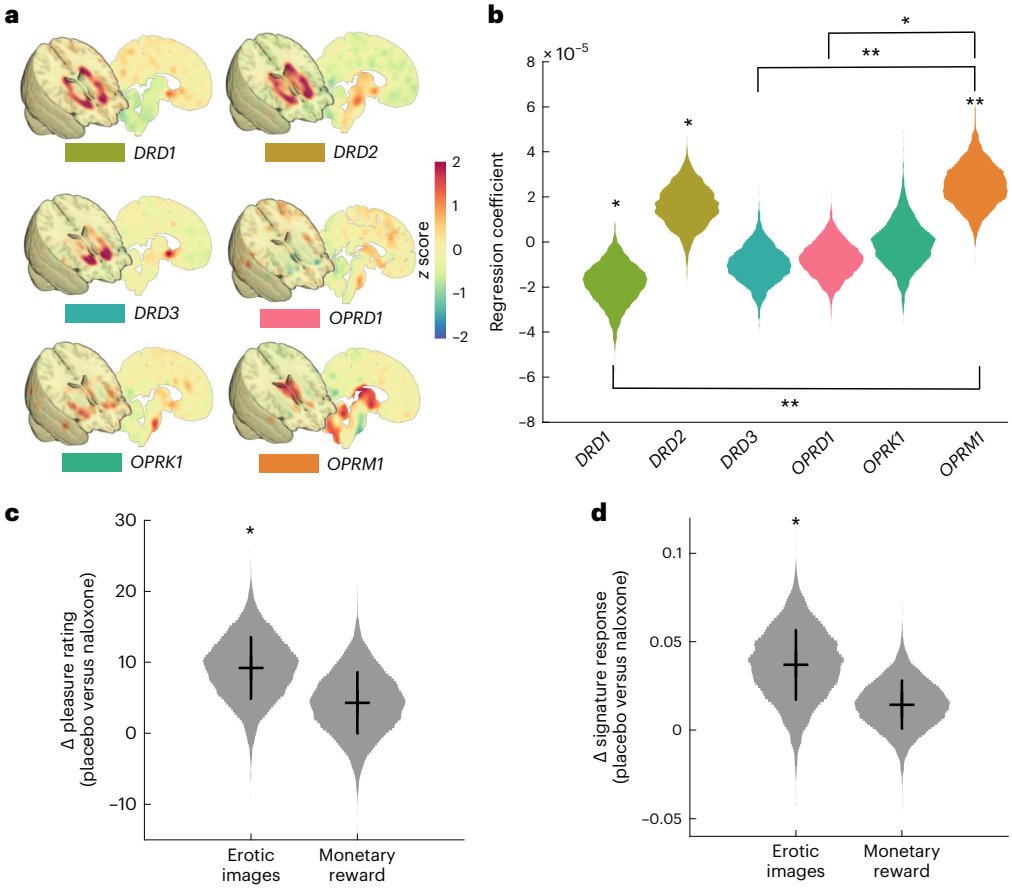


Fig. 4 | Opioid contributions to the pleasure signature. a, Normalized gene expression maps from the Allen Brain Atlas used to evaluate the spatial correspondence between neurotransmitter gene expression and the pleasure signature coefficients. **b**, Beta estimates from spatial regression indicate that *DRD2* and *OPRM1* expression was uniquely associated with coefficients of the pleasure signature (two-sided, uncorrected bootstrap tests). Complete statistics are reported in Supplementary Table 6. **c**, Naloxone challenge reduced self-reported pleasure to erotic images, but not monetary rewards ($n = 21$ independent participants, two-sided, uncorrected bootstrap tests). **d**, Signature response (cosine similarity) was lower with naloxone compared with placebo. Error bars depict the standard error of the mean ($n = 21$ independent participants, two-sided, uncorrected bootstrap tests). Grey shaded regions depict bootstrap distributions ($b = 10{,}000$). *$P < 0.05$, **$P < 0.001$.

losses in a mixed gambles task, and the magnitude of immediate and delayed monetary rewards in a delay discounting task) that typically evoke positive affect[69–71]. And even though the training data largely involved visual images (8/10 studies) and passive reward acquisition (8/10 studies), the signature generalized to pleasant taste/olfaction and a novel decision-making task involving humour. As such, the signature we developed is a step towards more precisely characterizing hedonic brain function in humans, taking the critical step of including types of pleasurable stimuli that cannot be easily studied in animals, such as music and humour.

Importantly, the ability of the model to generalize across different types of pleasurable experiences in independent samples does not necessarily suggest that pleasure is undifferentiated or that it is modular in nature. Pleasure is inherently multidimensional, with a hierarchical structure[43] in which unitary pleasure can be differentiated in terms of antecedent events, sensations and emotional responses. By design, we trained our model to characterize the apex of this hierarchy so that it would capture generalizable aspects of pleasure rather than stimulus- or situation-specific features. Future work focused on variation within and between different types of pleasure (for example, sensory, physical, aesthetic and social) is needed to determine how sensory information is transformed into a common representation.

Our results are consistent with neurobiological accounts that characterize affect as an emergent feature of coordinated population activity in distributed neural networks[3]. Rather than requiring only a single region, or depending on large-scale, global signals, we found that the signature's ability to accurately predict pleasure was driven by local topography within regions. Although prior meta-analytic summaries have proven invaluable for identifying neural correlates of affect, the present findings suggest that coordinate-based methods lack the precision necessary to discriminate positively and negatively valenced states. It will probably be necessary to move from coordinate-based assessments of the literature to pattern-based frameworks to accurately assess the brain basis of affective phenomena.

The spatial layout of the pleasure signature is consistent with prior neuroimaging summaries of positive affect, assessments of mu-opioid receptor availability[72,73], and observations of hedonic hotspots identified in rodent studies[9,12]. Beyond supporting existing descriptions of pleasure systems, it provides new insight into cortical areas not typically associated with hedonic function. For instance, we observed a peak in signature coefficients in ventral midcingulate cortex (along areas 24a′ and 33′), adjacent to the corpus callosum. Compared with dorsal aspects of the midcingulate, ventral portions of midcingulate have distinct cytoarchitecture[74] and functional connectivity[75] and are not consistently engaged by aversive and cognitively demanding tasks. Consistent with our findings, stimulation of the callosum near this area has been found to increase spontaneous expression of positive affect in awake humans[76]. Similar to variables encoded in the activity of adjacent

populations in midcingulate cortex involved in pain affect and reward, hedonic signals could be used to compute the expected value and drive learning about rewards[77], although this possibility remains to be tested.

A somewhat surprising result was the signature's relatively weak categorization of monetary reward (Fig. 2b). Monetary reward tasks are widely deployed in population neuroimaging research and robustly modulate behavioural responses and corticostriatal activity[78,79]. Many such tasks have been explicitly designed to distinguish neural activity related to reward anticipation or decision making and activity related to rewarding outcomes (for example, see refs. 80,81). Importantly, the latter 'consummatory' phase of these tasks is often presumed to reflect activity driven by affective responses to monetary reward receipt. Contrary to this interpretation, we found little evidence to suggest that either the anticipatory or consummatory phases of the validation tasks assessed were associated with strong pleasure signals as compared with either primary sensory rewards or secondary pleasure derived from humour. Moreover, an alternative signature developed using a monetary reward task was substantially better at classifying monetary reward from loss (Extended Data Fig. 4). This suggests that monetary reward tasks may primarily capture neural encoding related to reinforcement and instrumental actions, rather than pleasurable affective states. If true, this could have important implications for neuroimaging studies of psychiatric disorders associated with apathy and anhedonia[82,83].

In sum, the current work identifies a distributed pattern of brain activity that is both sensitive and specific to pleasurable experiences. Strikingly, this signature shares many features with hedonic systems identified in non-human animals, including its anatomical distribution, sensitivity to mu-opioid receptor expression and function, and distinction from non-pleasure reward signals previously linked to dopaminergic pathways. This signature offers new insight into the distributed neural architecture underlying pleasure and can serve as the foundation for more sophisticated models and measures of hedonic function in humans.

Methods

Study selection and contrast specification

Because theories of hedonic function focus on the variety of sensory states that can produce similarly pleasurable subjective experiences, we used a mega-analytic approach to develop a generalizable signature for pleasure. This approach enables comparisons between a larger number of experimental conditions and generalization across scanners and populations. To identify a distinct pattern of fMRI activity associated with pleasure, we systematically sampled neuroimaging data from 28 independent studies that manipulated affective valence or engaged cognitive control in healthy individuals (total $n = 494$). No statistical methods were used to pre-determine the sample sizes for each study, as they come from existing datasets.

Studies involving manipulations of positive affect were chosen to include five types of reward (images of appetizing food, visual cues of monetary rewards, pleasant music, images and videos depicting pleasant social interactions, and erotic images). Two studies of each reward type were selected for training (10 studies in total, $n = 224$). Activation maps included contrasts between images of appetizing food and baseline[84,85], the average response to reward cues during a temporal discounting task and baseline[86], linear variation in reward magnitude during a mixed gambles task[87], contrasts between pleasant music and resting baseline[88,89], activation as mothers viewed videos of their infants versus baseline[90], images of social activities versus baseline following 24 h of social deprivation[91], and contrasts between erotic images and baseline[92,93].

