Multivariate neural signatures for health neuroscience: Assessing spontaneous regulation during food choice

Danielle Cosme[1], Dagmar Zeithamova[1], Eric Stice[2], and Elliot T. Berkman[1]

[1]Department of Psychology, University of Oregon
[2]Department of Psychiatry, Stanford University

Establishing links between neural systems and health can be challenging since there isn't a one-to-one mapping between brain regions and psychological states. Building sensitive and specific predictive models of health-relevant constructs using multivariate activation patterns of brain activation is a promising new direction. We illustrate the potential of this approach by building two "neural signatures" of food craving regulation using multivariate machine learning and, for comparison, a univariate contrast. We applied the signatures to two large validation samples of overweight adults who completed tasks measuring craving regulation ability and valuation during food choice. Across these samples, the machine learning signature was more reliable. This signature decoded craving regulation from food viewing and higher signature expression was associated with less craving. During food choice, expression of the regulation signature was stronger for unhealthy foods and inversely related to subjective value, indicating that participants engaged in craving regulation despite never being instructed to control their cravings. Neural signatures thus have the potential to measure spontaneous engagement of mental processes in the absence of explicit instruction, affording greater ecological validity. We close by discussing the opportunities and challenges of this approach, emphasizing what machine learning tools bring to the field of health neuroscience.

Introduction

A primary goal of health neuroscience is to identify links between neural systems and health outcomes. Recent neuroimaging research has begun shifting from functional localization to prediction of health-relevant outcomes (Bzdok & Ioannidis, 2019; Yarkoni & Westfall, 2017). For example, "brain-as-predictor" approaches (Berkman & Falk, 2013) have found connections between brain activity in regions of interest (ROIs) and future health outcomes such as eating behavior, weight gain, and smoking (Demos et al., 2012; Falk et al., 2011; Giuliani et al., 2015; Hall et al., 2018; Lopez et al., 2014). Yet, several barriers limit the predictive utility of this approach.

First, functional neuroimaging yields massively multivariate data, making it challenging to model using traditional predictive modeling methods, which necessitate a limited set of predictors. Second, selecting ROIs to reduce the number of predictors relies on reverse inference that is rarely valid since brain regions are seldom selective for a single psychological process (Poldrack, 2011). Instead, the vast majority of cognitive functions are achieved in a distributed fashion. Third, neural activation recorded during fMRI tasks with low ecological validity is unlikely to generalize to behavior outside the lab. Despite a push to use more naturalistic stimuli (Burger & Stice, 2012; Rapuano et al., 2016), the tradeoff between experimental control and ecological validity makes it challenging to elicit engagement in target psychological processes in unstructured contexts.

An alternative approach that overcomes these limitations is to develop predictive models of health-relevant psychological processes by examining distributed patterns of activity (Woo et al., 2017). These "neural signatures" can be defined as whole-brain multivariate predictive models that are sensitive and specific predictors of a target psychological process. As described below, the signatures are built using machine learning algorithms that identify multivariate patterns of activation that reliably discriminate between two or more conditions. They are developed in standard space across subjects, allowing the models to be applied to predict outcomes (e.g., picture-induced negative emotion or pain) in new individuals (Chang et al., 2015; Wager et al., 2013). Another benefit of these models is that they can be shared, applied to new or existing data, and iteratively and collectively validated. Ultimately, these models might increase sensitivity to detect within and between person differences as a function of time, health status, or health behavior change intervention. Critically, they also enable measurement of spontaneous engagement in a psychological process absent instruction (Doré et al., 2019), facilitating the use of more ecologically valid, naturalistic paradigms.

However, the best method for developing neural signatures remains unclear; some studies have used machine learning to identify patterns of neural activity associated with psychological processes (Chang et al., 2015; Wager et al., 2013), while others have computed correspondence with meta-analytic maps derived using univariate methods (Doré et al., 2017; Shahane et al., 2018). Here, we compare these two approaches and illustrate potential applications to health neuroscience focusing on craving regulation and food choice

in overweight adults.

Cue-induced food craving is a major driver of unhealthy eating behavior and weight gain (Boswell & Kober, 2016), so we focus on craving regulation as a promising intervention target (Boswell et al., 2018; Stice et al., 2015). Specifically, we focus on cognitive reappraisal, the process of reframing a stimulus to change its affective meaning (Gross, 1998). Reappraisal can be used to flexibly increase the perceived costs of food consumption (e.g., negative health consequences) or benefits of abstinence (e.g., positive health consequences) and decreases self-reported food cravings and the subjective value of food (Boswell et al., 2018; Hutcherson et al., 2012). Reappraising one's responses to food cues, versus passive viewing of such cues, consistently engages regions in the frontoparietal control network (Giuliani et al., 2014; Kober et al., 2010; Yokum & Stice, 2013), making this task well-suited for developing a neural signature of craving regulation.

