A neuromarker for drug and food craving distinguishes drug users from non-users

Leonie Koban[∞1]
Tor D. Wager[*∞2]
Hedy Kober[*∞3]

[1] Paris Brain Institute (ICM), Inserm, CNRS, Sorbonne Université, Paris, France
[2] Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH
[3] Department of Psychiatry and Psychology, Yale University, New Haven, CA

Running Head: A BRAIN MARKER FOR DRUG AND FOOD CRAVING

* These authors contributed equally
∞Please address correspondence to:

Leonie Koban, PhD
Paris Brain Institute (ICM), Inserm U 1127, CNRS UMR 7225
Sorbonne University
Leonie.koban@icm-institute.org

OR

Tor D. Wager, PhD
Department of Psychological and Brain Sciences
Dartmouth College
Tor.D.Wager@dartmouth.edu

OR

Hedy Kober, PhD
Department of Psychiatry
Yale University
hedy.kober@yale.edu

Abstract

Craving is a core feature of substance use disorders. It is a strong predictor of substance use and relapse, and linked to overeating, gambling, and other maladaptive behaviors. Craving is measured via self-report, which is limited by introspective access and sociocultural contexts. Neurobiological markers of craving are both needed and lacking, and it remains unclear whether craving for drugs and food involve similar mechanisms. Across three fMRI studies (N=99), we identified a cross-validated neuromarker that predicts self-reported intensity of cue-induced drug and food craving (p<0.0002). This pattern, the Neurobiological Craving Signature (NCS), includes ventromedial prefrontal and cingulate cortices, ventral striatum, temporal/parietal association areas, mediodorsal thalamus, and cerebellum. NCS responses to drug versus food cues discriminate drug users versus non-users with 82% accuracy. The NCS is also modulated by a self-regulation strategy. Transfer between separate neuromarkers for drug and food craving suggests shared neurobiological mechanisms. Future studies can assess the discriminant and convergent validity of the NCS, and test whether it responds to clinical interventions and predicts long-term clinical outcomes.

Introduction

Craving—a strong desire to use drugs or to eat—has long been considered a core factor driving overeating and substance use[1], thereby contributing to the top three preventable causes of disease and death[2]. In 2013, it was added as a diagnostic criterion for substance use disorders (SUDs) in the DSM-5[3], underscoring its clinical significance. Importantly, cue-induced craving – which arises in response to drug- or food-related stimuli – is known to prospectively predict eating unhealthy foods (i.e., ultra-processed foods high in sugar and saturated fat), weight gain, drug use, and relapse (for meta-analyses, see[4–6]). Because it is a common predictor across multiple conditions (including SUDS, obesity, eating disorders) and maladaptive behaviors, it may constitute a transdiagnostic risk factor.

While self-reported craving has been useful clinically as well as experimentally, there is growing recognition of the need to understand its biological basis. Human neuroimaging studies have identified circuits related to substance use risk, incidence, and sequelae[7,8]. Some circuits, including ventromedial prefrontal cortex (vmPFC), ventral striatal/nucleus accumbens (VS/NAc), and insula, have been identified across SUDs, outcomes, and risky behaviors, and have been identified as key regions underlying addiction in animal models[9–13]. Alongside homeostatic circuits in hypothalamus and brainstem, these regions have been identified in studies of food valuation and dietary decision-making[14,15], and appear to be functionally related to weight gain and obesity[16,17]. These circuits can be targeted by neurostimulation, pharmacological, psychological, and behavioral interventions[18], providing new avenues for therapeutic intervention.

Nevertheless, although great strides have been made in our understanding of substance misuse, overeating, and related phenomena, our understanding of the neural basis of craving is still incomplete, and neural targets for monitoring craving and SUDs and for examining the efficacy of interventions are lacking. While the neuroimaging literature on craving is growing, craving cannot be directly measured in nonhumans[19]. In addition, understanding that any specific brain region is involved in craving or other outcomes does not imply that we can *decode* craving from the brain, or that we have a sufficiently precise measurement model to allow for monitoring of individual people or testing whether treatments engage their intended craving-related neural targets[20]. It is increasingly apparent that many mental states and related outcomes have a highly distributed brain basis, including emotion[21,22], pain[23], perception[24], object recognition[25],

memory retrieval[26], sustained attention[27], semantics[28], and autonomic responses[29]. Accordingly, measures that integrate across brain systems can provide sensitive, specific, and generalizable characterizations of the neurophysiological underpinnings of behavior[30]. They can also predict health-related outcomes with larger effect sizes than measures based on single regions in many cases[31].

Such predictive models—also called 'neuromarkers' or 'signatures'—have multiple potential uses[32–34]. They can predict risk for future disorders, identify subtypes (or biotypes) that predict who will respond to a treatment, and perhaps most importantly serve as mechanistic targets for interventions. They can also outperform subjective measures in predicting human choices[35], and can be linked with systems and cellular neuroscience to develop new biological treatments in 'reverse translation' approaches[36]. Accordingly, there is increasing recognition of the need to develop biomarkers based on human systems that can be compared with animal models[33,34,37]. However, such an approach has rarely been applied in addiction[38], and has not yet been applied to craving.

Here, we take a first step towards a neuromarker that predicts the intensity of drug and food craving in clinical and matched control samples. We integrated data from 5 different cohorts in 3 fMRI studies across different types of drug users (cigarettes, alcohol, cocaine), and non-users (total of 469 contrast images from N=99 participants). Across studies, participants were presented with visual cues of drugs and highly palatable food items. We then used machine-learning to identify a distributed functional brain activity pattern that predicted the intensity of craving.

We term the resulting pattern the *Neurobiological Craving Signature* (NCS), and we hope that this name would reduce ambiguity and provide a reference point for the pattern's future reuse and testing in new samples. Analyses related to the NCS allow us to address scientific questions related to the organization of craving-related brain systems across drugs and food (or other rewarding stimuli), and their susceptibility to cognitive, pharmacological, and other interventions. Further, recent perspectives have proposed a common neurophysiology for SUDs and obesity, and of drug and food craving more specifically[39–41], but this view has been challenged[42]. The NCS allows us to test whether craving for several types of drugs, including stimulants (nicotine, cocaine) and sedatives (alcohol), and for highly palatable foods are based on different or shared neurophysiological patterns. We further assess whether the brain systems involved in cue-induced craving are affected by cognitive regulation strategies, highlighting the malleability

of craving-related brain patterns to interventions and thus opening avenues for developing further interventions and improving existing ones.

Results

Data overview

A total of 469 contrast images from 99 participants and 5 independent cohorts were used for training and testing the pattern to predict drug and food craving (two drug using cohorts, two of their matched controls, and another sample of drug users with no matched controls). All participants viewed images of drugs and food under two instructions conditions: a craving instruction and an instruction to use a cognitive strategy to reduce craving (see Methods). Contrast images were computed for the onset of the visual drug and food cues (see **Figure 1a**), separately for each level of craving (1-5 Likert scale) for every participant (see Suppl. Figure 1) and were rescaled by the image-wise L2-norm to remove any differences in scale between participants and scanners.

fMRI results

Description of the Neurophysiological Craving Signature (NCS). Parallel to previous studies on fMRI-based prediction of pain and emotion[21,23], LASSO-PCR and study-stratified 10-fold cross-validation was used to predict the level of craving based on fMRI contrast images. The advantage of this approach is that it does not require a similar level of craving across food and drugs (or across participants and studies), because it predicts continuous, dimensional craving intensity ratings. Variance in self-reported craving, both within and between participants, is beneficial for the LASSO-PCR algorithm. Model training identifies a pattern of weights across voxels such that the weighted average activity is optimized to predict craving in a training sample of participants, and its predictive accuracy is validated in independent participants. The NCS is a model that consists of the weights (plus an overall intercept), which can be applied to any brain image to obtain a weighted average over brain voxels, yielding a single score per test image. If weights in a brain area are positive, more activity indicates higher predicted craving. If they are negative, more activity indicates lower predicted craving. **Figure 2** presents a thresholded display of the resulting weight map based on bootstrapping. While the unthresholded map (see Suppl. Figure 2) is used for prediction, the thresholded map illustrates the brain areas that most robustly contribute positive or negative weights to the predictive pattern. Areas with positive weights included ventromedial prefrontal cortex, dorsal anterior cingulate

cortex, subgenual cingulate/ventral striatum, retrosplenial cortex, parietal and temporal areas, cerebellum, and amygdala. Negative weights were found in visual areas, lateral prefrontal, parietal and somatomotor areas, among others (see **Table 1** for a list of FDR-corrected coordinates). Of note, many areas, including somatomotor cortex, parietal, temporal, and bilateral insula included clusters of both positive and negative weights.

Predictive performance of the NCS. The trained pattern resulted in a cross-validated prediction-outcome correlation of $r = 0.53$ (S.D. +-0.46) within-person and $r = 0.34$ across all data points, with a mean absolute error of 1.30 points on the 1-5 Likert scale. A multi-level general linear model (GLM) confirmed a strong relationship between out-of-sample predicted and actual level of craving with a large effect size ($\beta = 0.38$, STE = 0.04, $t(98) = 9.21$, $p < 0.0001$, Cohen's $d = 0.93$, see **Figure 3**). The strength of the predictive performance varied across datasets, but was significant in all 5 cohorts, with effect sizes (Cohen's *d*) ranging from 0.55-1.48 (see **Table 2**). Statistically controlling for white matter and ventricle signal did not alter these results (i.e., the relationship between craving ratings and NCS remained significant, $\beta = 0.35$, STE = 0.05, $t(97) = 6.81$, $p < 0.0001$, while the relationship of WM and ventricle signals with craving was not, $\beta = 2.53$, STE = 1.77, $t(97) = 1.43$, $p = 0.16$). Adding scanner site, group (drug users versus controls), sex, age, education, BMI, and average head motion as 2nd-level covariates also did not alter the relationship between craving and NCS ($\beta = 0.38$, STE = 0.04, $t(79) = 9.17$, $p < 0.0001$). Though the NCS predicts craving in both users and non-users, users versus non-users had significantly higher overall NCS responses ($\beta = 0.38$, STE = 0.18, $t(79) = 2.13$, $p = 0.036$) and smaller effect of craving ratings on out-of-sample NCS-predicted responses (i.e., a smaller slope, $\beta = -0.14$, STE = 0.05, $t(79) = -2.97$, $p = 0.004$), potentially because users may report ratings less reliably or have less neurotypical brains. None of the other covariates significantly affected the NCS or interacted with ratings to affect NCS responses. NCS responses did not significantly change over time during the experiment (see Suppl. Figure 3).

Classification of high versus low craving. We next assessed the accuracy of the NCS to differentiate between high versus low levels of craving. Forced-choice binary classification of highest versus lowest levels of craving per participant was achieved with a cross-validated accuracy of 81% +-4.0%STE, binomial test $p < 0.0001$

(sensitivity = 81%, specificity = 81%, area under the curve [AUC] = 0.91, see **Figure 3**). Even across subjects (single-interval classification), this pattern separated brain responses to the highest versus lowest individual levels of craving with 72% cross-validated accuracy (+-3.4%STE, binomial test $p < 0.0001$, sensitivity = 64%, specificity = 80%, AUC = 0.76). While this level of predictive accuracy does not provide perfect separation of high versus low craving, it is remarkable, since all stimuli were drugs or highly palatable food items, thus differences in classification performance were not driven by external stimulus characteristics but by the personal history and internal motivational states of the participants.