We additionally included activation maps from an archival dataset[42] of 18 studies ($n = 270$) involving pain, cognitive control and negative affect. Activation maps from pain studies included contrasts between high (painful) and low (not painful) levels of thermal stimulation[94], high levels of painful thermal stimulation and baseline[95], rectal distension trials and baseline[96,97], and pressure applied to the thumb and baseline[98]. Activation maps from studies involving cognitive control included contrasts between blocks of an *N*-back task and a fixation baseline[99,100], trials in stop signal tasks compared with baseline[101,102], and congruent and incongruent trials from studies using the Eriksen Flanker[103] and Simon[104] tasks. Activation maps for the negative emotions domain include contrasts between negative and neutral pictures from the International Affective Picture System[105], negative pictures and baseline[106], pictures of ex-partners and pictures of close friends[107], images of others in pain and baseline[108], and listening to unpleasant affective sounds and baseline[42].

All participants in the studies described above provided informed consent in line with local ethics and institutional review boards. Supplementary Table 8 contains descriptions of the ethics approval, image acquisition and analysis, and demographics for each study. Data collection and analysis were not performed blind to the conditions of the experiments. Information about subject compensation, data acquisition and experimental paradigms are available in full detail in the corresponding references. Data were accessed and processed using software from SPM12 v7771 (https://github.com/spm/spm12), CanlabCore Tools v1 (https://github.com/canlab/CanlabCore, accessed on 8 December 2021) in addition to custom MATLAB code (see 'Code availability').

Feature selection

When selecting features for the development of PLS models, we included regions known to contain hedonic hotspots, namely NAc[45], VeP[45], insular cortex[109] and ventromedial prefrontal cortex including orbitofrontal cortex[42]. We also incorporated several other regions that have connectivity with hedonic hotspots, and several that are involved in processing variables that are often correlated with hedonic impact (for example, reward value, motivation and attention) and other effortful and/or aversive affective states. These regions include medial prefrontal cortex (subgenual cingulate, perigenual cingulate, anterior midcingulate, posterior midcingulate, ventromedial prefrontal cortex and dorsomedial prefrontal cortex)[42], amygdala (central amygdala, basolateral amygdala, superficial amygdala and amygdalostriatal area)[110], basal forebrain structures (extended amygdala and mammillary nucleus)[45], striatum (caudate, putamen and globus pallidus)[45], hypothalamus[45], habenula[45] and multiple midbrain nuclei (red nucleus, ventral tegmental area and substantia nigra)[45].

To assess the average response in each domain, activation maps showing the average effect of each domain controlling for study were estimated in separate mass-univariate analyses. These regression models included an intercept and separate terms modelling the effect of each study that were centred and scaled to unit variance. Consistent with the objective of feature selection (selecting studies and regions consistently activated by rewards), these regression models revealed that studies manipulating positive affect generally yielded increases in signal across all regions. All domains exhibited increased activity in the striatum, insula, and midcingulate cortex (Supplementary Fig. 1).

PLS specification and estimation

To identify a single pattern of brain activity associated with diverse instances of pleasure that does not respond during other manipulations of affect, we specified a PLS regression model to predict different levels of a functional hierarchy. The hierarchy included four levels: subject, study, subdomain and domain. The input data matrix consisted of contrasts from all 494 subjects in the development sample and the output matrix consisted of 46 dummy coded variables (28 studies, 14 subdomains and 4 domains, with values of +1/−1 based on inclusion/exclusion for each term, Supplementary Fig. 2). PLS models were fit using SIMPLS[111] as implemented in MATLAB. This produced a matrix of PLS regression coefficients of size 35,292 (the number of voxels selected for training plus an intercept term) by 46 outcome variables. The pleasure signature comprised the pattern of coefficients corresponding to the dummy coded variable for the positive affect domain. A block

bootstrap procedure was used for inference on PLS regression coefficients (3,000 samples). In this procedure, observations from individual studies were resampled with replacement to account for dependencies within studies. Normal approximations were made on the basis of the mean and standard deviation bootstrap distributions for each voxel, producing *z*-maps and associated *P* values. PLS regression coefficient maps were thresholded using false discovery rate correction on *P* values (two-sided) from the bootstrap procedure ($q < 0.05$).

Estimation of spatial gradients

We used linear regressions to examine whether coefficients exhibited gradients similar to those identified subcortical structures in non-human animals. In particular, the rodent NAc exhibits a rostro-caudal gradient, with rostral involvement in defensive behaviour, and caudal involvement in 'liking'. Translating these findings to humans, neuroimaging studies show that rewards activate more medial 'core-like' portions of the NAc, whereas aversive stimuli activate more lateral 'shell-like' areas[48,49]. If the pleasure signature we developed is consistent with this gradient, then a linear regression of signature coefficients onto their distance from midline (mm in MNI space) should produce a negative beta estimate. We tested this hypothesis by taking the full bootstrap distribution used for voxel-wise inference and running a spatial regression on each bootstrap sample, producing a bootstrap distribution of beta estimates that were used to compute *z*-scores and corresponding *P* values using normal approximation. The same procedure was used to identify anterior-to-posterior gradients in the VeP, which would produce a positive beta estimate if consistent with rodent topography, and in insular cortex, which would produce a negative beta estimate, although this is contentious as the structure, connectivity and cytoarchitecture of the insula differ between species[9].

Generalization tests

To evaluate the generalizability of the pleasure signature, we examined its response in four independent datasets (Studies 1–4 in Supplementary Table 9), making classifications based on cosine similarity between PLS regression coefficients estimated during training and test data. The first two of these studies were selected to test model specificity, as they included contrasts that primarily differed in terms of reward value rather than hedonic impact. They included gain versus loss trials in a standard reinforcement learning task[57] and high versus low reward trials in a task requiring participants to make decisions about expending effort for rewards of varying magnitude[112]. The second pair of studies was chosen to evaluate the sensitivity of the model, as they included contrasts between decision-making about humourous versus non-humourous content[55] and between pleasant versus neutral tastes[53]. Cohen's *d* and area under the receiver operating characteristic curve (AUROC) were used to index discrimination within individual studies, using randomization tests with 10,000 iterations.

To evaluate the signature response across all four validation studies, a linear mixed effects model was specified with study and a condition × study interaction as fixed effects, and random intercepts for subjects nested within studies. The interaction term in this model was specified to test whether the signature response was larger between conditions for studies that included a manipulation of pleasure (pleasant versus neutral tastes and humour versus neutral captioning) and those that did not (large versus small reward cues and gain versus loss feedback). The model was fit using maximum likelihood estimation through MATLAB's fitglme function. Inference was made using a two-sided *t*-test against zero and confirmed using a parametric bootstrap (10,000 iterations).

Spatial mappings with neurotransmitter receptor gene expression

We compared the pleasure signature with gene expression maps for dopamine receptors (*DRD1*, *DRD2* and *DRD3*) and opioid receptors (*ORPM1*, *ORPK1* and *ORPD1*) from the Allen Brain Atlas[61]. We performed a multiple regression using the 6 normalized gene expression maps (35,291 voxels by 6 genes) to predict the PLS regression coefficients that define the pleasure signature (35,291 voxels), producing beta estimates that reflect the unique association between gene expression for each receptor type and the pleasure signature. To estimate the variability of betas estimates, a bootstrap procedure was performed using the same bootstrap distribution used to make inference on PLS regression coefficients. Inference was performed using a two-sided test with normal approximation following visual inspection (full distributions are shown in Fig. 3).

Signature response to naloxone challenge

To assess the effects of opioids on the pleasure signature, we evaluated the effect of naloxone on mesocorticolimbic activity and self-reported pleasure in a placebo-controlled crossover study[66]. As with the other four generalization tests, we computed the cosine similarity between the signature and maps contrasting the placebo manipulation and for erotic images and visual cues indicating they would receive money after the scanning session. Because some activation maps varied in signal quality and coverage, images that were extreme outliers based on Mahalanobis distance (three for erotic image contrasts and two for monetary reward contrasts) were excluded from this analysis. We computed differences in average pattern expression and self-reported pleasure between the naloxone and saline sessions for both types of stimuli. Bootstrap resampling with normal approximation was performed for both measures, using two-sided tests for inference.

Representational similarity analysis

We constructed model-based representational dissimilarity matrices (RDMs) reflecting the psychological domains and subdomains involved in each study (following methods developed in ref. 42, Supplementary Fig. 3). For each RDM, we calculated dissimilarity as 1-Pearson's *r* between multivoxel patterns of brain activity. First, we modelled each of the 28 studies individually to assess differences in pattern generalizability across studies. Then, we modelled the 14 subdomains (food reward, musical reward, monetary reward, social reward, sexual reward, visceral stimulation, thermal stimulation, mechanical stimulation, response conflict, response selection, working memory, visual negative emotion, social negative emotion and auditory negative emotion) to assess patterns that generalize across studies but differ across subdomains. Lastly, we modelled each of the four psychological domains (positive affect, pain, cognitive control and negative affect) independently to account for response patterns that generalize across studies and subdomains but differ across the four general domains.