Here, we take a step towards developing a neural signature that is a sensitive and specific indicator of food craving regulation and can measure spontaneous regulation in unstructured contexts. We explore opportunities and challenges of this approach by comparing two signatures of food craving regulation–one created using machine learning and one created using a univariate contrast–derived from pooled data from four neuroimaging studies. We validate these signatures in two independent samples of overweight adults with healthy eating goals who completed two distinct tasks. We assess trial-by-trial expression of the regulation signatures during a craving regulation task, where participants explicitly control their cravings, as well as during an incentive-compatible food valuation task, where they bid on healthy and unhealthy foods. Bidding on actual food is expected to produce goal conflicts in these participants because they were strongly motivated to change their eating behavior, and therefore provides an excellent opportunity to assess spontaneous craving regulation.

Our aims were to investigate the degree to which these signatures 1) generalize to new samples completing the same craving regulation task, 2) generalize to distinct, yet related task contexts, and 3) differentiate individuals. Our purpose was to explore the feasibility of this approach, so we focus on descriptive statistics and effect size estimation and do not report null-hypothesis significance tests (Cumming, 2014). We discuss the results in the broader context of what predictive models offer health neuroscience. All code is available online (https://osf.io/7jf82/).

Methods

Participants.
We used data from 296 participants who completed a craving regulation (CR) task, divided into three samples: the neural signature development sample (N = 166), a partial validation sample (N = 94), of which 50 participants were also included in the neural signature development sample, and a complete validation sample (N = 86) (see Table 1; Figure 3A). The neural signature development sample included ado-

lescents and adults from four studies on craving regulation; two are previously published (Giuliani et al., 2014; Giuliani & Pfeifer, 2015). The partial validation sample (Cosme et al., in press) included overweight and obese adults enrolled in a healthy eating intervention who completed the CR task and a food valuation (FV) task. Because some of the participants (N = 50) in this sample were included the development of the neural signature, it allowed us to test an "upper" bound for generalizability within, and most interestingly, between tasks, due to the sample overlap. Participants in the complete validation sample (N = 86) were also overweight and obese adults enrolled in an ongoing healthy eating intervention. They completed the CR and FV tasks at baseline, prior to intervention. Because no data from this sample were used in signature development, it offers an estimate of a "lower" bound for within- and between-tasks signature generalizability.

All participants were right-handed, MRI-eligible, native English speakers. Participants were excluded per task if they exhibited excessive motion (defined below; $CR_{partial} = 2$, $CR_{complete} = 4$, $FV_{partial} = 1$, $FV_{complete} = 6$), had low data quality due to visual artifacts ($CR_{complete} = 2$, $FV_{complete} = 1$), were missing responses due to a technical error ($CR_{partial} = 2$, $CR_{complete} = 3$, $FV_{partial} = 1$, $FV_{complete} = 2$), did not comply with task instructions ($CR_{partial} = 2$, $FV_{partial} = 2$), had structural abnormalities ($CR_{partial} = 1$, $FV_{partial} = 1$), or did not complete a given task ($FV_{complete} = 1$). All available data was used when participants were missing a task run or trials. This yielded the total sample sizes listed in Table 1. These studies were approved by the University of Oregon Institutional Review Board and participants gave written informed consent (or had parental consent for minors) and were compensated for participation.

Craving regulation (CR) task.
Participants completed a CR task while undergoing functional MRI. On each trial, participants either responded naturally ("look") or reappraised their desire ("regulate") by visualizing the negative consequences of eating that food (e.g., risk for diabetes, weight gain, stomach ache). The task consisted of 80 trials (Figure 1). Craved and not-craved food categories were selected by each participant prior to the task, and each participant viewed a unique set of food images during the task to maximize generalizability (Westfall et al., 2016). Participants reappraised images of craved foods or responded naturally to images of high-calorie craved and not-craved foods, or neutral low-calorie foods (20 trials each). The analyses reported here include only craved food trials.