Our studies did not include in-scanner ratings other than craving ratings. We therefore assessed whether the NCS does indeed predict something specific to craving that is not predicted by other brain signatures, which are trained to predict other types of affect ratings. For this purpose, we applied five recently developed brain signatures[43]—trained to predict four different types of negative affect (mechanical pain, thermal pain, aversive sounds, and unpleasant pictures) and domain-general negative affect—to the data from Studies 1-3 and tested whether these other brain signatures would predict high versus low craving with comparable accuracy as the NCS. The results of this control analysis confirmed that other signatures trained to predict affective ratings did not significantly predict high versus low cravings but were at chance level (46%-52% accuracy, see Suppl. Figure 4).

Differentiating drug users from non-users. We next tested whether individual craving-pattern responses to drug and food cues could be used to predict whether a participant was a drug user or non-user (**see Figure 4a** for group averages, **Figure 4b** for individual effects, and **Figure 4c** for ROC plots). While pattern expression in brain responses to food cues did not significantly differentiate drug users from non-users (60% accuracy +-4.9% STE, $P = 1.00$, AUC = 0.40), NCS pattern responses to drug cues significantly classified drug users from non-users, with 75% accuracy (+-4.4% STE, $P = 0.002311$, sensitivity = 86%, specificity = 57%, AUC = 0.76). When testing the pattern response to the drug>food contrast, the response in the NCS separated drug users from non-users with 82% accuracy (+-3.9% STE, $P < 0.001$, sensitivity = 97%, specificity = 60%, AUC = 0.87, see **Figure 4c**).

Given slight but significant differences in years of education between users and non-users, we used a general linear model to control for years of education, other basic

demographic variables (age and biological sex) in predicting individual differences in NCS response. This showed that drug users had stronger NCS responses to drug cues than non-users ($t(94) = 4.22$, $p < 0.001$, 96%-CI: [0.57, 1.55], Cohen's $d = 0.87$) and stronger NCS responses to drug>food cues ($t(94) = 7.04$, $p < 0.001$, 96%-CI: [0.90, 1.60], Cohen's $d = 1.45$), while education, age, and sex were not associated with NCS responses to drugs or drug>food cues (all p's > 0.20).

In addition, we tested whether the classification of drug users versus non-users could be driven by any single study (or user group) alone or whether they are significant in each study independently. We performed the classification analysis separately on Studies 1 and 3 (note that Study 2 did not include non-users). The results showed that NCS responses to drug cues and drug>food cues (but not food cues) significantly separated users from non-users in both Study 1 and Study 3, separately (see Suppl. Figure 5 for ROC plots and full results). Average craving ratings and NCS responses for each study and cue type are also shown in Suppl. Figure 6.

Drug and food craving are predicted by shared brain patterns. An important debate concerns the question whether drug and food craving are based on similar brain processes[40,42]. If drug and food craving are driven by shared brain processes, then drug craving should be predictable based on a pattern that is trained to predict food craving, and food craving should be predictable based on a pattern that is trained to predict drug craving – at least in drug users. Conversely, if drug and food craving are based on dissociable brain processes, then better predictive accuracy would be gained by training drug- and food-specific (compared to craving-general) brain patterns.

We therefore repeated the procedures described above and tested whether training on drug and food data separately would improve prediction of craving, and whether food craving could be predicted based on a pattern trained on drug data only, and vice versa (**Figure 5**). Food craving was predicted similarly well by the overall pattern (76% out-of-sample accuracy +-4.3% STE, $P < 0.001$, AUC = 0.82) as by a craving pattern trained on food cues only (79% +-4.1% STE, $P < 0.001$, AUC = 0.88). Food craving was also significantly predicted by a pattern trained on drug cues only, but with somewhat lower accuracy across both drug-using and non-using participants (65% +-4.8% STE, $P = 0.005$, AUC = 0.68). For the prediction of drug craving, the results indicated no substantial improvements for training only on modality-specific (drug) cue trials (69% +-4.9% STE, $P < 0.001$, AUC = 0.75) compared to all cues (70% +-4.9% STE,

$P < 0.001$, AUC = 0.78). Drug craving was also significantly predicted by a pattern that was trained only on food trials (66% +-5.1% STE, $P = 0.004$, AUC = 0.74). Thus, we did not find evidence for a double dissociation between drug and food craving, but rather significant cross-prediction of drug and food craving. Most importantly, the NCS performed as well as the two cue-specific patterns. Together, this supports the hypothesis of shared representations between drug and food craving – and across drug types.

Modulation by cognitive regulation strategies. Finally, we used a general linear model to assess how craving ratings and responses of the NCS were modulated by the cognitive regulation of craving task that was employed in all five studies. Across all participants, craving ***ratings*** were influenced by cue type (drug vs. food, $F_{(1,388)} = 95.5$, $p < 0.001$, 95% CI: [-0.44, -0.29]) and regulation instruction (NOW vs LATER, $F_{(1, 388)} = 97.6$, $p < 0.001$, 95% CI: [0.32, 0.49]). Drug users reported greater overall craving ($F_{(1, 388)} = 74.2$, $p < 0.001$, 95% CI: [0.36, 0.58]) and this group effect interacted with both regulation instruction ($F_{(1, 388)} = 4.51$, $p = 0.034$, 95% CI: [0.01, 0.17]) and cue type ($F_{(1, 388)} = 191.5$, $p < 0.001$, 95% CI: [0.44, 0.59]), such that drug users craved drugs ($t_{(97)} = 14.5$, $p < 0.001$, 95% CI: [1.69, 2.23], Cohen's d = 2.94) but not food ($t_{(97)} = -0.67$, $p = 0.50$, 95% CI: [-0.34, 0.17], Cohen's d = 0.14) more than non-users, and that they showed slightly higher regulation effects than non-users. Importantly, these effects were qualified further by a significant three-way interaction between group, cue type, and regulation condition ($F_{(1,388)} = 21.7$, $p < 0.001$, 95% CI: [0.05, 0.12], see **Figure 3c**). While the regulation effect was significant for both drug and food cues in both users and non-users, the difference between NOW and LATER condition was significantly smaller in the drug condition compared to the food condition in non-users ($t_{(37)} = -4.22$, $p < 0.001$, 95% CI: [-0.77, -0.27], Cohen's d = 0.67), who reported low craving for drugs overall. Consistently, drug users had a somewhat larger regulation effect (difference between NOW and LATER condition) for drug compared to food cues ($t_{(58)} = 2.14$, $p = 0.037$, 95% CI: [0.01, 0.35], Cohen's d = 0.28).

Similar to craving ratings, ***responses of the NCS*** were influenced by cue type (drug vs. food, $F_{(1, 388)} = 70.4$, $p < 0.001$, 95%-CI: [-0.51, -0.83]) and by regulation instruction (NOW vs LATER, $F_{(1, 388)} = 35.5$, $p < 0.001$, 95%-CI: [0.28, 0.55]), suggesting that cognitive regulation strategies modify NCS responses. Drug users versus non-users had marginally greater NCS responses overall ($F_{(1, 388)} = 3.0$, $p = 0.085$, 95%-CI: [-0.05, 0.75]). Drug users' versus non-users' signature response differed with respect to cue type

($F_{(1, 388)}$ = 57.5, p < 0.001, 95%-CI: [0.90, 1.53]), such that drug users had higher NCS responses to drug cues than non-users ($t_{(97)}$ = 4.39, p < 0.001, 95%-CI: [0.56, 1.49], Cohen's d = 0.90), whereas NCS responses to food cues did not significantly differ ($t_{(97)}$ = -1.13). Further, drug users' versus non-users' signature response differed with respect to regulation condition ($F_{(1, 388)}$ = 9.15, p = 0.003, 95%-CI: [0.15, 0.69]), such that users had larger NCS responses than non-users in the NOW condition ($t_{(97)}$ = 2.53, p = 0.013, 95%-CI: [0.12, 1.00], Cohen's d = 0.52), but not in the LATER condition ($t_{(97)}$ = 0.71), see **Figure 3c**), which was likely driven by more room to down-regulate craving in users compared to non-users. The three-way-interaction between group, regulation and cue type was not significant for NCS responses ($F_{(1,388)}$ = 0.0).

Affective stimulus characteristics. We next explored how self-reported craving and the NCS were related to intrinsic craving-related image features of the different food and drug cues. For this purpose, we employed a deep-learning neural network that has been previously trained to detect 20 different affective states, including craving, in visual images ("Emonet"[44]). This allowed us to test whether single-trial craving ratings and NCS responses were associated with the Emonet visual 'craving' output unit on a stimulus-by-stimulus basis. Emonet 'craving' output is a probability score indicating the predicted probability that humans will label an image as 'craving' related and reflects a high-level abstraction of visual input. A multilevel GLM confirmed that both stimulus-to-stimulus craving ratings (β = 0.04, STE = 0.00, t(95) = 10.5, p < 0.001) and NCS responses (β = 0.02, STE = 0.00, t(95) = 6.80, p < 0.001) were strongly and positively associated with the automatic Emonet 'craving' scores for the stimuli (see Supplementary Figure 7 for additional results). Importantly, the association between craving ratings and NCS remained highly significant when controlling for 'craving' stimulus features (β = 0.18, STE = 0.02, t(95) = 8.60, p < 0.001), ruling out stimulus features as the main or only source of NCS variability. Instead, the NCS significantly mediated the effects of Emonet's 'craving' output on self-reported craving (p = 0.011, see **Figure 6**).

Discussion

Craving contributes to multiple behaviors that are detrimental to physical and mental health in the long term, including smoking, alcohol drinking, overeating, and gambling[4,5], and is arguably one of the most central processes in SUDs[6]. Like other key transdiagnostic processes – and human behavior more broadly – craving results from

brain function. However, it is typically assessed using subjective measures that require introspection and are sensitive to context[6]; thus, there is a strong need for biomarkers, and particularly neuromarkers based on brain function[37,38,45–47]. Such biomarkers can identify mechanistic targets that can aid in monitoring disease progression (monitoring biomarkers according to the FDA), identifying individuals at risk for SUDs and future weight gain (prognostic biomarkers), predicting treatment response (predictive biomarkers), and serving as targets for neuromodulatory and behavioral interventions[34].

Here we used machine learning to identify a distributed brain pattern – which we term the NCS as a reference for future use – that tracks the degree of craving when applied to new individuals, across different diagnostic groups, scanners, and scanning parameters. Importantly, this pattern separated drug users from non-users based on brain responses to drug cues, but not food cues. Thus, it is an important step towards a diagnostic neuromarker of substance use. Further, given the role of self-reported craving in predicting outcomes[4,5], this brain-based pattern may function as both a diagnostic and predictive biomarker with potential utility in predicting clinically-relevant individual differences and future outcomes. Future studies could build on these findings to test whether the NCS responds to therapeutic interventions that reduce craving and/or drug use and whether it has predictive value for long-term clinical outcomes such as drug relapse or weight gain. Further, we found that the NCS is sensitive to cognitive regulation strategies, indicating that it may be psychologically modifiable. This is important because psychological and behavioral interventions can be effective for SUDs, but their mechanisms are poorly understood. Furthermore, current interventions are associated with high rates of relapse and could be improved[48]. Future models could also be developed based on other data types (e.g., resting-state fMRI, imaging in animals) or their combination[38,46].