We used binary vectors based on study membership to model the observed brain RDMs as a linear combination of individual studies (28 RDMs), subdomains (14 RDMs) and psychological domains (4 RDMs). We then created a set of vectors from the intersubject dissimilarities of these 46 RDMs and a constant RDM, which were used as regressors in a linear regression model. On-diagonal elements, which have zero dissimilarity, were excluded from all analyses. We used a block bootstrap procedure[42] to obtain *P* values because the general linear model assumes independent errors while dissimilarity matrices exhibit complex dependencies. Positive regression coefficients thus reflect similarities in brain responses that generalize across a psychological domain and cannot be explained by features unique to any subdomain, study or individual. Two-sided tests were performed for inference.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data used to train and validate the signature are available at https://osf.io/vs84r/. Data from the Allen Brain Atlas are available at https://neurosynth.org/genes/ and http://portal.brain-map.org/.

Code availability

Code for reproducing the findings presented in this manuscript is available at https://github.com/ecco-laboratory/PMA. SPM can be downloaded from https://www.fil.ion.ucl.ac.uk/spm/software/download/, and the CanlabCore Tools are available at https://github.com/canlab/CanlabCore.

References

1. McMahon, D. M. *Happiness: A History* (Grove Press, 2006).
2. Kringelbach, M. L. & Berridge, K. C. Towards a functional neuroanatomy of pleasure and happiness. *Trends Cogn. Sci.* **13**, 479–487 (2009).
3. Berridge, K. C. & Kringelbach, M. L. Pleasure systems in the brain. *Neuron* **86**, 646–664 (2015).
4. Schultz, W. Neuronal reward and decision signals: from theories to data. *Physiol. Rev.* **95**, 853–951 (2015).
5. de Araujo, I. E., Schatzker, M. & Small, D. M. Rethinking food reward. *Annu. Rev. Psychol.* **71**, 139–164 (2020).
6. Cabanac, M. Pleasure: the common currency. *J. Theor. Biol.* **155**, 173–200 (1992).
7. Peng, Y. et al. Sweet and bitter taste in the brain of awake behaving animals. *Nature* **527**, 512–515 (2015).
8. Dolensek, N., Gehrlach, D. A., Klein, A. S. & Gogolla, N. Facial expressions of emotion states and their neuronal correlates in mice. *Science* **368**, 89–94 (2020).
9. Castro, D. C. & Berridge, K. C. Opioid and orexin hedonic hotspots in rat orbitofrontal cortex and insula. *Proc. Natl Acad. Sci. USA* **114**, E9125–E9134 (2017).
10. Smith, K. S. & Berridge, K. C. Opioid limbic circuit for reward: interaction between hedonic hotspots of nucleus accumbens and ventral pallidum. *J. Neurosci.* **27**, 1594–1605 (2007).
11. Smith, K. S. & Berridge, K. C. The ventral pallidum and hedonic reward: neurochemical maps of sucrose "liking" and food intake. *J. Neurosci.* **25**, 8637–8649 (2005).
12. Peciña, S. & Berridge, K. C. Opioid site in nucleus accumbens shell mediates eating and hedonic 'liking' for food: map based on microinjection Fos plumes. *Brain Res.* **863**, 71–86 (2000).
13. Castro, D. C. & Berridge, K. C. Opioid hedonic hotspot in nucleus accumbens shell: mu, delta, and kappa maps for enhancement of sweetness "liking" and "wanting". *J. Neurosci.* **34**, 4239–4250 (2014).
14. Berridge, K. C. Affective valence in the brain: modules or modes? *Nat. Rev. Neurosci.* **20**, 225–234 (2019).
15. Kühn, S. & Gallinat, J. The neural correlates of subjective pleasantness. *NeuroImage* **61**, 289–294 (2012).
16. Clithero, J. A. & Rangel, A. Informatic parcellation of the network involved in the computation of subjective value. *Soc. Cogn. Affect. Neurosci.* **9**, 1289–1302 (2014).
17. Liu, X., Hairston, J., Schrier, M. & Fan, J. Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neurosci. Biobehav. Rev.* **35**, 1219–1236 (2011).
18. Nummenmaa, L. et al. μ-Opioid receptor system mediates reward processing in humans. *Nat. Commun.* **9**, 1500 (2018).
19. Carlén, M. What constitutes the prefrontal cortex? *Science* **358**, 478–482 (2017).
20. Pizzagalli, D. A. & Roberts, A. C. Prefrontal cortex and depression. *Neuropsychopharmacology* **47**, 225–246 (2022).
21. Korb, S. et al. Dopaminergic and opioidergic regulation during anticipation and consumption of social and nonsocial rewards. *eLife* **9**, e55797 (2020).
22. Richard, J. M. & Berridge, K. C. Nucleus accumbens dopamine/glutamate interaction switches modes to generate desire versus dread: D1 alone for appetitive eating but D1 and D2 together for fear. *J. Neurosci.* **31**, 12866–12879 (2011).
23. Gore, F. et al. Neural representations of unconditioned stimuli in basolateral amygdala mediate innate and learned responses. *Cell* **162**, 134–145 (2015).
24. Chen, R. et al. Decoding molecular and cellular heterogeneity of mouse nucleus accumbens. *Nat. Neurosci.* **24**, 1757–1771 (2021).
25. Kahnt, T. A decade of decoding reward-related fMRI signals and where we go from here. *NeuroImage* **180**, 324–333 (2018).
26. Chikazoe, J., Lee, D. H., Kriegeskorte, N. & Anderson, A. K. Population coding of affect across stimuli, modalities and individuals. *Nat. Neurosci.* **17**, 1114–1122 (2014).
27. Poldrack, R. A. & Farah, M. J. Progress and challenges in probing the human brain. *Nature* **526**, 371–379 (2015).
28. Haynes, J.-D. & Rees, G. Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* **7**, 523–534 (2006).
29. Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J. & Barrett, L. F. The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature. *Cereb. Cortex* **26**, 1910–1922 (2016).
30. Miskovic, V. & Anderson, A. K. Modality general and modality specific coding of hedonic valence. *Curr. Opin. Behav. Sci.* **19**, 91–97 (2018).
31. Keren, H. et al. Reward processing in depression: a conceptual and meta-analytic review across fMRI and EEG studies. *Am. J. Psychiatry* **175**, 1111–1120 (2018).
32. Shackman, A. J. & Fox, A. S. Two decades of anxiety neuroimaging research: new insights and a look to the future. *Am. J. Psychiatry* **178**, 106–109 (2021).
33. Lozano, A. M. et al. Deep brain stimulation: current challenges and future directions. *Nat. Rev. Neurol.* **15**, 148–160 (2019).
34. Fenoy, A. J., Quevedo, J. & Soares, J. C. Deep brain stimulation of the "medial forebrain bundle": a strategy to modulate the reward system and manage treatment-resistant depression. *Mol. Psychiatry* **27**, 574–592 (2022).
35. Krystal, A. D. et al. A randomized proof-of-mechanism trial applying the 'fast-fail' approach to evaluating κ-opioid antagonism as a treatment for anhedonia. *Nat. Med.* **26**, 760–768 (2020).
36. Chang, L. J., Gianaros, P. J., Manuck, S. B., Krishnan, A. & Wager, T. D. A sensitive and specific neural signature for picture-induced negative affect. *PLoS Biol.* **13**, e1002180 (2015).
37. Speer, S. P. H. et al. A multivariate brain signature for reward. *NeuroImage* **271**, 119990 (2022).
38. Reddan, M. C., Lindquist, M. A. & Wager, T. D. Effect size estimation in neuroimaging. *JAMA Psychiatry* **74**, 207–208 (2017).
39. Marek, S. et al. Reproducible brain-wide association studies require thousands of individuals. *Nature* **603**, 654–660 (2022).
40. Kragel, P. A., Han, X., Kraynak, T. E., Gianaros, P. J. & Wager, T. D. Functional MRI can be highly reliable, but it depends on what you measure: a commentary on Elliott et al. (2020). *Psychol. Sci.* **32**, 622–626 (2021).
41. Han, X. et al. Effect sizes and test–retest reliability of the fMRI-based neurologic pain signature. *NeuroImage* **247**, 118844 (2022).
42. Kragel, P. A. et al. Generalizable representations of pain, cognitive control, and negative emotion in medial frontal cortex. *Nat. Neurosci.* **21**, 283–289 (2018).
43. Dubé, L. & Le Bel, J. The content and structure of laypeople's concept of pleasure. *Cogn. Emot.* **17**, 263–295 (2003).
44. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **58**, 109–130 (2001).
45. Pauli, W. M., Nili, A. N. & Tyszka, J. M. A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. *Sci. Data* **5**, 180063 (2018).
46. Gründemann, J. et al. Amygdala ensembles encode behavioral states. *Science* **364**, eaav8736 (2019).