Food valuation (FV) task.
Participants in both validation samples completed an incentive-compatible willingness-to-pay task to assess subjective valuation of snack foods (Figure 2). The task is an economic auction where participants view images of healthy and unhealthy foods (60 total in the partial validation sample, 64 in the complete validation sample) and bid on each. Participants in the partial validation sample viewed the same images, whereas complete validation participants saw unique

Table 1
Sample Characteristics

	neural signature development sample		partial validation sample		complete validation sample	
Craving regulation task	N = 166		N = 87		N = 77	
Food valuation task			N = 89		N = 76	
	M	*SD*	*M*	*SD*	*M*	*SD*
Age	30	9.54	39.22	3.53	36.47	12.1
BMI	27.78	8.05	31.48	3.92	31.4	4.91
	females	males	females	males	females	males
Sex	146	20	76	15	48	31
Race / Ethnicity						
American Indian or Alaskan Native	0.0%		1.1%		1.3%	
Asian	2.4%		1.1%		1.3%	
Black or African American	0.6%		3.3%		2.5%	
Hispanic or Latinx	4.2%		7.7%		5.0%	
More than one race	1.8%		0.0%		6.3%	
Unknown or no response	2.4%		3.3%		8.8%	
White	88.6%		83.5%		75%	

Note. The sample sizes (N) reported here are the final ns after exclusions.
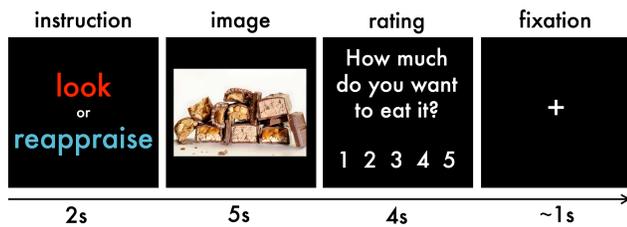
Fig. 1. Craving regulation (CR) task design for the neural signature development and partial validation samples. Each trial consisted of a 2s instruction period, a 5s image presentation, and a 4s craving rating period (1 = no desire to eat the food, 5 = strong desire to eat the food). Participants in the complete validation sample had 2.5s to rate foods and used a 4-point scale with the same anchors. Between trials, participants viewed a jittered fixation cross (*M* = 1s for neural signature development and partial validation samples, *M* = 4.1s for the complete validation sample).

Fig. 2. Food valuation (FV) task design for the partial validation sample. Each trial consisted of a 4s snack food presentation, followed by a 4s bid period. Snack foods were either healthy (e.g., carrot sticks, yogurt) or unhealthy (e.g., candy, chips). In the complete validation sample, the bid period lasted 2.5s and they made bids ranging from $ 0-$ 1.50. All trials ended with a jittered fixation cross (*M* = 4.38s).

images based on pre-session ratings. They rated the palatability of 100 food images and their top 16 healthy and 16 unhealthy images became their "liked" foods, whereas their bottom 16 healthy and 16 unhealthy images became their "disliked" foods. Participants were given $ 2 ($ 1.50 in the complete validation sample) to buy a snack. One trial was randomly selected and enacted. Participants received the foods if their bid was greater than or equal to the randomly selected bid value. Otherwise, they only received the money. The optimal strategy is to bid the true amount one is willing to pay for each item.

Neuroimaging data acquisition and preprocessing.
Neuroimaging data were acquired on either a 3T Siemens Allegra or Skyra scanner at the University of Oregon Lewis Center for Neuroimaging. Both validation samples were collected on the Skyra. Sequence parameters appear in Supplementary material. Neuroimaging data were preprocessed using fMRIPrep 1.1.4 (Esteban et al., 2019). Preprocessing details appear in Supplementary material, but briefly, anatomical images were segmented and normalized to MNI space using FreeSurfer Fischl, 2012); functional images were susceptibility distortion corrected, realigned, and coregistered to

the normalized anatomical images. Normalized functional data were then smoothed (6mm^3 FWHM) in SPM12.

Neural signatures of craving regulation.
Event-related condition effects were estimated in first-level analyses using a fixed-effects general linear model and a canonical hemodynamic response function. Regressors modeled each experimental condition (regulate craved, look craved, look not-craved, look neutral) during image presentation. Additional regressors of no interest were added for the instruction and rating periods. Five motion regressors were modeled as covariates of no interest. Realignment parameters were transformed into Euclidean distance for translation and rotation separately; we also included the displacement derivative of each. Another "trash" regressor marked images with motion artifacts (e.g., striping) identified via automated motion assessment (Cosme et al., 2018) and visual inspection. Data were high-pass filtered at 128seconds and temporal autocorrelation was modeled using FAST (Corbin et al., 2018). The resulting contrast maps for regulate craved > baseline and look craved > baseline were then used to develop two predictive models (Figure 3).