Our results also offer new insight into a long-standing debate concerning the question whether craving of drugs of abuse and food share common underlying brain processes, especially in motivation-related circuits[40]. We show that craving of various types of drugs and food can be predicted by largely shared whole-brain activity patterns. Indeed, the results demonstrate that craving-related responses to cues for legal and illegal drugs and for highly palatable food items are surprisingly similar and not dissociable at the fMRI pattern level in both drug users and non-drug using adults. This is noteworthy, especially as most of the non-users in the present studies were not obese or “food addicted,” but rather healthy controls. Importantly, this overlap is consistent with models

suggesting that drug craving depends on systems evolved for seeking highly palatable food and other primary rewards[40]. Future research could test whether the NCS also responds to less palatable or healthy food items, and to other types of primary and secondary rewards.

Some areas in the NCS, including the vmPFC and ventral striatum/accumbens, have been broadly implicated in reward and valuation[49,50], and have long been associated with craving and substance use across species. Several prior studies and meta-analyses[51,41,39] have demonstrated a central role of vmPFC, VS, amygdala, insula, and posterior cingulate cortex in drug and food cue reactivity and craving (although findings across meta-analyses are inconsistent). The vmPFC has been targeted in rTMS studies to successfully reduce drug craving[52]. The positive peaks of the NCS in this area could thus serve as a more precise target for neurostimulation. Future studies can test whether successful neurostimulation of vmPFC also reduces NCS expression and alters connectivity of the vmPFC with other NCS core areas such as the ventral striatum.

The insula is connected to many regions of the NCS and has been previously associated with craving[53]. Lesions in various insular locations have been shown to reduce the urge to use drugs and facilitate smoking cessation[54], which could reflect the role of the insula in the interoceptive component of drug craving[55,56]. The NCS has positive weights (at uncorrected thresholds) in the mid and posterior insula, in line with these previous reports. However, the anterior insula also displayed negative weights in the NCS (at uncorrected thresholds), revealing a potentially more complex role of different insula subregions in craving. Further, the insula might be more prominent to bodily cues of withdrawal, craving, and negative affect[53], as well as for nutrient-related reward signals[57], whereas areas such as amygdala or vmPFC (which are more prominent in the NCS) are related to craving evoked by external cues[53] such as those employed in the present datasets.

The NCS's weights were largely negative in lateral prefrontal cortex, lateral parietal areas, somatosensory cortex, and precuneus, indicating that activity in these areas is associated with reduced craving. Lateral prefrontal cortex particularly is known to be involved in cognitive control and emotion regulation[58], including the cognitive regulation of craving (e.g., as shown previously in the same data sets[59,60]) and by others[61–64]. This area is also involved in the regulation of dietary decision-making, such as when focusing more on health aspects and long-term consequences of foods[65]. The negative weights of the NCS in these areas are thus consistent with these previous findings and recent simulation

studies[66] that suggest a causal role for lateral PFC in the regulation of drug and food craving.

Finally, predictive NCS features were also found in occipital and parietal brain areas associated with visual processing and attention allocation. Our control analyses demonstrated that those effects may not be due to differences in low-level visual stimulus features. The application of a deep neural network[44] showed that both behavioral craving ratings and NCS responses were partially driven by complex, craving-related stimulus features, as captured by Emonet's 'Craving' output. However, the NCS was associated with craving ratings above and beyond elementary and craving-related image features, and partially mediated their effects on ratings, ruling out that this association was driven purely or primarily by low-level or complex image features or content. We also note that NCS weights in visual and attentional areas may reflect the effects of recurrent connections and top-down (content- and meaning-related) effects on visual processing.

In sum, the NCS further extends prior work in several ways: First, it includes strong positive and negative weights in brain areas not previously associated with craving, such as the cerebellum, lateral temporal, and parietal areas. These areas are connected to regions more traditionally associated with craving and might constitute new targets for investigation and intervention. Second, the NCS is a precise and replicable pattern, including relative activity levels across voxels within key regions and relative activity across networks. Thus, it constitutes a reproducible brain model[30,67] of craving that can be empirically quantified and validated in any new brain imaging study or dataset.

The present findings have some limitations that could be addressed in future studies. The included studies used a limited set of highly appetitive cues. Future studies could use a larger range of stimuli, including less palatable (and healthier) food items or non-craving related (neutral) cues. Greater variation in craving ratings should – in principle – lead to increased discrimination accuracy between low and high craving. We also note that hunger ratings were available in Study 2, and did not correlate with NCS responses. Nevertheless, future work is needed to characterize how hunger or food deprivation modulates NCS responses to food (and other) cues, or how NCS responses might differ in overweight or obese participants. Future studies could also test other modalities of drug and food cues (such as cigarette smoke or food smells, videos). The present study used craving ratings as the predicted outcome and did not have a non-craving control condition in the same group of participants. While our supplemental analyses show that the NCS is distinct from other signatures that predict other types of affect ratings, the discriminant

validity of the NCS should be further evaluated in future studies. Another important future direction will be to validate whether the NCS predicts other correlates of craving such as psychophysiological responses to drug cues, event-related potentials[68], and other types of behavioral measures[69]. In addition, fMRI has an inherently limited spatial resolution that cannot pinpoint the cellular or microstructural processes associated with craving, or different types of craving. However, craving cannot be directly assessed in animals, and this work fills a crucial gap across species and brain systems, which is important for translating neuroscientific findings for human clinical use. It is also important for future translational applications of MRI-based neuromarkers, which will inevitably use different scanners, hardware, and processes that evolve over time—thus requiring a focus on large-scale patterns that are generalizable across studies, scanners, groups, different preprocessing protocols, and other factors.

In both Western and Eastern philosophy, craving has been considered a source of suffering and unhappiness. While craving is an important feature of SUDs, eating disorders, and other psychiatric conditions, it is also a general aspect of human experience. Identifying the neurobiological basis of this important driver of human behavior is thus an important step in mapping brain circuits to basic affective and mental processes. In this paper we introduce the *NCS* — the first fMRI-based neuromarker of drug and food craving, which classifies drug users from non-users based on responses to drug, but not food cues. As such, it offers a promising target for future research and clinical interventions.

Acknowledgments

This work was funded by grants from NIDA (R01 DA043690 to HK) and a Campus France Marie-Sklodowska-Curie co-fund fellowship (PRESTIGE-2018-2-0023 to LK). Matlab code for analyses is available at: https://github.com/canlab. Collection of the datasets included herein was funded by R01DA022541 (PI: Kevin Ochsner), P50AA012870 (PI: John Krystal), P20DA027844 (PI: Marc Potenza). We thank everyone who supported the collection of these data. We also thank Nilofar Vafaie for help with data management and transfer.

Author Contributions Statement

HK designed the experiments and collected the data. LK analyzed the data, created the figures, and wrote the first draft of the manuscript. HK and TW conceived the project,

obtained funding, and supervised the project. All authors contributed to the writing and editing of the paper.

Competing Interest Statement

None of the authors have competing financial or non-financial interests.

Tables

Table 1. Clusters of FDR-corrected bootstrapped weights of the NCS.

Region name	Vol (mm^3)	X	Y	Z	max(Z)	Atlas region (see note)	Large-scale network / structure (see note)
Positive weights							
Postcentral gyrus / somatosensory cortex	405	-42	-27	60	4.7279	3b_L	SomatomotorA
Inferior temporal gyrus	189	57	-48	-15	4.6653	TE1p_R	Fronto_ParietalB
Cerebellum	297	-48	-63	-39	4.659	Cblm_CrusI_L	Cerebellum
Subcallosal gyrus / ventral striatum	135	12	6	-21	4.5203	pOFC_R	Limbic
Superior frontal gyrus / dlPFC	270	-24	33	54	4.4567	8BL_L	Default_ModeB
Rostral gyrus / vmPFC	162	-6	51	3	4.3707	a24_L	Default_ModeA
Retrosplenial cortex	27	-3	-54	15	4.2439	v23ab_L	Default_ModeA
Inferior parietal lobule	108	30	-63	45	4.235	IP1_R	Dorsal_AttentionA
Supramarginal gyrus	54	57	-33	45	4.2191	PF_R	Ventral_AttentionA
Supraparietal lobule	27	-12	-60	60	4.0332	7Am_L	Dorsal_AttentionB
Thalamus	108	0	-9	6	4.0137	Thal_MD	Diencephalon
Angular gyrus	27	57	-36	21	3.9646	PSL_R	Temporal_Parietal
Middle frontal gyrus	27	-42	39	18	3.8987	46_L	Ventral_AttentionB
Lateral occipital	27	51	-72	-18	3.8826	PH_R	Dorsal_AttentionA
Negative weights							
Postcentral gyrus / somatosensory cortex	2025	-51	-21	51	-7.5193	1_L	SomatomotorA
Middle temporal gyrus	243	54	-63	6	-4.8288	TPOJ2_R	Dorsal_AttentionA
Superior occipital gyrus	108	18	-84	45	-4.5147	V6A_R	Visual_Peripheral
Angular gyrus	162	42	-60	33	-4.5105	Pgi_R	Default_ModeC
Visual cortex	81	39	-72	-18	-4.4118	PIT_R	Visual_Central
Superior temporal gyrus	27	54	-3	0	-4.187	PBelt_R	SomatomotorB
Precuneus	27	9	-63	33	-4.052	POS2_R	Fronto_ParietalC
Superior parietal lobule	27	-12	-57	75	-3.9052	7AL_L	Dorsal_AttentionB
Visual cortex	27	0	-84	6	-3.8653	V1_R	Visual_Peripheral
Angular gyrus	27	-51	-60	54	-3.8516	PFm_L	Fronto_ParietalB

Note. Significant positive and negative weights contributing to the NCS (FDR corrected $q < 0.05$ across the whole brain gray-matter mask). Cortical atlas regions are labeled based on a combination of parcellations available on Github (see Methods for details): https://github.com/canlab/Neuroimaging_Pattern_Masks/tree/master/Atlases_and_parcellations/2018_Wager_combined_atlas. This repository includes multiple atlases and other meta-analytic and multivariate maps. Tools for manipulating and analyzing this and other atlases are in the CANlab Core Tools repository: https://github.com/canlab/CanlabCore

Table 2. Predictive performance of the NCS and effect sizes for each study sample.

Dataset	Cues	Sample	N	Prediction-outcome (glmfit_multilevel)	Effect size Cohen's *d*
Study 1a	Cigarette and food cues	Cigarette smokers	21	β=0.22, STE=0.09, t(20)=2.45, p = 0.024	d = 0.55
Study 1b	Cigarette and food cues	Non-smokers	22	β=0.46, STE=0.07, t(21)=6.77, p < 0.001	d = 1.48
Study 2	Alcoholic drinks and food cues	Alcohol users	17	β=0.32, STE=0.11, t(16)=2.76, p = 0.014	d = 0.69
Study 3a	Cocaine and food cues	Cocaine users	21	β=0.36, STE=0.09 , t(20)=3.87, p = 0.001	d = 0.87
Study 3b	Cocaine and food cues	Non-users	18	β=0.58, STE=0.09, t(17)=6.11, p < 0.001	d = 1.48
All Studies			99	β=0.38, STE=0.04, t(98)=9.21, p < 0.001	d = 0.93

Figure Legends

Figure 1. Study design and analytic approach. (a) In the Regulation of Craving (ROC) task, participants were presented with a series of photographs depicting either drugs (cigarettes, alcoholic drinks, or cocaine) or highly palatable food items. Before presentation of the cues, participants were instructed (2s written cue) to either consider the immediate consequences of consumption of the items ('NOW' condition), or their negative (typically long-term) consequences ('LATER' condition). At the end of each trial, participants rated their craving ('How much do you want this?'), using a 1-5 Likert scale. **(b)** The present study employed the pooled data from 3 previous studies (5 groups of participants). Study 1 tested the ROC task (displaying cigarette and food cues) in 21 heavy smokers (Study 1a) and 22 non-smokers (Study 1b; see details in methods). Study 2 tested the ROC task (displaying alcohol and food cues) in participants fulfilling diagnostic criteria of alcohol use disorder (N=17, see details in methods). Study 3 tested the ROC task (displaying cocaine and food cues) in 21 individuals with cocaine use disorder (Study 3a) and 18 matched non-users (Study 3b; see details in methods). **(c)** For each participant from all five studies, we computed brain activation images (β-estimates) for each level of craving (1-5). These images were then used in a LASSO-PCR (least absolute shrinkage and selection operator – principal component regression) machine-learning algorithm to predict level of craving (1-5) based on brain activity. Cross-validation (ten folds stratified for studies and participants population) allowed assessment of (1) predictive accuracy of the pattern for craving; (2) whether it was differentially activated for drug versus food cues; (3) whether it was differentially activated for the two regulation conditions (NOW vs. LATER); and (4) whether the pattern can differentiate drug users from non-users. **(d)** Permutation test results. Null distributions are plotted in blue bars, observed accuracy measures as red lines. For all measures, none of the permutation samples performed as well as the observed results (all p's < 0.0002). MSE = mean squared error; RMSE = root mean squared error; MAE = mean absolute error.