47. Ye, L. et al. Wiring and molecular features of prefrontal ensembles representing distinct experiences. *Cell* **165**, 1776–1788 (2016).
48. Baliki, M. N. et al. Parceling human accumbens into putative core and shell dissociates encoding of values for reward and pain. *J. Neurosci.* **33**, 16383–16393 (2013).
49. Woo, C.-W. et al. Quantifying cerebral contributions to pain beyond nociception. *Nat. Commun.* **8**, 14211 (2017).
50. Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C. & Taylor, P. A. FMRI clustering in AFNI: false-positive rates redux. *Brain Connect.* **7**, 152–171 (2017).
51. Berridge, K. C. & O'Doherty, J. P. in *Neuroeconomics* 2nd edn (eds Glimcher, P. W. & Fehr, E.) 335–351 (Academic Press, 2014).
52. Rozin, P. in *Well-being: The Foundations of Hedonic Psychology* (eds Kahneman, D. et al.) 109–133 (Russell Sage Foundation, 1999).
53. Dalenberg, J. R., Weitkamp, L., Renken, R. J., Nanetti, L. & Ter Horst, G. J. Flavor pleasantness processing in the ventral emotion network. *PLoS ONE* **12**, e0170310 (2017).
54. Mobbs, D., Greicius, M. D., Abdel-Azim, E., Menon, V. & Reiss, A. L. Humor modulates the mesolimbic reward centers. *Neuron* **40**, 1041–1048 (2003).
55. Admon, R. & Pizzagalli, D. A. Corticostriatal pathways contribute to the natural time course of positive mood. *Nat. Commun.* **6**, 10065 (2015).
56. O'Doherty, J. P. The problem with value. *Neurosci. Biobehav. Rev.* **43**, 259–268 (2014).
57. Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J. & Frith, C. D. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* **442**, 1042–1045 (2006).
58. Blain, B. & Rutledge, R. B. Momentary subjective well-being depends on learning and not reward. *eLife* **9**, e57977 (2020).
59. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis—connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* https://doi.org/10.3389/neuro.06.004.2008 (2008).
60. Skerry, A. E. & Saxe, R. A common neural code for perceived and inferred emotion. *J. Neurosci.* **34**, 15997–16008 (2014).
61. Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
62. Gregoriou, G. C., Kissiwaa, S. A., Patel, S. D. & Bagley, E. E. Dopamine and opioids inhibit synaptic outputs of the main island of the intercalated neurons of the amygdala. *Eur. J. Neurosci.* **50**, 2065–2074 (2019).
63. Ford, C. P., Mark, G. P. & Williams, J. T. Properties and opioid inhibition of mesolimbic dopamine neurons vary according to target location. *J. Neurosci.* **26**, 2788–2797 (2006).
64. Eikemo, M. et al. Sweet taste pleasantness is modulated by morphine and naltrexone. *Psychopharmacology* **233**, 3711–3723 (2016).
65. Koepp, M. J. et al. Evidence for endogenous opioid release in the amygdala during positive emotion. *NeuroImage* **44**, 252–256 (2009).
66. Buchel, C., Miedl, S. & Sprenger, C. Hedonic processing in humans is mediated by an opioidergic mechanism in a mesocorticolimbic system. *eLife* **7**, e39648 (2018).
67. Olds, J. & Milner, P. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J. Comp. Physiol. Psychol.* **47**, 419–427 (1954).
68. Heath, R. G. Pleasure and brain activity in man. *J. Nerv. Ment. Dis.* **154**, 3–18 (1972).
69. Samanez-Larkin, G. R. et al. Anticipation of monetary gain but not loss in healthy older adults. *Nat. Neurosci.* **10**, 787–791 (2007).
70. Knutson, B., Taylor, J., Kaufman, M., Peterson, R. & Glover, G. Distributed neural representation of expected value. *J. Neurosci.* **25**, 4806–4812 (2005).
71. Charpentier, C. J., De Neve, J.-E., Li, X., Roiser, J. P. & Sharot, T. Models of affective decision making: how do feelings predict choice? *Psychol. Sci.* **27**, 763–775 (2016).
72. Karjalainen, T. et al. Opioidergic regulation of emotional arousal: a combined PET–fMRI study. *Cereb. Cortex* **29**, 4006–4016 (2019).
73. Pecina, M., Love, T., Stohler, C. S., Goldman, D. & Zubieta, J.-K. Effects of the Mu opioid receptor polymorphism (OPRM1 A118G) on pain regulation, placebo effects and associated personality trait measures. *Neuropsychopharmacology* **40**, 957–965 (2015).
74. Palomero-Gallagher, N., Vogt, B. A., Schleicher, A., Mayberg, H. S. & Zilles, K. Receptor architecture of human cingulate cortex: evaluation of the four-region neurobiological model. *Hum. Brain Mapp.* **30**, 2336–2355 (2009).
75. Beckmann, M., Johansen-Berg, H. & Rushworth, M. F. Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *J. Neurosci.* **29**, 1175–1190 (2009).
76. Bijanki, K. R. et al. Cingulum stimulation enhances positive affect and anxiolysis to facilitate awake craniotomy. *J. Clin. Invest.* **129**, 1152–1166 (2019).
77. Shenhav, A., Botvinick, M. M. & Cohen, J. D. The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* **79**, 217–240 (2013).
78. Somerville, L. H. et al. The Lifespan Human Connectome Project in Development: a large-scale study of brain connectivity development in 5–21 year olds. *NeuroImage* **183**, 456–468 (2018).
79. Barch, D. M. et al. Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage* **80**, 169–189 (2013).
80. Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L. & Hommer, D. Dissociation of reward anticipation and outcome with event-related fMRI. *NeuroReport* **12**, 3683–3687 (2001).
81. Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C. & Fiez, J. A. Tracking the hemodynamic responses to reward and punishment in the striatum. *J. Neurophysiol.* **84**, 3072–3077 (2000).
82. Cooper, J. A., Arulpragasam, A. R. & Treadway, M. T. Anhedonia in depression: biological mechanisms and computational models. *Curr. Opin. Behav. Sci.* **22**, 128–135 (2018).
83. Pizzagalli, D. A. Toward a better understanding of the mechanisms and pathophysiology of anhedonia: are we ready for translation? *Am. J. Psychiatry* **179**, 458–469 (2022).
84. Smeets, P. A., Kroese, F. M., Evers, C. & de Ridder, D. T. Allured or alarmed: counteractive control responses to food temptations in the brain. *Behav. Brain Res.* **248**, 41–45 (2013).
85. Iranpour, J., Morrot, G., Claise, B., Jean, B. & Bonny, J.-M. Using high spatial resolution to improve BOLD fMRI detection at 3T. *PLoS ONE* **10**, e0141358 (2015).
86. Castrellon, J. J. et al. Mesolimbic dopamine D2 receptors and neural representations of subjective value. *Sci. Rep.* **9**, 20229 (2019).
87. Tom, S. M., Fox, C. R., Trepel, C. & Poldrack, R. A. The neural basis of loss aversion in decision-making under risk. *Science* **315**, 515–518 (2007).
88. Kragel, P. A. & LaBar, K. S. Multivariate neural biomarkers of emotional states are categorically distinct. *Soc. Cogn. Affect. Neurosci.* **10**, 1437–1448 (2015).
89. Lepping, R. J. et al. Neural processing of emotional musical and nonmusical stimuli in depression. *PLoS ONE* **11**, e0156859 (2016).
90. Laurent, H. K., Wright, D. & Finnegan, M. Mindfulness-related differences in neural response to own infant negative versus positive emotion contexts. *Dev. Cogn. Neurosci.* **30**, 70–76 (2018).
91. Tomova, L. et al. Acute social isolation evokes midbrain craving responses similar to hunger. *Nat. Neurosci.* **23**, 1597–1605 (2020).

92. Kragel, P. A., Reddan, M. C., LaBar, K. S. & Wager, T. D. Emotion schemas are embedded in the human visual system. *Sci. Adv.* **5**, eaaw4358 (2019).
93. Dalenberg, J. R., Weitkamp, L., Renken, R. J. & Ter Horst, G. J. Valence processing differs across stimulus modalities. *NeuroImage* **183**, 734–744 (2018).
94. Atlas, L. Y., Bolger, N., Lindquist, M. A. & Wager, T. D. Brain mediators of predictive cue effects on perceived pain. *J. Neurosci.* **30**, 12964–12977 (2010).
95. Wager, T. D. et al. An fMRI-based neurologic signature of physical pain. *N. Engl. J. Med.* **368**, 1388–1397 (2013).
96. Kano, M. et al. Altered brain and gut responses to corticotropin-releasing hormone (CRH) in patients with irritable bowel syndrome. *Sci. Rep.* **7**, 12425 (2017).
97. Rubio, A. et al. Uncertainty in anticipation of uncomfortable rectal distension is modulated by the autonomic nervous system—a fMRI study in healthy volunteers. *NeuroImage* **107**, 10–22 (2015).
98. Čeko, M., Kragel, P. A., Woo, C.-W., López-Solà, M. & Wager, T. D. Common and stimulus-type-specific brain representations of negative affect. *Nat. Neurosci.* **25**, 760–770 (2022).
99. DeYoung, C. G., Shamosh, N. A., Green, A. E., Braver, T. S. & Gray, J. R. Intellect as distinct from openness: differences revealed by fMRI of working memory. *J. Pers. Soc. Psychol.* **97**, 883 (2009).
100. Van Ast, V. et al. Brain mechanisms of social threat effects on working memory. *Cereb. Cortex* **26**, 544–556 (2016).
101. Xue, G., Aron, A. R. & Poldrack, R. A. Common neural substrates for inhibition of spoken and manual responses. *Cereb. Cortex* **18**, 1923–1932 (2008).
102. Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J. & Poldrack, R. A. Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *J. Neurosci.* **27**, 3743–3752 (2007).
103. Kelly, A. C., Uddin, L. Q., Biswal, B. B., Castellanos, F. X. & Milham, M. P. Competition between functional brain networks mediates behavioral variability. *NeuroImage* **39**, 527–537 (2008).
104. Mennes, M., Kelly, C., Colcombe, S., Castellanos, F. X. & Milham, M. P. The extrinsic and intrinsic functional architectures of the human brain are not equivalent. *Cereb. Cortex* **23**, 223–229 (2013).
105. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* **8**, 665–670 (2011).
106. Gianaros, P. J. et al. An inflammatory pathway links atherosclerotic cardiovascular disease risk to neural activity evoked by the cognitive regulation of emotion. *Biol. Psychiatry* **75**, 738–745 (2014).
107. Kross, E., Berman, M. G., Mischel, W., Smith, E. E. & Wager, T. D. Social rejection shares somatosensory representations with physical pain. *Proc. Natl Acad. Sci. USA* **108**, 6270–6275 (2011).
108. Krishnan, A. et al. Somatic and vicarious pain are represented by dissociable multivariate brain patterns. *eLife* **5**, e15166 (2016).
109. Glasser, M. F. et al. A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
110. Amunts, K., Mohlberg, H., Bludau, S. & Zilles, K. Julich-Brain: a 3D probabilistic atlas of the human brain's cytoarchitecture. *Science* **369**, 988–992 (2020).
111. de Jong, S. SIMPLS: an alternative approach to partial least squares regression. *Chemom. Intell. Lab. Syst.* **18**, 251–263 (1993).
112. Arulpragasam, A. R., Cooper, J. A., Nuutinen, M. R. & Treadway, M. T. Corticoinsular circuits encode subjective value expectation and violation for effortful goal-directed behavior. *Proc. Natl Acad. Sci. USA* **115**, E5233–E5242 (2018).