A samples

neural signature development sample

| n = 57 | n = 30 | n = 29 | n = 50 |

ROC

| partial validation sample | complete validation sample |

| n = 87 | n = 89 | n = 77 | n = 76 |

ROC   FV   ROC   FV

B neural signature development
subject-level analyses

participant 1

look   regulate

participant 2

look   regulate

⋮

participant n

look   regulate

machine learning classifier

univariate contrast

multivariate signature

univariate signature

C neural signature validation
trial-level analyses

task beta-series

trial n

trial n

⋮

trial n

dot product
signature · beta

pattern expression (PEV)

pattern expression (PEV)

ROC task
multivariate PEVs
univariate PEVs

FV task
multivariate PEVs
univariate PEVs

Fig. 3. Samples and analytic overview. A) The neural signature development sample included participants from four different neuroimaging studies who completed the craving regulation (CR) task. One of these studies included participants (n = 50; dotted rectangle) from the partial validation sample. Participants in the partial validation sample and the complete validation sample completed the CR task as well as the food valuation (FV) task. B) Average instruction effects (i.e., mean look > baseline and regulate > baseline) from participants in the neural signature development sample were used to create two neural signatures. The multivariate signature was developed using a machine learning classifier and 5-fold cross-validation to decode instruction; the univariate signature was created using the group-level regulate > look contrast. C) The resulting neural signatures were applied to trial-level data from the partial validation and complete validation samples for the CR and FV tasks by taking the dot product of each signature and each trial beta image. This process yielded scalar pattern expression values for each trial for each participant and task.

Multivariate machine learning signature. We trained three machine learning algorithms (logistic regression, ridge regression, support vector machine) to distinguish look from regulate using NLTools (https://github.com/cosanlab/nltools). Inputs were average effects for each instruction condition for each participant and we used 5-fold cross-validation. These classifiers had equivalent accuracy (Table 2); we selected the logistic regression classifier to facilitate interpretability.

Univariate contrast signature. One aim was to begin to identify the conditions under which maps derived using classifiers yield different predictions from those derived using univariate methods. The group-level regulate > look contrast serves as the univariate analysis-based neural signature (Figure 3). We also compare these neural signatures to one created from the group-level regulate > baseline contrast in Supplementary material.

Neuroimaging first-level analysis in the validation samples.
First-level statistical analyses for the tasks were conducted in MNI space in SPM12. For each task, trials were entered in the models as separate regressors (rather than grouped by condition). For the CR task, trial duration was the image

presentation duration (5s); for the food valuation task, trial duration was image onset to bid response. All task runs with >10% unusable volumes were excluded ($CR_{partial} = 3$, $CR_{complete} = 3$, $FV_{complete} = 3$). Participants with 2 or more runs excluded were excluded completely. All data were high-pass filtered at 128seconds and temporal autocorrelation was modeled using FAST. The resulting statistical maps for each trial were concatenated to create a beta-series for each task (Rissman et al., 2004). Because motion artifacts may persist in the beta-series, we calculated the mean global intensity for each beta map and excluded trials that were more than 3 SDs from the grand mean across participants within each sample and task separately ($CR_{partial} = 2.3\%$, $CR_{complete} = 1.3\%$, $FV_{partial} = 1.2\%$, $FV_{complete} = 1.1\%$). We also visually inspected each image and excluded trials with visible striping artifacts ($CR_{partial} = 3.8\%$, $CR_{complete} = 1.0\%$, $FV_{partial} = 2.1\%$, $FV_{complete} = 1.6\%$).

Pattern expression values.
To assess the degree to which participants expressed the neural signatures, we treated each neural signature and the trial-level beta maps as vectors of weights and took the dot product for each trial, task, and participant (Doré et al., 2017; Shahane et al., 2018). This process yields one scalar "pattern expression value" for each trial. We standardized the pattern

Table 2

Classification Accuracy for Candidate Multivariate Neural Signatures

Classifier	Accuracy [95% CI]	Sensitivity	Specificity	AUC	PPV
logistic regression	0.82 [0.74, 0.90]	0.81	0.83	0.90	0.83
ridge regression	0.82 [0.74, 0.90]	0.81	0.83	0.90	0.83
support vector machine	0.83 [0.73, 0.93]	0.86	0.80	0.90	0.81

Note. AUC = area under the curve, PPV = positive predictive value

Table 3

Descriptive Statistics for Craving Ratings in the Regulation of Craving Task

	partial validation sample			complete validation sample		
Instruction	*n*	*M*	*SD*	*n*	*M*	*SD*
look	1688	3.64	1.15	1460	3.25	0.86
regulate	1688	2.39	1.08	1425	2.28	0.83

Note. n = number of trials. The partial validation sample used a 1-5 rating scale; the complete validation sample used a 1-4 rating scale.

expression values by dividing by the SD within participant and neural signature to account for different variances across participants. This maintains the decision or contrast boundary, such that values above 0 indicate evidence for "regulate," whereas values below 0 indicate evidence for "look."