Figure 2. Thresholded display of the NCS. Note that unthresholded patterns are used for prediction; this thresholded pattern is shown for illustration at $p < 0.005$ uncorrected. (a) Medial, lateral, and insula displays of the most consistent pattern weights. (b) Pop-out rectangles show the multivariate pattern for selected clusters of interest. Warm (yellow-

red) color indicates positive weights, cold (cyan-purple) color indicates negative weights in predicting drug and food craving. P-values are based on bootstrapping and indicate the areas that contribute most consistently with positive or negative weights. See **Table 1** for a list of FDR corrected weights. The NCS weight map and code to apply it to new data are available for download at https://github.com/canlab/Neuroimaging_Pattern_Masks/tree/master/Multivariate_signature_patterns/2022_Koban_NCS_Craving.

Figure 3. Predictive performance of the NCS. (a) Receiver-operating characteristic (ROC) plot for the prediction of highest versus lowest levels of craving (forced-choice discrimination, N=99). **(b)** Individual datapoints and slopes for the relationship between craving levels and NCS for all five datasets (significant positive association in each study see **Table 2**). **(c)** Average levels of craving ratings (on the x-axis) and NCS responses (on the y-axis) for each of the four experimental conditions (drug versus food cues, NOW versus LATER instruction) and within each dataset. Gray lines show individual slopes across the four conditions. Dots indicate individual data points for each condition and participant. Horizontal and vertical error bars indicate standard errors of the mean (SEM) for ratings and NCS pattern expression, respectively (Study 1a: n=21, Study 1b: n=22, Study 2: n=17, Study 3a: n=21, Study 3b: n=18).

Figure 4. Classification of drug users versus non-users based on NCS responses to drugs and food. (a) Out-of-sample responses of the NCS to drug and food cues, in drug users (N=59) and non-users (N=40). Data are presented as mean values +/- SEM, dots show individual data points. (**b)** Differences in NCS responses to drug minus food cues, in drug users versus non-users. **(c)** Response-operating characteristic (ROC) plots for NCS-based prediction of drug use. Whereas NCS responses to food cues did not significantly separate users from non-users (in blue), NCS responses to drug cues were able to significantly separate users from non-users (in red). Differential NCS responses to drug-food cues (in purple) had the highest classification accuracy (82%, see text for details).

Figure 5. Cross-prediction of a) drug and b) food craving based on drug- and food-based brain patterns. Compared to the NCS (trained across all conditions, red ROC plots), training on drug or food cues separately (gray ROC plots) does not improve

accuracy, suggesting shared predictive patterns for cue-induced drug and food craving. Numbers indicate prediction accuracy for each brain classifier (all were significant, as indicated by stars).

Figure 6. The NCS partially mediates the effects of intrinsic visual craving features on craving ratings. (a) A deep convolutional neural network **(**Emonet[44], on the left) was used to extract visual craving-related features from the drug and food cue stimuli. We were most interested in the Emonet 'craving' output, shaded in orange. **(b)** A mediation analysis confirmed that (1) higher Emonet 'craving' output was associated with higher NCS responses (path *a*, $\beta = 0.01$, STE = 0.00, $p < 0.001$); (2) NCS responses significantly predicted craving ratings when controlling for visual features (path *b*, $\beta = 0.11$, STE = 0.01, $p < 0.001$); and (3) the NCS partially mediated the effect of visual craving features on ratings (path *ab*, $\beta = 0.0009$, STE = 0.0003, $p = 0.011$). Note that the direct path remained significant as well (path *c'*, $\beta = 0.03$, STE=0.00, $p < 0.001$).

References

1. O'Brien, C. P., Childress, A. R., Ehrman, R. & Robbins, S. J. Conditioning factors in drug abuse: can they explain compulsion? *J. Psychopharmacol. Oxf. Engl.* **12**, 15–22 (1998).
2. Janssen, F., Trias-Llimós, S. & Kunst, A. E. The combined impact of smoking, obesity and alcohol on life-expectancy trends in Europe. *Int. J. Epidemiol.* **50**, 931–941 (2021).
3. Diagnostic and Statistical Manual of Mental Disorders. *DSM Library* https://dsm.psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596.
4. Vafaie, N. & Kober, H. Association of Drug Cues and Craving With Drug Use and Relapse: A Systematic Review and Meta-analysis. *JAMA Psychiatry* (2022) doi:10.1001/jamapsychiatry.2022.1240.
5. Boswell, R. G. & Kober, H. Food cue reactivity and craving predict eating and weight gain: a meta-analytic review. *Obes Rev* **17**, 159–177 (2016).
6. Sayette, M. A. The Role of Craving in Substance Use Disorders: Theoretical and Methodological Issues. *Annu Rev Clin Psychol* **12**, 407–433 (2016).
7. Jasinska, A. J., Stein, E. A., Kaiser, J., Naumer, M. J. & Yalachkov, Y. Factors modulating neural reactivity to drug cues in addiction: a survey of human neuroimaging studies. *Neurosci Biobehav Rev* **38**, 1–16 (2014).
8. Heilig, M. *et al.* Addiction as a brain disease revised: why it still matters, and the need for consilience. *Neuropsychopharmacology* 1–9 (2021) doi:10.1038/s41386-020-00950-y.
9. Berridge, K. C. & Robinson, T. E. Liking, wanting, and the incentive-sensitization theory of addiction. *Am Psychol* **71**, 670–679 (2016).
10. Bedi, G. *et al.* Incubation of cue-induced cigarette craving during abstinence in human smokers. *Biol Psychiatry* **69**, 708–711 (2011).
11. Schoenbaum, G. & Shaham, Y. The role of orbitofrontal cortex in drug addiction: a review of preclinical studies. *Biol Psychiatry* **63**, 256–262 (2008).
12. Everitt, B. J., Dickinson, A. & Robbins, T. W. The neuropsychological basis of addictive behaviour. *Brain Res Brain Res Rev* **36**, 129–138 (2001).
13. Lüscher, C., Robbins, T. W. & Everitt, B. J. The transition to compulsion in addiction. *Nat Rev Neurosci* **21**, 247–263 (2020).

14. Farr, O. M., Li, C.-S. R. & Mantzoros, C. S. Central nervous system regulation of eating: Insights from human brain imaging. *Metabolism* **65**, 699–713 (2016).
15. Plassmann, H., Schelski, D. S., Simon, M.-C. & Koban, L. How we decide what to eat: Toward an interdisciplinary model of gut-brain interactions. *Wiley Interdiscip Rev Cogn Sci* **13**, e1562 (2022).
16. Schmidt, L. *et al.* Resting-state connectivity within the brain's reward system predicts weight loss and correlates with leptin. *Brain Commun.* **3**, fcab005 (2021).
17. Lawrence, N. S., Hinton, E. C., Parkinson, J. A. & Lawrence, A. D. Nucleus accumbens response to food cues predicts subsequent snack consumption in women and increased body mass index in those with reduced self-control. *Neuroimage* **63**, 415–422 (2012).
18. Ekhtiari, H. *et al.* Transcranial electrical and magnetic stimulation (tES and TMS) for addiction medicine: A consensus paper on the present state of the science and the road ahead. *Neurosci Biobehav Rev* **104**, 118–140 (2019).
19. Venniro, M., Banks, M. L., Heilig, M., Epstein, D. H. & Shaham, Y. Improving translation of animal models of addiction and relapse by reverse translation. *Nat Rev Neurosci* (2020) doi:10.1038/s41583-020-0378-z.
20. Whelan, R. & Garavan, H. When optimism hurts: inflated predictions in psychiatric neuroimaging. *Biol Psychiatry* **75**, 746–748 (2014).
21. Krishnan, A. *et al.* Somatic and vicarious pain are represented by dissociable multivariate brain patterns. *Elife* **5**, (2016).
22. Woo, C.-W. *et al.* Separate neural representations for physical pain and social rejection. *Nat Commun* **5**, 5380 (2014).
23. Wager, T. D. *et al.* An fMRI-based neurologic signature of physical pain. *N Engl J Med* **368**, 1388–1397 (2013).
24. Kay, K. N., Naselaris, T., Prenger, R. J. & Gallant, J. L. Identifying natural images from human brain activity. *Nature* **452**, 352–355 (2008).
25. Huth, A. G., Nishimoto, S., Vu, A. T. & Gallant, J. L. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* **76**, 1210–1224 (2012).
26. Rissman, J., Greely, H. T. & Wagner, A. D. Detecting individual memories through the neural decoding of memory states and past experience. *Proc. Natl. Acad. Sci.* **107**, 9849–9854 (2010).

27. Rosenberg, M. D. *et al.* A neuromarker of sustained attention from whole-brain functional connectivity. *Nat. Neurosci.* **19**, 165–171 (2016).
28. Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E. & Gallant, J. L. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**, 453–458 (2016).
29. Eisenbarth, H., Chang, L. J. & Wager, T. D. Multivariate Brain Prediction of Heart Rate and Skin Conductance Responses to Social Threat. *J. Neurosci.* **36**, 11987–11998 (2016).
30. Kragel, P. A., Koban, L., Barrett, L. F. & Wager, T. D. Representation, Pattern Information, and Brain Signatures: From Neurons to Neuroimaging. *Neuron* **99**, 257–273 (2018).
31. Reddan, M. C., Lindquist, M. A. & Wager, T. D. Effect Size Estimation in Neuroimaging. *JAMA Psychiatry* **74**, 207–208 (2017).
32. Gabrieli, J. D. E., Ghosh, S. S. & Whitfield-Gabrieli, S. Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron* **85**, 11–26 (2015).
33. Davis, K. D. *et al.* Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: challenges and opportunities. *Nat Rev Neurol* **16**, 381–400 (2020).
34. FDA-NIH Biomarker Working Group. *BEST (Biomarkers, EndpointS, and other Tools) Resource*. (Food and Drug Administration (US), 2016).
35. Knutson, B. & Genevsky, A. Neuroforecasting Aggregate Choice. *Curr Dir Psychol Sci* **27**, 110–115 (2018).
36. Venniro, M., Banks, M. L., Heilig, M., Epstein, D. H. & Shaham, Y. Improving translation of animal models of addiction and relapse by reverse translation. *Nat Rev Neurosci* (2020) doi:10.1038/s41583-020-0378-z.
37. Volkow, N. D., Koob, G. & Baler, R. Biomarkers in Substance Use Disorders. *ACS Chem. Neurosci.* **6**, 522–525 (2015).
38. Steele, V. R. *et al.* Machine Learning of Functional Magnetic Resonance Imaging Network Connectivity Predicts Substance Abuse Treatment Completion. *Biol Psychiatry Cogn Neurosci Neuroimaging* **3**, 141–149 (2018).
39. Noori, H. R., Cosa Linan, A. & Spanagel, R. Largely overlapping neuronal substrates of reactivity to drug, gambling, food and sexual cues: A comprehensive meta-analysis. *Eur Neuropsychopharmacol* **26**, 1419–1430 (2016).