Acknowledgements

This project was supported by grants R01MH126083 and R00MH102355 to M.T.T. D.A.P. was partially supported by grants P50MH119467 and R37MH068376. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

Conceptualization, methodology, formal analysis, writing—original draft, writing—review and editing, validation and visualization, P.A.K.; conceptualization, review and editing, and resources, M.T.T.; writing—review and editing, and resources, R.A.; writing—review and editing, and resources, D.A.P.; conceptualization, writing—original draft, review and editing, and formal analysis, E.C.H.

Competing interests

The authors declare the following competing interests: in the past 3 years M.T.T. has served as a paid consultant to Neumora Therapeutics (formerly BlackThorn Therapeutics) and Boehringer Ingelheim. Over the past 3 years, D.A.P. has received consulting fees from Albright Stonebridge Group, Boehringer Ingelheim, Compass Pathways, Engrail Therapeutics, Neumora Therapeutics (formerly BlackThorn Therapeutics), Neurocrine Biosciences, Neuroscience Software, Otsuka, Sunovion and Takeda; he has received honoraria from the Psychonomic Society and the American Psychological Association (for editorial work) and Alkermes; he has received research funding from the Brain and Behavior Research Foundation, the Dana Foundation, Millennium Pharmaceuticals, National Institute of Mental Health (NIMH) and Wellcome Leap; he has received stock options from Compass Pathways, Engrail Therapeutics, Neumora Therapeutics and Neuroscience Software. No funding from these entities was used to support the current work, and all views expressed are solely those of the authors. The remaining authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41562-023-01639-0.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-023-01639-0.

Correspondence and requests for materials should be addressed to Philip A. Kragel.

Peer review information *Nature Human Behaviour* thanks Junichi Chikazoe, Dylan Nielson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

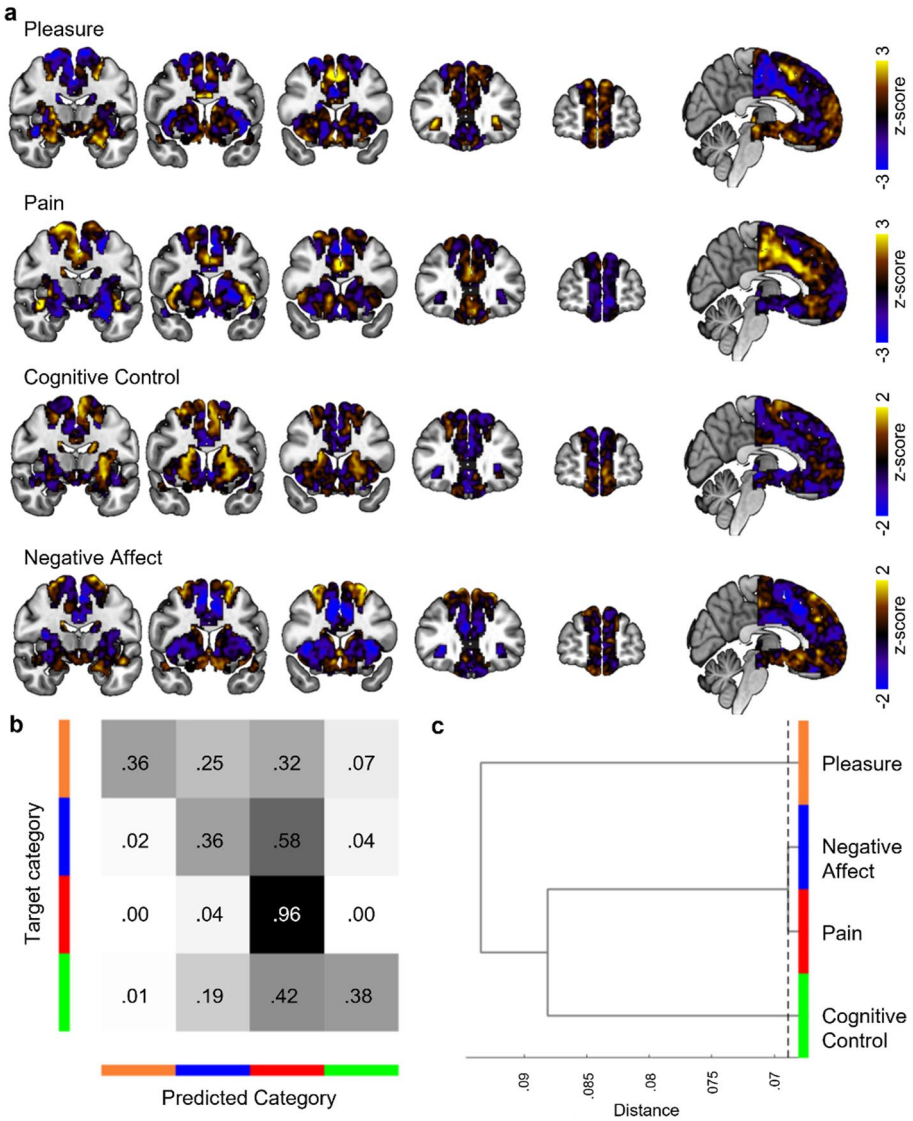
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

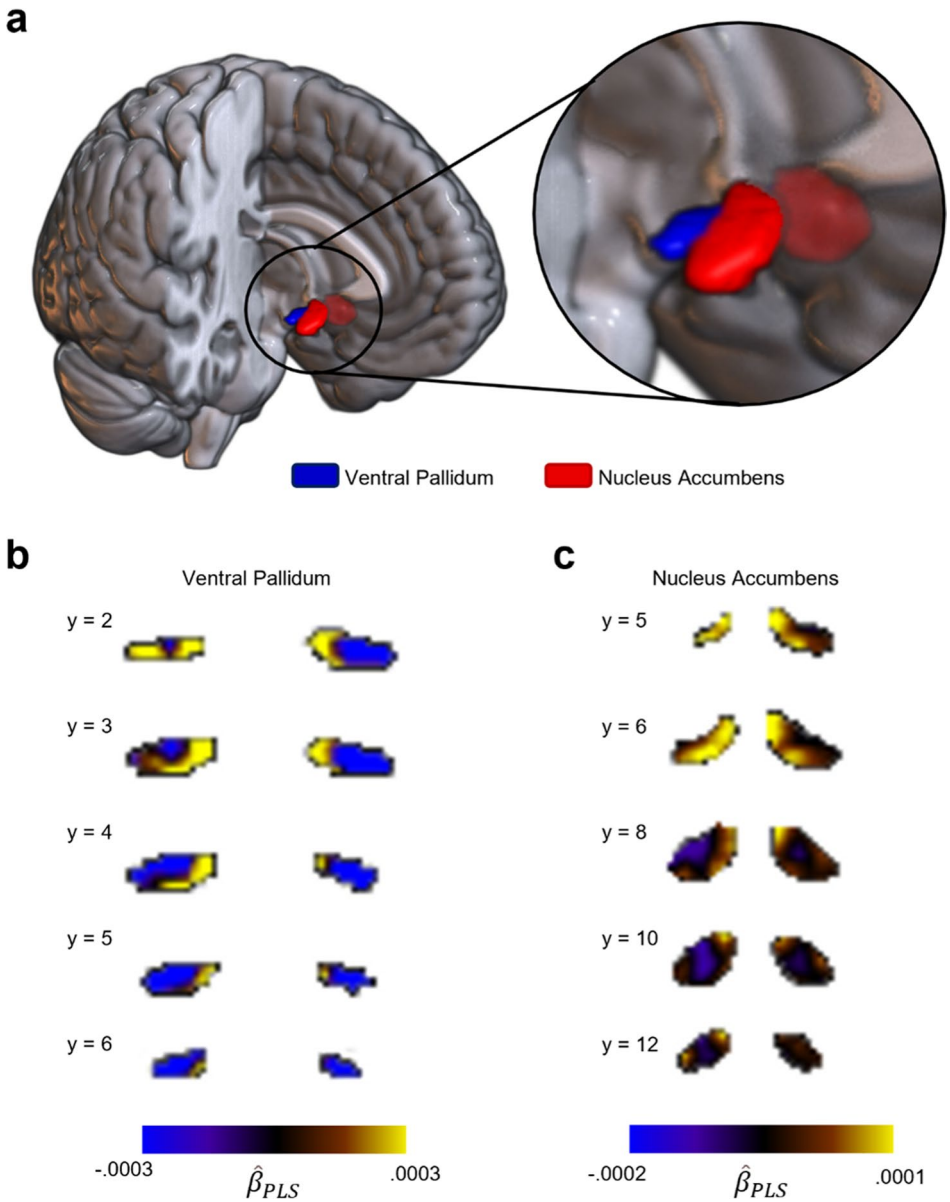
© The Author(s), under exclusive licence to Springer Nature Limited 2023

https://doi.org/10.1038/s41562-023-01639-0



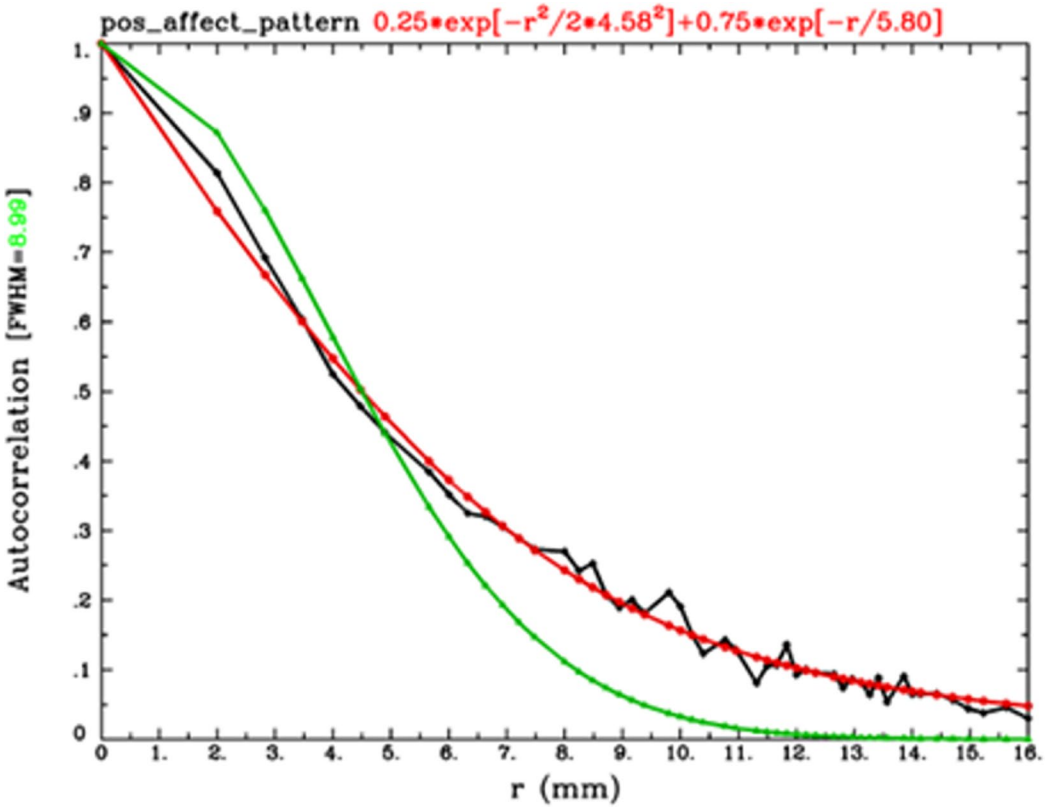
Extended Data Fig. 1 | Multivariate models discriminate brain states during manipulations of pleasure, pain, cognitive control, and negative affect. (**a**) Rendering of z-scores for beta estimates from Partial Least Squares regression fit on training data ($n = 499$) overlaid on the ICBM152 template. Warm colors are positively associated with predictions of each domain, whereas cool colors are negatively associated with each domain. (**b**) Confusion matrix estimated using stratified 5-fold cross-validation in the training dataset (4-way accuracy = 47.17%, all four classes are statistically distinguishable at $p < .05$). Rows have been normalized to sum to 1. (**c**) Clustering of domains based on classification errors. Dendrogram shows clustering of errors using Ward's linkage. Dashed vertical line depicts the optimal cut point, in which all four domains are assigned to separate clusters.

https://doi.org/10.1038/s41562-023-01639-0



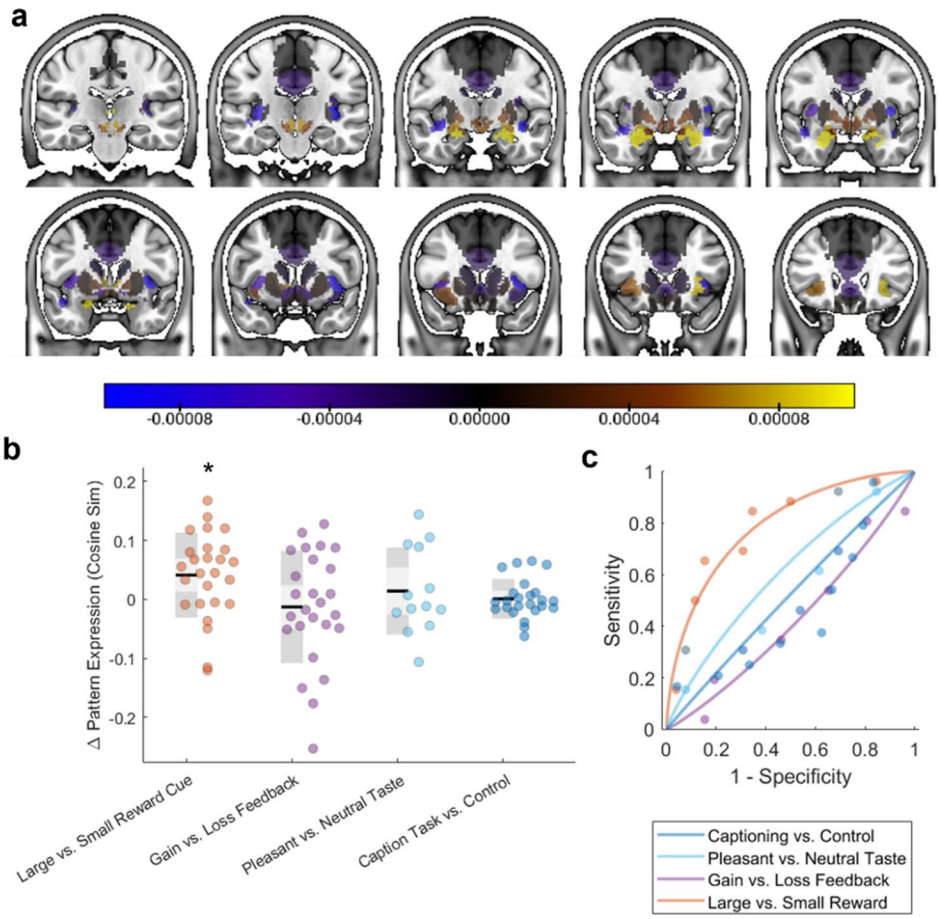
Extended Data Fig. 2 | Signature coefficients within ventral pallidum and nucleus accumbens. (**a**) Volumetric rendering of anatomically defined regions of interest overlaid on the ICBM152 template. (**b**) Signature coefficients (beta estimates from Partial Least Squares regression) within the ventral pallidum that predict states of pleasure. Warm colors are positively associated with predictions of pleasure, whereas cool colors are negatively associated with pleasure. MNI coordinates (mm in the y dimension) are shown next to each section. (**c**) Signature coefficients in the nucleus accumbens.