671 40. Volkow, N. D., Wise, R. A. & Baler, R. The dopamine motive system: implications for
672 drug and food addiction. *Nat. Rev. Neurosci.* **18**, 741–752 (2017).
673 41. Tang, D. W., Fellows, L. K., Small, D. M. & Dagher, A. Food and drug cues activate
674 similar brain regions: a meta-analysis of functional MRI studies. *Physiol Behav* **106**,
675 317–324 (2012).
676 42. Ziauddeen, H., Farooqi, I. S. & Fletcher, P. C. Obesity and the brain: how convincing
677 is the addiction model? *Nat Rev Neurosci,* **13**, 279-286 (2012).
678 43. Čeko, M., Kragel, P. A., Woo, C.-W., López-Solà, M. & Wager, T. D. Common and
679 stimulus-type-specific brain representations of negative affect. *Nat Neurosci* **25**,
680 760–770 (2022).
681 44. Kragel, P. A., Reddan, M. C., LaBar, K. S. & Wager, T. D. Emotion schemas are
682 embedded in the human visual system. *Sci. Adv.* **5**, eaaw4358 (2019).
683 45. Cosme, D., Zeithamova, D., Stice, E. & Berkman, E. T. Multivariate neural
684 signatures for health neuroscience: Assessing spontaneous regulation during food
685 choice. *Soc Cogn Affect Neurosci* **15**, 1120-1134 (2020).
686 46. Whelan, R. *et al.* Neuropsychosocial profiles of current and future adolescent alcohol
687 misusers. *Nature* **512**, 185–189 (2014).
688 47. Kühn, S. *et al.* Predicting development of adolescent drinking behaviour from whole
689 brain structure at 14 years of age. *Elife* **8**, (2019).
690 48. Potenza, M. N., Sofuoglu, M., Carroll, K. M. & Rounsaville, B. J. Neuroscience of
691 Behavioral and Pharmacological Treatments for Addictions. *Neuron* **69**, 695–712
692 (2011).
693 49. Bartra, O., McGuire, J. T. & Kable, J. W. The valuation system: a coordinate-based
694 meta-analysis of BOLD fMRI experiments examining neural correlates of subjective
695 value. *Neuroimage* **76**, 412–427 (2013).
696 50. Kable, J. W. & Glimcher, P. W. The neural correlates of subjective value during
697 intertemporal choice. *Nat Neurosci* **10**, 1625–1633 (2007).
698 51. Kühn, S. & Gallinat, J. Common biology of craving across legal and illegal drugs–a
699 quantitative meta-analysis of cue-reactivity brain response. *Eur J Neurosci,* **33**,
700 1318-1326 (2011).
701 52. Hanlon, C. A. *et al.* What goes up, can come down: Novel brain stimulation
702 paradigms may attenuate craving and craving-related neural circuitry in substance
703 dependent individuals. *Brain Res* **1628**, 199–209 (2015).

53. Naqvi, N. H., Gaznick, N., Tranel, D. & Bechara, A. The insula: a critical neural substrate for craving and drug seeking under conflict and risk. *Ann N Acad Sci* **1316**, 53–70 (2014).
54. Naqvi, N. H., Rudrauf, D., Damasio, H. & Bechara, A. Damage to the insula disrupts addiction to cigarette smoking. *Science* **315**, 531–534 (2007).
55. Garavan, H. Insula and drug cravings. *Brain Struct Funct* **214**, 593–601 (2010).
56. (Bud) Craig & D, A. How do you feel — now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* **10**, 59–70 (2009).
57. de Araujo, I. E., Schatzker, M. & Small, D. M. Rethinking Food Reward. *Annu Rev Psychol* **71**, 139–164 (2020).
58. Buhle, J. T. *et al.* Cognitive reappraisal of emotion: a meta-analysis of human neuroimaging studies. *Cereb Cortex* **24**, 2981–2990 (2014).
59. Kober, H. *et al.* Prefrontal-striatal pathway underlies cognitive regulation of craving. *Proc Natl Acad Sci U A* **107**, 14811–14816 (2010).
60. Suzuki, S. *et al.* Regulation of Craving and Negative Emotion in Alcohol Use Disorder. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **5**, 239–250 (2020).
61. Goldstein, R. Z. & Volkow, N. D. Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nat Rev Neurosci* **12**, 652–669 (2011).
62. Yokum, S. & Stice, E. Cognitive regulation of food craving: effects of three cognitive reappraisal strategies on neural response to palatable foods. *Int J Obes* **37**, 1565–1570 (2013).
63. Giuliani, N. R., Mann, T., Tomiyama, A. J. & Berkman, E. T. Neural systems underlying the reappraisal of personally craved foods. *J Cogn Neurosci* **26**, 1390–1402 (2014).
64. Silvers, J. A. *et al.* Curbing craving: behavioral and brain evidence that children regulate craving when instructed to do so but have higher baseline craving than adults. *Psychol Sci* **25**, 1932–1942 (2014).
65. Hutcherson, C. A., Plassmann, H., Gross, J. J. & Rangel, A. Cognitive regulation during decision making shifts behavioral control between ventromedial and dorsolateral prefrontal value systems. *J Neurosci* **32**, 13543–13554 (2012).
66. Steele, V. R. Transcranial Magnetic Stimulation as an Interventional Tool for Addiction. *Front Neurosci* **14**, 592343 (2020).

67. Rosenberg, M. D. & Finn, E. S. How to establish robust brain–behavior relationships without thousands of individuals. *Nat Neurosci* **25**, 835-837 (2022).
68. Franken, I. H. A., Hulstijn, K. P., Stam, C. J., Hendriks, V. M. & van den Brink, W. Two new neurophysiological indices of cocaine craving: evoked brain potentials and cue modulated startle reflex. *J Psychopharmacol* **18**, 544–552 (2004).
69. Creswell, K. G. *et al.* Assessing Cigarette Craving With a Squeeze. *Clin Psychol Sci* **7**, 597–611 (2019).

Methods

Participants

The present analysis used the pooled data from N=99 participants (33 female, M_{AGE}=34.1 years, SD_{AGE}=10.8) collected across three independent neuroimaging studies (five different participant groups, three scanners, see **Figure 1** and Supplementary Table 1). Two additional participants in Study 1 were excluded in the original study and for the present analyses due to vomiting and not following task instructions. In Study 2, four additional participants were excluded in the original study and for the present analysis due to unanticipated claustrophobia and non-completion of the task, one due to providing false information during the screening, one due to no responses in some runs of the task, two due to having been scanned in the morning, and one due to excessive movement artefacts. In Study 3, three participants in the user group were excluded due to not understanding or not completing the task, one due to high anxiety and large movement artefacts, one due to being a past but not current cocaine user, and one control participant due to high cocaine craving. Other details regarding the methods and procedures as well as other results from these data sets, focusing on the effects of regulation on behavior and univariate brain responses, are[59,60] or will be (Schafer, Potenza, & Kober, unpublished data) reported elsewhere.

Analyses reported here were not reported previously, and the three studies have not been previously combined. Across studies, participants were recruited using flyers and ads (in newspapers, online bulleting boards, etc.) from communities around Yale and Columbia Universities. Participants were included in drug using groups (N=59, M_{AGE}=34.6 years, SD_{AGE}=11.2, 18 female) based on verified clinical measures (e.g., structured clinical interviews for diagnosis and/or Fagerström test of nicotine dependence). Information on the severity and duration of use is presented in Supplementary Table 1. Individuals were included in “healthy control” groups (N=40, M_{AGE}=33.4 years, SD_{AGE}=10.5, 15 female) if they were (1) age-, sex-, and race-matched to the SUD group in each respective study, (2) did not qualify for any SUD diagnosis or primary psychiatric diagnoses, and (3) did not regularly consume the substance of the SUD group in each respective study (i.e., matched healthy controls for the cigarette-smoking group did not regularly smoke). Participants in the drug use group in Study 1 were heavy daily smokers who smoked an average of 15.7 cigarettes every day. Participants in the drug use groups in Studies 2 and 3 completed diagnostic interviews and fulfilled DSM-IV criteria for substance use disorder (alcohol and

cocaine, respectively). None of the participants were recruited for a treatment study. In Studies 2 and 3, participants were excluded if they were seeking treatment for their drug use. Drug users did not significantly differ from non-users in age, sex, or racial/ethnic background. Compared to non-users, drug (especially cocaine) users had significantly lower years of education (see Supplementary Table 1, 15.5 vs. 14.0 years, $p < 0.001$). We therefore checked that the resulting NCS was not related to education level above and beyond drug use status.

To avoid alterations in brain responses and to ensure craving, we made sure that participants (drug users or controls) were not intoxicated and drug-negative at the time of scanning. In Study 1 (cigarette smokers and their matched controls), participants were asked not to smoke, eat, or drink for at least two hours prior to their study appointment (resulting in a 3–4-hour abstinence at the time of scanning). We then used a breathalyzer to measure exhaled carbon monoxide to verify that participants indeed abstained from smoking, as instructed. Questions were used to verify their abstinence from eating and drinking in the absence of suitable biological verification methods. In addition, participants completed a standard urine toxicology test prior to the scan to verify abstinence from other drugs (opioids, amphetamines, methamphetamines, cocaine, barbiturates, benzodiazepines, PCP [phencyclidine], and THC [the primary psychoactive ingredient in marijuana]). Participants whose test results indicated recent drug or alcohol use were not scanned. In Study 2 (individuals with alcohol use disorder), participants were told to not drink alcohol since the night before, and not to eat or drink anything for at least two hours prior to their study appointment. We then used a breathalyzer to measure exhaled alcohol (the most common proxy for blood alcohol level) to verify that participants indeed abstained from drinking alcohol, as instructed (questions were used to verify their abstinence from eating and non-alcohol drinking). In addition, participants completed a standard urine toxicology test prior to the scan to verify abstinence from other drugs. Again, participants whose test results indicated drug or alcohol use were not scanned. In Study 3 (individuals with cocaine use disorder and their matched controls), participants were part of a larger study, and had spent the prior several nights on an inpatient research unit, where they did not have any access to drugs or alcohol. Drug (and alcohol) abstinence at the time of scan was thus verified by observation. They were also asked not to eat or drink for at least two hours prior to the study participation and were accompanied to the scan directly from the clinical research unit by a research assistant. Thus, no participant was intoxicated during the experiment.

All participants provided informed consent and were paid for their participation in the study. The studies have been approved by the institutional review boards of Columbia and Yale universities and were conducted in compliance with all relevant ethical regulations.