https://doi.org/10.1038/s41562-023-01639-0



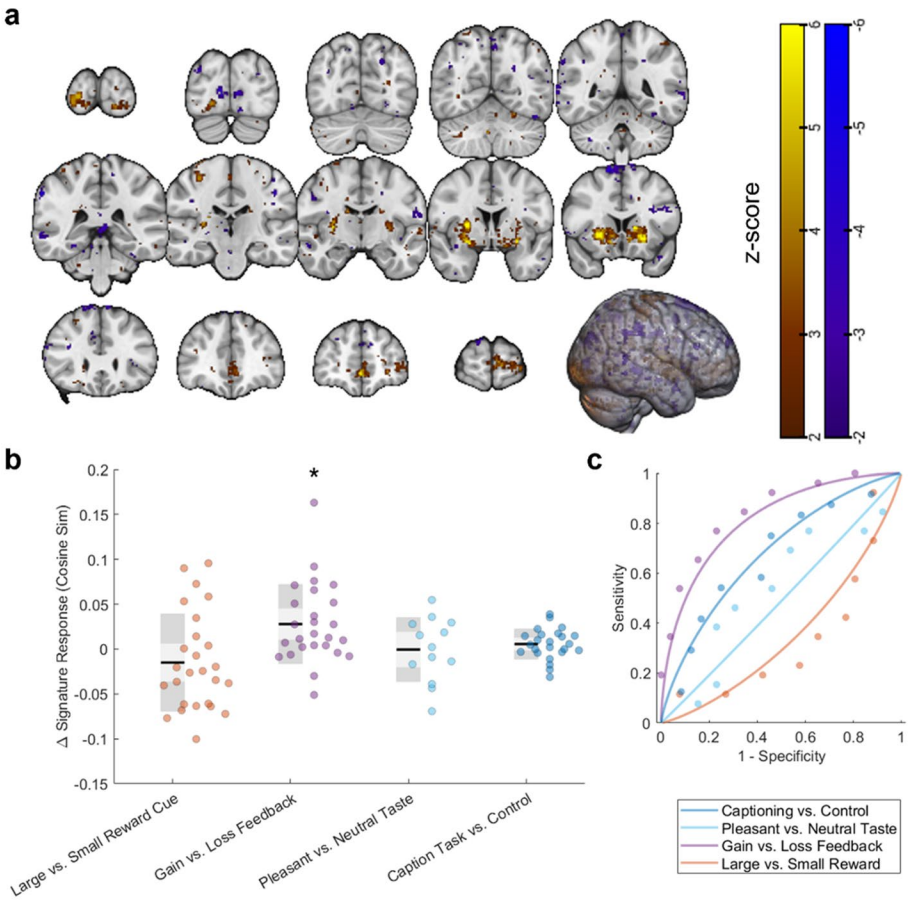
Extended Data Fig. 3 | Estimated spatial smoothness of the pleasure signature. The empirical spatial autocorrelation of the pleasure signature (black) and estimates using both Gaussian (green) and mono-exponential fit (red) are shown. Figure generated from the AFNI program 3dFWHMx.

https://doi.org/10.1038/s41562-023-01639-0



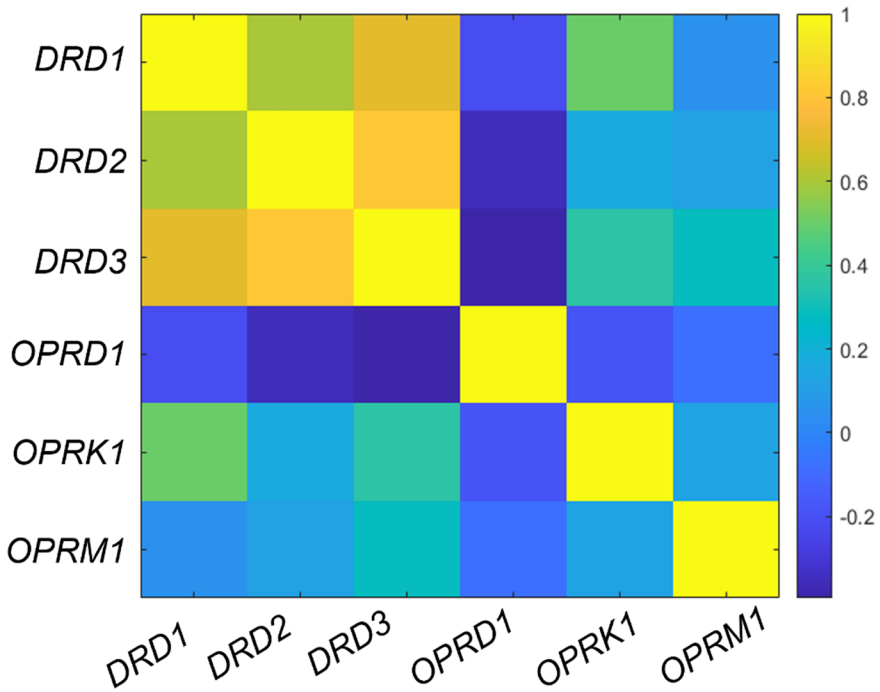
Extended Data Fig. 4 | A simplified, region-average model of the signature response is neither sensitive nor specific to pleasure. (**a**) The simplified model defined by the average of coefficients from the optimized signature. Warm colors indicate regions in which increases in brain activity contribute to predictions of pleasure, whereas cool colors indicate regions in which increased brain activity leads to fewer classifications of pleasure. (**b**) Box and whisker plot shows differences in the region-average signature response for each study (*ns* = 26, 26, 13, 24 independent participants; *mean = .0412, t_{25} = 2.948, p = .00685, *d* = .578, 95% Confidence Interval = [.0138 .0686], uncorrected two-sided paired t-test). Black lines depict the mean response, light-shaded regions depict one standard deviation, and darker-shaded regions two standard errors. Each point corresponds to the response of a single subject. (**c**) Receiver operating characteristic curves for each of the four studies.

https://doi.org/10.1038/s41562-023-01639-0



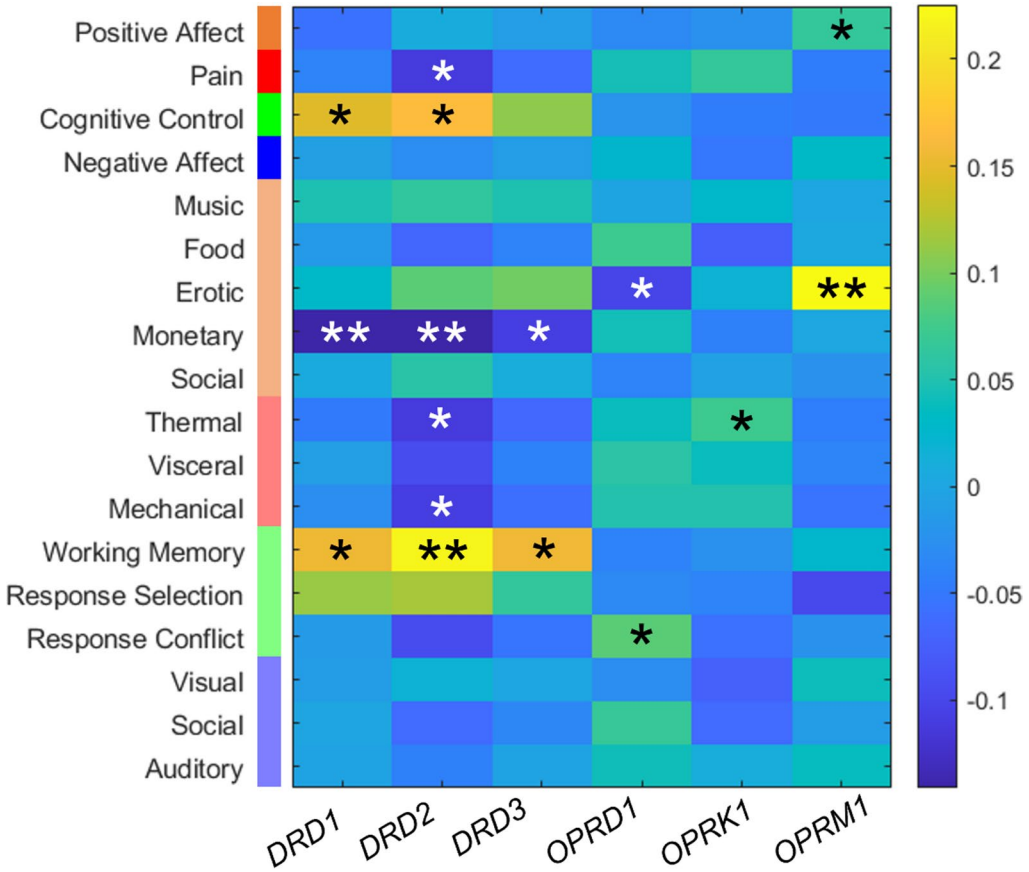
Extended Data Fig. 5 | Brain reward signature is sensitive to reward feedback, but not pleasure. (**a**) Reward signature trained to discriminate gains and losses in a monetary incentive delay task. Warm colors indicate regions in which increases in brain activity contribute to predictions of greater reward, whereas cool colors indicate regions in which increases in brain activity lead to lower levels of reward. (**b**) Box and whisker plot shows differences in the region-average signature response for each study (*n*s = 26, 26, 13, 24 independent participants; *mean = .0279, t_{25} = 3.221, p = .00353, d = .632, 95% Confidence Interval = [.0109 .0448], uncorrected two-sided paired t-test). Black lines depict the mean response, light-shaded regions depict one standard deviation, and darker-shaded regions two standard errors. Each point corresponds to the response of a single subject. (**c**) Receiver operating characteristic curves for each of the four studies.

https://doi.org/10.1038/s41562-023-01639-0



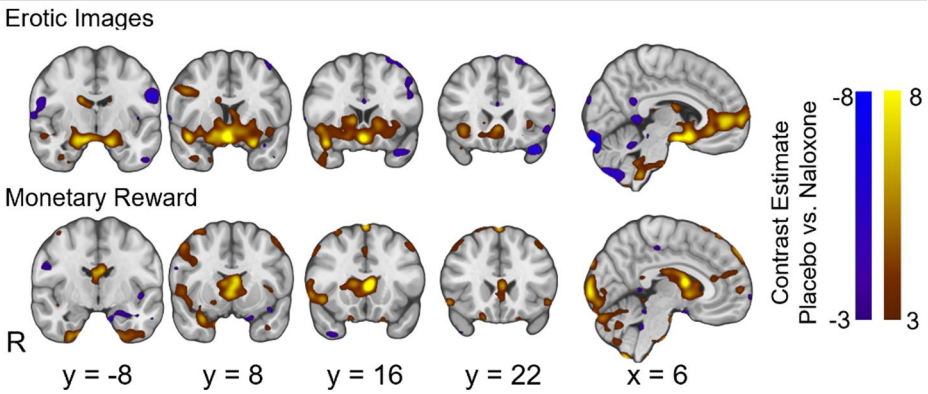
Extended Data Fig. 6 | Spatial correlation of neurotransmitter gene expression maps from the Allen Brain Atlas. Heatmap depicts the Pearson correlation coefficient between pairs of gene expression maps, with cool colors indicating negative correlations and warm colors positive correlations.

https://doi.org/10.1038/s41562-023-01639-0



Extended Data Fig. 7 | Spatial correlation between Partial Least Squares regression coefficients and neurotransmitter gene expression maps from the Allen Brain Atlas. *$p < .05$; **$q_{FDR} < .05$, two-sided bootstrap test. Full statistics are reported in Supplementary Table 7.

https://doi.org/10.1038/s41562-023-01639-0



Extended Data Fig. 8 | Group-average contrast maps showing the effect of naloxone on brain activity during the presentation of erotic images and feedback about monetary rewards. Warm colors indicate greater activity during saline placebo administration compared to naloxone, whereas cool colors indicate a greater response during naloxone administration compared to placebo.

nature portfolio

Corresponding author(s): Philip Kragel

Last updated by author(s): 4/22/2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Archival data were accessed from Neurovault.org and Openneuro.org using the Amazon Web Services command line interface, and through CanlabCore Tools v1 command line interface (https://github.com/canlab/CanlabCore).
Data analysis	SPM12 v7771 (https://github.com/spm/spm12), CanlabCore Tools v1 (https://github.com/canlab/CanlabCore, accessed on Dec 8, 2021) in addition to custom MATLAB code were used to analyze data (available at https://github.com/ecco-laboratory/PMA). The SIMPLS algorithm for PLS regression implemented in the Statistics and Machine Learning Toolbox as introduced in MATLAB R2008a.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data used to train and validate the signature are available at https://osf.io/vs84r/. Data from the Allen Brain Atlas are available at https://neurosynth.org/genes/ and http://portal.brain-map.org/.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Both the development sample and validation samples included data from males and females (self-reported sex). These studies were not designed to test sex as a biological variable. Because the proportion of females in each study range from 0 to 100% (Supplemental Tables 7 and 8), we did not include sex-based analysis as we would be underpowered for these post-hoc tests. The findings of this study generally apply to both sexes, with the exception of validation studies 1 and 5, which only involved male participants.
Population characteristics	We sampled data from 34 studies of healthy adults (28 studies for model development and 6 for validation). The mean sample size was 30.8 participants (sd = 31.8, range = 10-183). The mean proportion of females was .533 (sd = .221, range = 0-1). The mean age was 25.9 years (sd = 4.21, range = 20.2-42.7).
Recruitment	N/A
Ethics oversight	This work involved secondary analysis of anonymized data. Study protocols for data collection were approved by ethics boards at the University Hospital of Clermont-Ferrand, University Medical Center Utrecht, Vanderbilt University, the University of California, Los Angeles, Duke University, University of Kansas Medical Center, University of Oregon, Massachusetts Institute of Technology, University Medical Center Groningen, University of Colorado Boulder, Columbia University, Tohoku University, Comité de Protection des Personnes Sud Est V, France, Washington University Medical Center, New York University, University of Pittsburgh, and Stanford University. Informed consent was obtained by participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No formal statistical procedure was used to determine sample size for individual studies, because this work analyzed existing datasets in a mega-analysis.
Data exclusions	Five activation maps from validation study 5 were excluded as outliers based on Mahalanobis distance (3 for erotic image contrasts and 2 for monetary reward contrasts). This was due to differences in signal quality and coverage. No other data were excluded.
Replication	Model performance was verified in four confirmatory validation tests that supported replication. The reliability of the model was assessed in longitudinal resting state data.
Randomization	This work is a mega-analysis of data pooled from existing studies. Randomization of subjects to different groups is not applicable.
Blinding	This work is a mega-analysis of data pooled from existing studies. Blinding the assignment of participants to different experimental groups is not applicable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Magnetic resonance imaging

Experimental design

Design type	Event-related task fMRI, resting-state fMRI
Design specifications	Varies by study. See Supplemental Table 4.
Behavioral performance measures	Behavioral measures of interest depended on the study. Readers are referred to the original publications for more details.

Acquisition

Imaging type(s)	BOLD fMRI
Field strength	1.5 and 3 Tesla
Sequence & imaging parameters	EPI (standard and multiband) and spiral in-out sequences were used for data acquisition. Readers are referred to the original publications for more details.
Area of acquisition	Whole brain scans were used.

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	SPM12 v7771 was used to preprocess fMRI data acquired from OpenNeuro. Preprocessing involved spatial realignment, coregistration of functional and structural images, unified segmentation and normalization of T1 weighted images to MNI152 space. Data obtained through Neurovault was preprocessed using SPM and custom code, with different versions used depending on the study. Please refer to the relevant publication for more details.
Normalization	Non-linear normalization to MNI space was performed using T1 weighted images for datasets obtained through OpenNeuro.
Normalization template	All data were aligned to MNI152 space. The templates used for normalization depended on the study (variants of the ICBM152 template or SPM's default tissue probability map).
Noise and artifact removal	Regression of motion parameters was performed in all studies. Nuisance covariates included either 6 parameters based on translation and rotation or 24 parameters including translation and rotation, and their derivatives, successive differences, and squared successive differences.
Volume censoring	Volume censoring was not performed.

Statistical modeling & inference

Model type and settings	We developed a predictive model using partial least squares (PLS) regression. The model was specified to predict different levels of a functional hierarchy including four levels: subject, study, subdomain, and domain. A block bootstrap procedure was used for inference on PLS regression coefficients, in this procedure, observations from individual studies were resampled with replacement to account for dependencies within studies, treating subject as a random effect. We performed representational similarity analysis (RSA) to test whether BOLD signals in regions of interest generalized across studies. We used binary vectors based on study membership to model the observed brain RDMs. We created a set of vectors from the intersubject dissimilarities of these RDMs and a constant RDM, which were used as regressors in a linear regression model. Inference was made using a block bootstrap procedure was performed to account for dependencies between elements of RDMs while treating subject as a random effect.
Effect(s) tested	The predictive model was developed to classify measures of brain activity as states of "positive affect" (studies involved food,

Effect(s) tested

monetary, music, social, and sexual reward; see Supplemental Table 4). We validated the model by comparing responses of the model to pleasant vs. neutral beverages, humor vs. neutral captions, visual cues of high vs. low The predictive model was developed to classify measures of brain activity as states of "positive affect" (studies involved food, monetary, music, social, and sexual reward; see Supplemental Table 4) versus other brain states. We validated the model by comparing responses of the model to pleasant vs. neutral beverages, humorous vs. neutral captions, visual cues of high vs. low monetary reward, visual indicators of monetary gains vs. losses. The effect of naloxone (vs. placebo) on model predictions was evaluated separately for monetary and primary rewards.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☒ Both

Anatomical location(s): Existing anatomical atlases and parcellations were used to define ROIs and for feature selection.

Statistic type for inference (See Eklund et al. 2016)

Voxel-wise thresholding.

Correction

FDR q < .05

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☐	☒ Multivariate modeling or predictive analysis

Multivariate modeling and predictive analysis

We developed a predictive model (PLS regression) to classify whether an individual is in a positive affective state using patterns of fMRI BOLD activity as input. This model included multiple dependent variables to enable predictions at multiple levels of a functional hierarchy. The hierarchy included four levels: subject, study, subdomain, and domain. The input data matrix consisted of contrasts from all 494 subjects in the development sample and the output matrix consisted of 46 dummy coded variables (28 studies, 14 subdomains, and 4 domains, with values of +1/–1 based on inclusion/exclusion for each term).

When selecting features to use for model development, we included regions known to contain hedonic hotspots, namely nucleus accumbens, ventral pallidum, insular cortex, and ventromedial prefrontal cortex including orbitofrontal cortex. We also incorporated several other regions that have connectivity with hedonic hotspots, and several that are involved in processing variables that are often correlated with hedonic impact (e.g., reward value, motivation, attention) and other effortful and/or aversive affective states. These regions include medial prefrontal cortex (subgenual cingulate, perigenual cingulate, anterior midcingulate, posterior midcingulate, ventromedial prefrontal cortex, and dorsomedial prefrontal cortex), amygdala (central amygdala, basolateral amygdala, superficial amygdala, and amygdalostriatal area), basal forebrain structures (extendeded amygdala, mammillary nucleus), striatum (caudate, putamen, globus pallidus), hypothalamus, habenula, and multiple midbrain nuclei (red nucleus, ventral tegmental area, and substantia nigra).

The PLS regression model was regularized by retaining 16 latent dimensions to minimize the potential for overfitting while affording the possibility to capture study- subdomain- and domain-specific effects related to positive affect. PLS models were fit using SIMPLS, and performance was assessed using two-alternative forced choice tests on cosine similarity. Scale free measures of effect size (Cohen's d and AUC) are reported in the main text.