Regulation of Craving task

The Regulation of Craving task is designed to evoke cue-induced craving of drug and food stimuli and test participants' ability to regulate craving[59]. Participants were shown images of drugs and food that were known to induce craving (see Supplementary Tables 2-4, each image was only shown once and order was randomized across and within participants). Additional analyses showed that luminance (β = 0.08, STE = 0.02, t(95) = 4.39, p < 0.001, Cohen's d = 0.45), but not stimulus entropy (β = 0.01, STE = 0.02, t(95) = 0.52, p = 0.60), was significantly associated with NCS responses. However, when controlling for low-level visual features (stimulus luminance and entropy), single trial NCS responses were still significantly associated with craving ratings (β = 0.19, STE = 0.02, t(95) = 9.44, p < 0.001, Cohen's d = 0.97), suggesting that the NCS does not opportunistically rely on these features for prediction of craving ratings.

On each trial, participants were instructed to observe these images in one of two ways. The NOW condition served as a craving baseline, whereby participants were instructed to consider the immediate positive consequences of consuming the pictured drug or food. In the LATER condition, participants were instructed to employ a cognitive strategy drawn from cognitive-behavioral treatments for substance use and obesity, and to consider the negative consequences of repeated consumption of the drug or food.

On each of 100 trials (50 drug, 50 food trials, presented in random order using E-Prime software), participants were presented with a 2-second instructional cue (NOW or LATER) followed by a 6-second presentation of the drug or food image. After a jittered delay period (around 3 second), participants indicated how much they craved the drug or food at that moment ("How much do you want this?") on a 1-5 Likert scale, on which 1 indicated the lowest ("not at all") and 5 the highest ("very much") level of craving. Trials were separated by jittered intervals that followed an exponential distribution, during which a fixation cross was displayed. Prior work[70–72] including the results from the pooled datasets[59,60] have confirmed that participants report less craving for food and drugs in the regulation (LATER) compared to the craving (NOW) condition.

847 ***fMRI data acquisition and preprocessing***

848 Data were collected on three different scanners at Columbia and Yale Universities using
849 different acquisition parameters. Data underwent standard preprocessing in SPM
850 (versions 5, 8, 12) including slice time correction, realignment, motion correction, warping,
851 and smoothing with a 6mm FWHM kernel. No data censoring was used. Differences in
852 acquisition and preprocessing across datasets are in fact helpful in the current context as
853 they ensure that our pooled findings are not dependent on such details[30,73].

854

855 ***fMRI single trial models***

856 For each participant, we first computed a first level general linear model (GLM) using
857 SPM8 and custom scripts (canlab.github.org). These models contained separate
858 regressors for trials in the same condition and rating level (1-5), per run (modeled at 8s
859 duration each). One additional regressor was added to model activity related to ratings
860 (3s) across all trials. Further, 24 movement regressors (estimates for displacement and
861 rotation in three dimensions, their derivatives, squared movement estimates, and
862 derivatives of squared movement estimates) and spike regressors (based on the
863 identification of global outliers, coded as 1 for the outlier time point and zero for all other
864 time points) were added as regressor of no interest to control for motion artifacts.

865 Next, we averaged the resulting beta-images for each participant within each rating
866 level. This resulted in up to five beta images per participant that reflected craving levels
867 from 1 to 5, respectively. If a participant did not have any ratings at a given level, a map
868 for that level was not created for that participant (18 participants had one missing craving
869 level, and four participants had two missing levels). To bring all images to the same scale
870 (thus increasing comparability across studies and scanners) and reduce the impact of
871 potential outliers, each trial-averaged beta image was scaled (divided) by the L2-norm. An
872 inclusive gray-matter-mask was applied to exclude voxels that likely contain white matter
873 or cerebrospinal fluid only.

874

875 ***Training and cross-validation of the NCS***

876 The resulting images for all five levels of craving for each training participant were then
877 used for linear prediction of craving using LASSO-PCR (least absolute shrinkage and
878 selection operator-principal component regression)[74] and default parameters (to avoid
879 overfitting). LASSO-PCR is a machine-learning algorithm that is well suited for prediction
880 of continuous outcomes based on large feature sets such as whole brain imaging data,

which is characterized by substantially higher number of potential predictive features (i.e., voxels) than outcome data points (e.g., rating levels by subjects), and by a non-independence of these features (i.e., voxel activity is strongly covaried across regions and functional networks). LASSO-PCR avoids overfitting by first performing data reduction using principal component regression, thereby identifying brain networks that are characterized by high covariation of voxels. It then performs the LASSO algorithm, which reduces the contribution of less important or more unstable components by shrinking their regression weights towards zero. Voxel weights can be reconstructed based on their scores for the different components, thus rendering the resulting classifier interpretable and applicable to new datasets.

We used a 10-fold cross-validation procedure to evaluate the predictive accuracy of the classifier. Thus, we divided the data into 10 folds that were stratified by studies. Beta images of any given participant (corresponding to all levels of craving) were always held out in the same fold. In each iteration, the classifier was trained on the remaining data and then tested on the hold out data by calculating predicted level of craving (or "NCS response") as the dot product of the trained NCS and each held out beta-image. This *out-of-sample* predicted level of craving was used to assess differences in NCS responses between low and high craving ratings, experimental conditions (instruction, cue type), and drug users versus controls. Since NCS responses reflect predicted ratings, they are in principle on the same scale as craving ratings, but not restricted to whole numbers between 1-5. For training and testing of drug- and food-craving patterns separately, the same procedure was repeated, but only using either drug or food contrast images, respectively.

Bootstrapping and thresholding

To assess the voxels with the most reliable positive or negative weights, we performed a bootstrap test. 10,000 samples with replacements were taken from the paired brain and outcome data and the LASSO-PCR was repeated for each bootstrap sample. Two-tailed, uncorrected *p*-values were calculated for each voxel based on the proportion of weights above or below zero[23,75]. False Discovery Rate (FDR) correction was applied to *p*-values to correct for multiple comparisons across the whole brain. Significant cortical clusters (see Table 1) were automatically labeled using a multimodal cortical parcellation[76], basal ganglia regions are based on[77], cerebellar regions based on[78], and brainstem regions

based on a combination of studies. Large-scale network names are based on an established resting-state parcellation[79].

Permutation tests

Statistical significance of the cross-validated prediction accuracy was assessed using permutation tests. In each of 5000 iterations, craving ratings within each cohort were randomly permuted and training and cross-validation was performed on the permuted data to establish a null-distribution for performance measures (mean square error, root mean square error, mean absolute error, prediction-outcome correlation). Observed performance measures were compared to these permutation-based null-distributions in order to obtain non-parametric *p*-values.

Classification analyses

We used binary receiver operating characteristic (ROC) plots to illustrate the ability of the NCS to separate high versus low levels of craving using forced-choice tests (**Figure 2**), where pattern-expression values (the dot-product of the hold-out beta-images with the classifier weights) were compared for each participant's highest and lowest level of craving and the higher value was chosen as highest level of craving. To separate drug users from non-users (**Figure 4b** and Suppl. Figure 5), pattern-expression values (separately for drug, food, or drug>food contrasts) or each participant were submitted to a single-interval test, thresholded for optimal overall accuracy. Area under the curve (AUC) is provided as a thresholded-independent measure of classification performance. Binomial tests were used to assess the statistical significance of classification accuracy.

Other statistical analyses

Data collection and analysis were not performed blind to the conditions of the experiments. General linear models and t-tests were used to assess NCS effects while statistically controlling for potential confounds such as age, sex, education, head motion, and signals from white matter and ventricles. General linear models and ANOVA were used to test the effects of regulation and cue type on behavioral ratings and NCS responses. Data distribution was assumed to be normal but this was not formally tested. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications[23].

Data Availability

Data, meta-data, and NCS weight maps are available for non-commercial aims at: https://github.com/canlab/Neuroimaging_Pattern_Masks/tree/master/Multivariate_signature_patterns/2022_Koban_NCS_Craving and at: https://doi.org/10.6084/m9.figshare.21174256.

Code Availability

Matlab code for analyses is available at: https://github.com/canlab. Custom code to train and apply the NCS is available at: https://github.com/canlab/Neuroimaging_Pattern_Masks/tree/master/Multivariate_signature_patterns/2022_Koban_NCS_Craving.

Methods-only references

70. Lopez, R. B., Ochsner, K. N. & Kober, H. Brief training in regulation of craving reduces cigarette smoking. J Subst Abuse Treat 138, 108749 (2022).

71. Naqvi, N. H. et al. Cognitive Regulation of Craving in Alcohol-Dependent and Social Drinkers. Alcohol. Clin. Exp. Res. 39, 343–349 (2015).

72. Boswell, R. G., Sun, W., Suzuki, S. & Kober, H. Training in cognitive strategies reduces eating and improves food choice. Proc. Natl. Acad. Sci. 115, E11238–E11247 (2018).

73. Woo, C.-W., Chang, L. J., Lindquist, M. A. & Wager, T. D. Building better biomarkers: brain models in translational neuroimaging | Nature Neuroscience. Nat Neurosci 20, 365–377 (2017).

74. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. J R Stat Soc Ser. B Stat Methodol 58, 267–288 (1996).

75. Wager, T. D., Atlas, L. Y., Leotti, L. A. & Rilling, J. K. Predicting individual differences in placebo analgesia: contributions of brain activity during anticipation and pain experience. J Neurosci 31, 439–452 (2011).

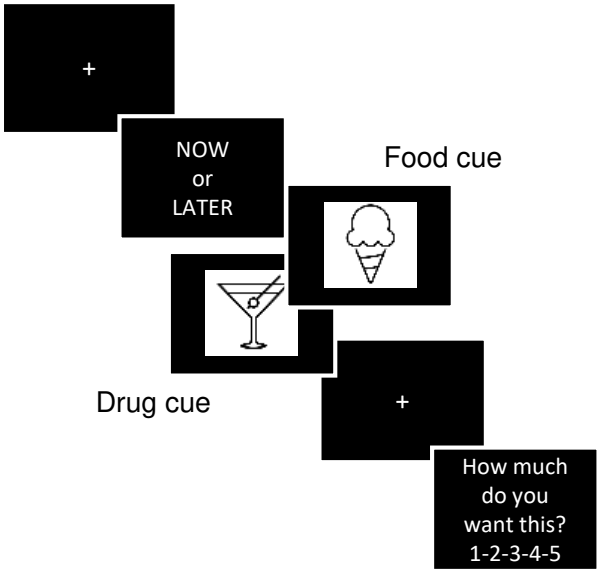
76. Glasser, M. F. et al. A multi-modal parcellation of human cerebral cortex. Nature 536, 171–178 (2016).

77. Pauli, W. M., O'Reilly, R. C., Yarkoni, T. & Wager, T. D. Regional specialization within the human striatum for diverse psychological functions. Proc Natl Acad Sci U A 113, 1907–1912 (2016).

78. Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E. & Ramnani, N. A probabilistic MR atlas of the human cerebellum. Neuroimage 46, 39–46 (2009).

79. Schaefer, A. et al. Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cereb Cortex 28, 3095–3114 (2018).

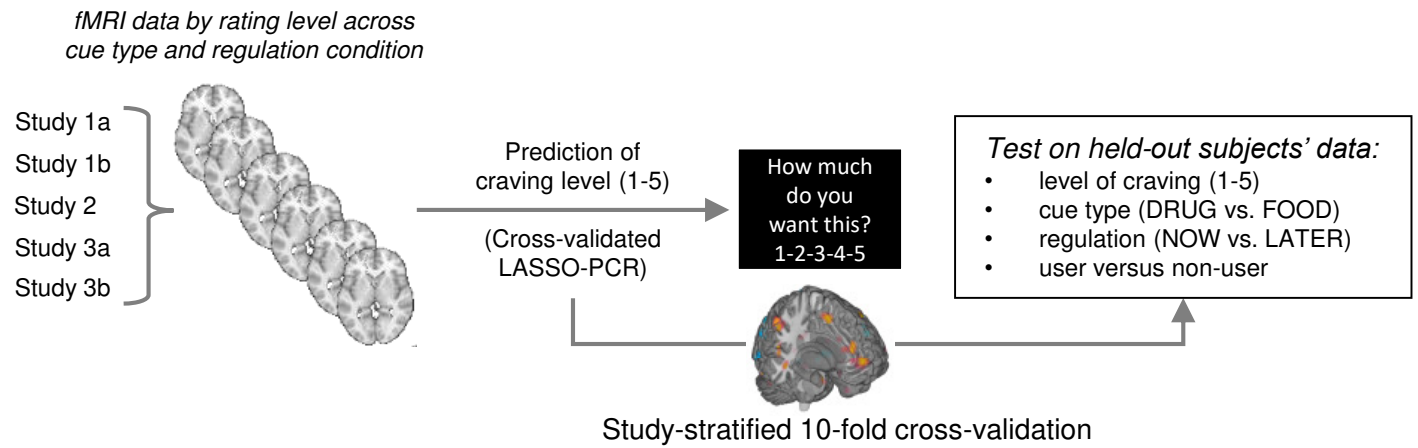
a Regulation of Craving Task design



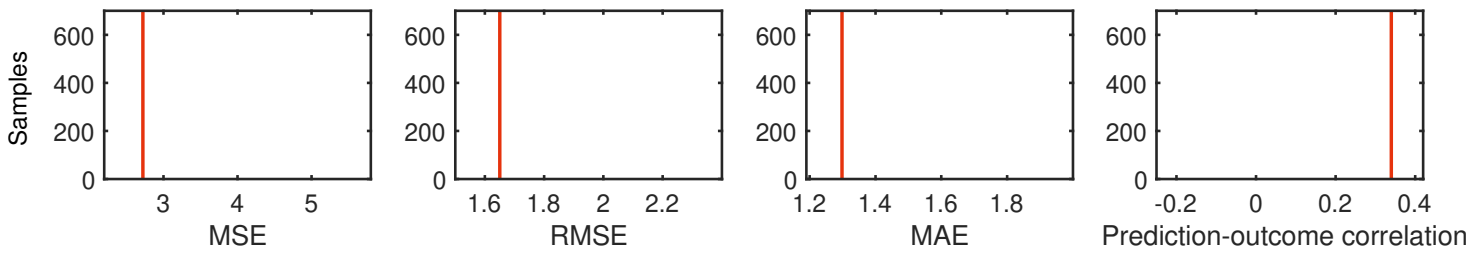
b Study samples

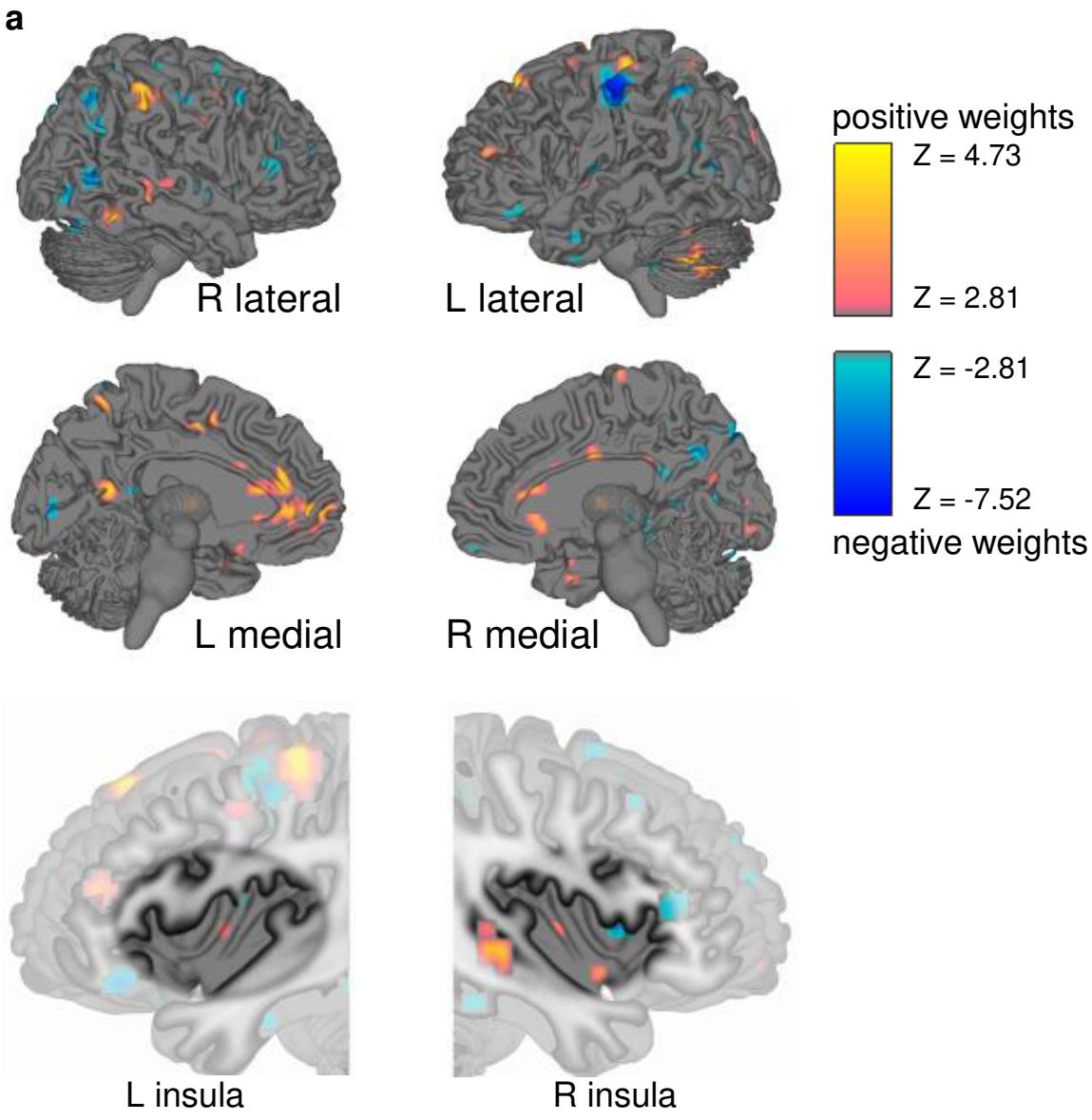
	Population	Cues
S1a	Cigarette smokers (N=21)	Photos of food and cigarettes
S1b	Non-smokers (N=22)	
S2	Alcohol users (N=17)	Photos of food and alcohol
S3a	Cocaine users (N=21)	Photos of food and cocaine
S3b	Non-users (N=18)	

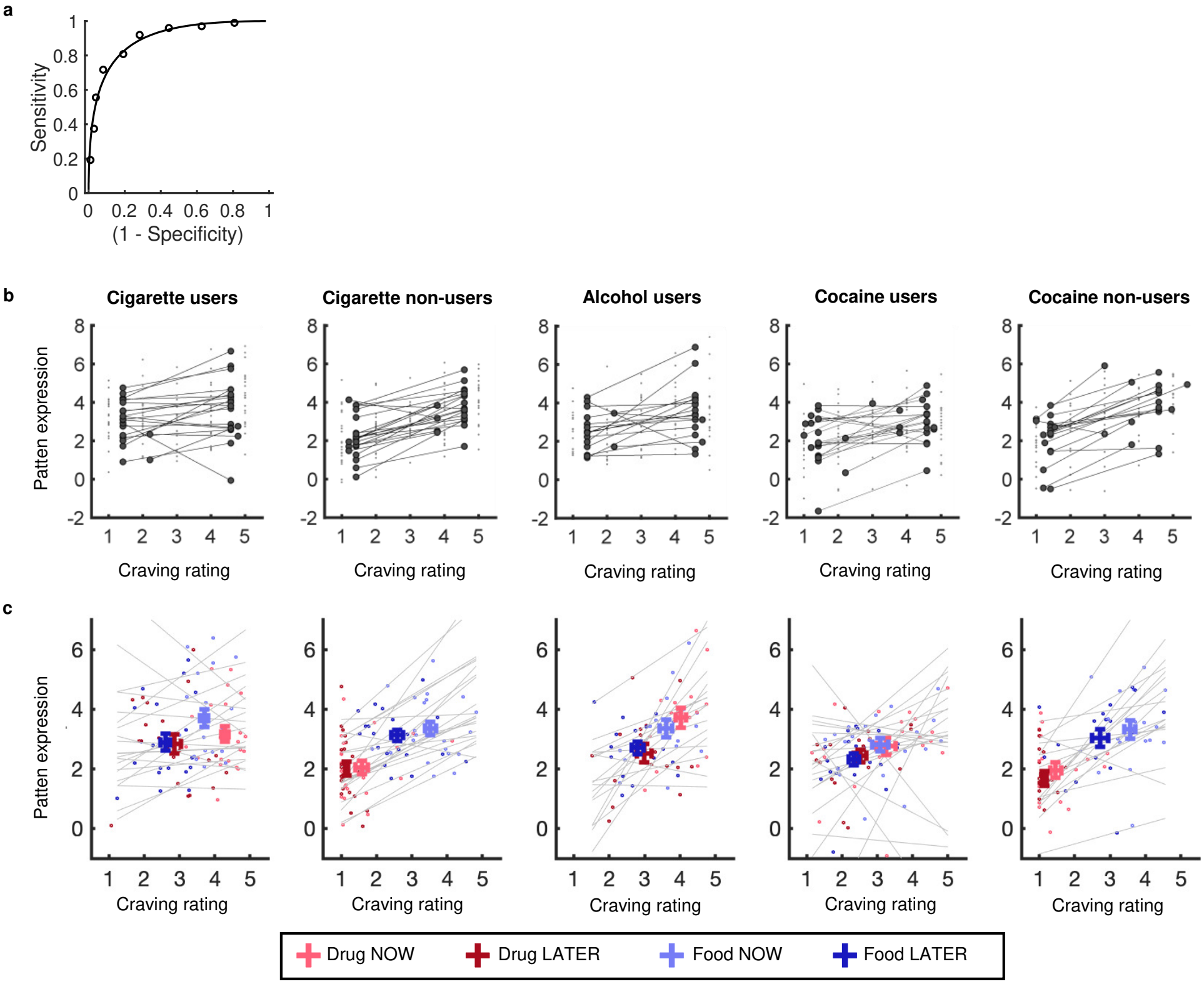
c fMRI-based prediction of craving

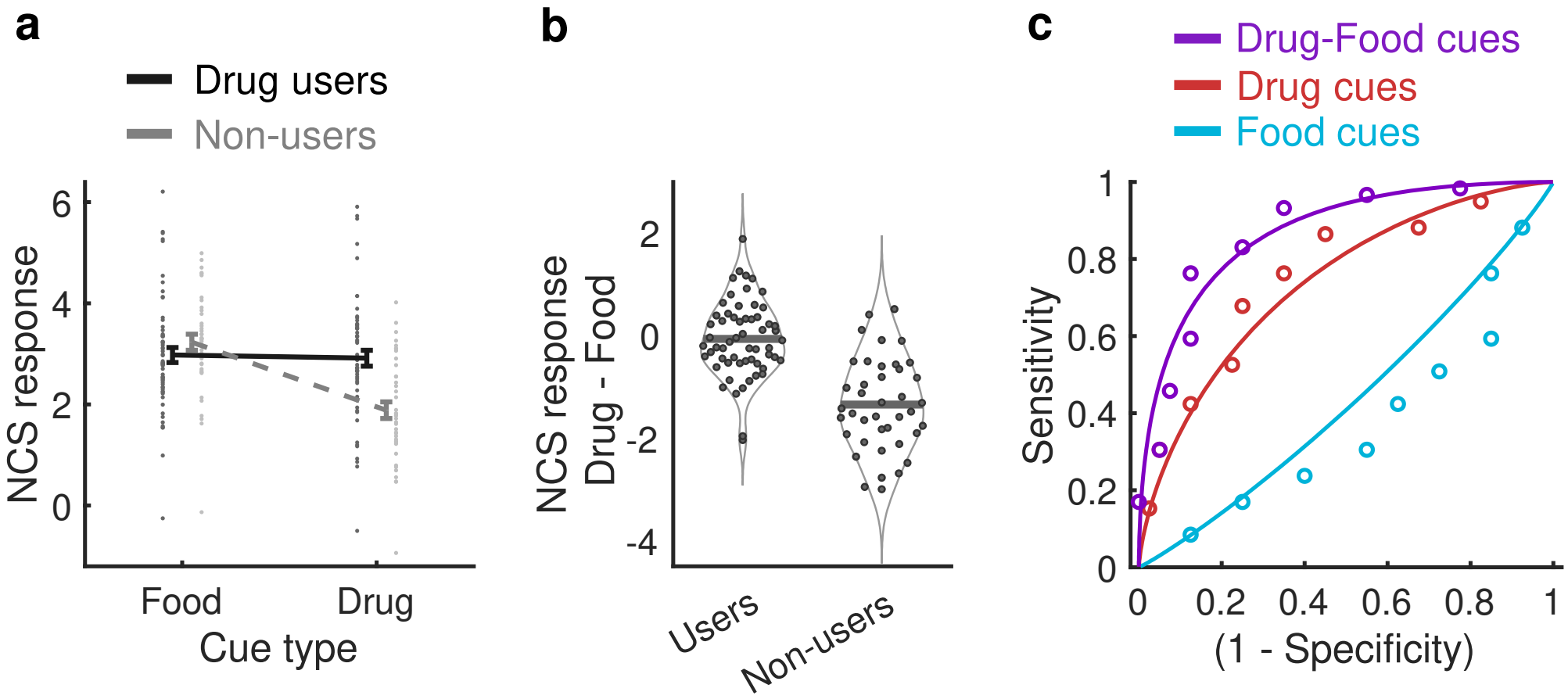


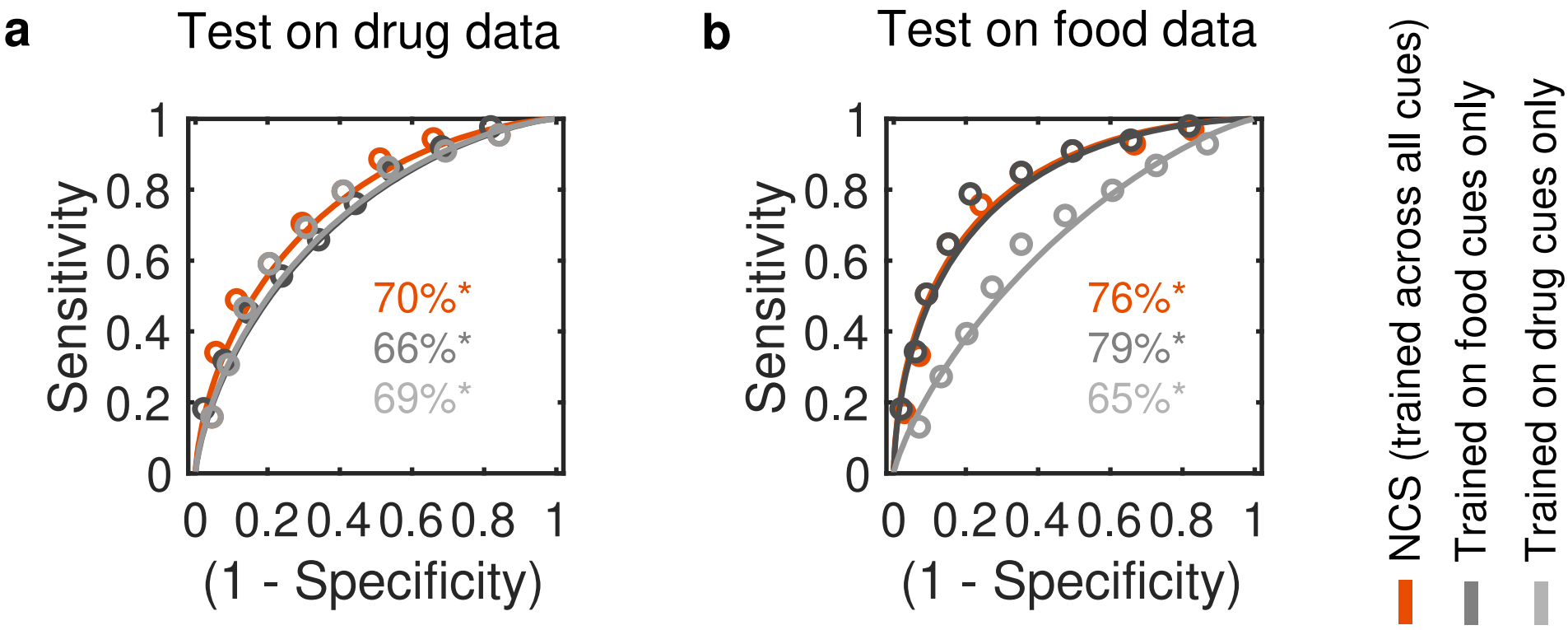
d Permutation test results

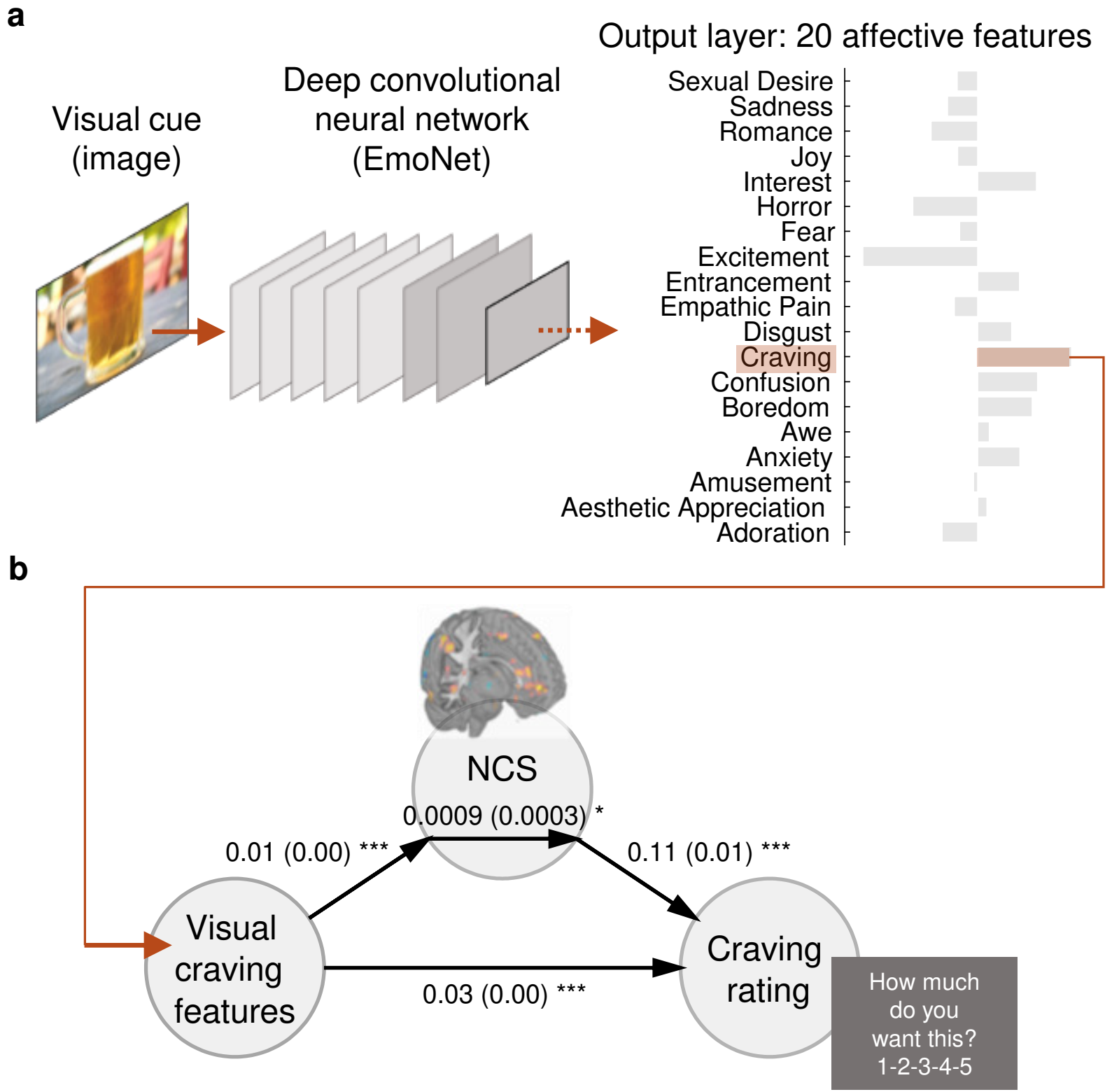



a
R lateral
L lateral
positive weights
Z = 4.73
Z = 2.81
Z = -2.81
Z = -7.52
negative weights
L medial
R medial
L insula
R insula
b
SMA
aMCC
vmPFC
sgACC/VS


a
Sensitivity
(1 - Specificity)
b
Cigarette users
Cigarette non-users
Alcohol users
Cocaine users
Cocaine non-users
Patten expression
Craving rating
c
Patten expression
Craving rating
Drug NOW
Drug LATER
Food NOW
Food LATER


a
Drug users
Non-users
NCS response
6
4
2
0
Food
Drug
Cue type
b
NCS response
Drug - Food
2
0
-2
-4
Users
Non-users
c
Drug-Food cues
Drug cues
Food cues
Sensitivity
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
(1 - Specificity)


a
Test on drug data
Sensitivity
1
0.8
0.6
0.4
0.2
0
0 0.2 0.4 0.6 0.8 1
(1 - Specificity)
70%*
66%*
69%*
b
Test on food data
Sensitivity
1
0.8
0.6
0.4
0.2
0
0 0.2 0.4 0.6 0.8 1
(1 - Specificity)
76%*
79%*
65%*
NCS (trained across all cues)
Trained on food cues only
Trained on drug cues only


a
Visual cue (image)
Deep convolutional neural network (EmoNet)
Output layer: 20 affective features
Sexual Desire
Sadness
Romance
Joy
Interest
Horror
Fear
Excitement
Entrancement
Empathic Pain
Disgust
Craving
Confusion
Boredom
Awe
Anxiety
Amusement
Aesthetic Appreciation
Adoration
b
NCS
0.0009 (0.0003) *
0.01 (0.00) ***
0.11 (0.01) ***
Visual craving features
Craving rating
0.03 (0.00) ***
How much do you want this? 1-2-3-4-5