Specification curves.
We conducted a series of specification curve analyses (Simonsohn et al., 2015) to determine 1) which neural signatures (if any) explained additional variance in trial-level criterion variables (i.e., craving ratings or bids) beyond task condition and 2) determine whether multivariate and univariate signatures account for unique or overlapping variance. For each sample and task, we specified 13 unique nested multilevel models regressing criterions on the fixed effects of condition (instruction in the CR task, health in the FV task), pattern expression value from each signature, and the interactions between condition and each signature. Participant intercepts were treated as random effects. We compared model fit to a base model including condition only using the Akaike Information Criterion (AIC). We then organized the models by AIC, plotted each model and which variables were included in it, and visualized which models had lower AIC values than the base model.

Results and discussion

Aim 1: Construct validity and generalization to new individuals in the CR task.
Participants reported lower cravings when regulating via cognitive reappraisal than when just viewing (Table 3). This is consistent with other evidence that instructed cognitive reappraisal reduces craving (Boswell et al., 2018; Giuliani et al., 2014).

We applied the neural signatures to each trial for each participant in the validation samples, and coded the standardized pattern expression values >0 as relative evidence for regulation and values <0 as relative evidence for the control condition (i.e., look). Classification accuracy was greater for
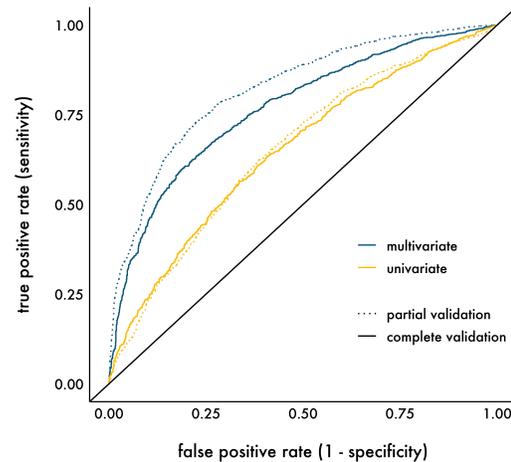
Fig. 4. Receiver operating characteristic curves as a function of signature type (multivariate classifier or univariate contrast), and sample (partial validation or complete validation sample).

the multivariate signature than the univariate signature, but both correctly classified instruction (look or regulate) above chance (Table 4; see Supplementary material for empirically derived chance accuracy estimates). Accuracy was higher in the partial validation sample than the complete validation sample for the multivariate classifier but not the univariate contrast. Receiver operating characteristic curves depicting sensitivity and specificity are visualized in Figure 4.

When averaged across trials, regulation pattern expression (i.e., the relative evidence for regulation) was higher in the regulate than look condition for the multivariate (partial validation M_{diff} = 1.09, 95% CI [1.03, 1.15]; complete validation M_{diff} = 0.93, 95% CI [0.87, 0.99]) and univariate signatures in both samples (partial validation M_{diff} = 0.49, 95% CI [0.43, 0.54]; complete validation M_{diff} = 0.51, 95% CI [0.44, 0.57]). These data are visualized in Figure 5; descriptive statistics are reported in Supplementary material.

We next examined whether the neural signatures were related to trial-level craving ratings. Collapsing across instruction, higher mean regulation pattern expression was associated with lower mean craving, for the multivariate (partial validation b = -0.26, 95% CI [-0.29, -0.23]; complete validation b = -0.35, 95% CI [-0.39, -0.31]) and univariate signatures (partial validation b = -0.08, 95% CI [-0.11, -0.06]; complete validation b = -0.15, 95% CI [-0.19, -0.11]) in both samples. This effect was strongest for the multivariate signature (Figure 6A,C). However, in the partial validation sample, this relationship was moderated by instruction (Figure 6B,D). When regulating, greater regulation pattern expression was associated with higher, rather than lower, cravings (multivari-

Table 4

Prediction Accuracy as a Function of Sample and Neural Signature Type

Sample	Signature Type	Sensitivity	Specificity	Accuracy [95% CI]
partial validation	multivariate	0.77	0.73	0.75 [0.74, 0.77]
	univariate	0.71	0.50	0.60 [0.59, 0.62]
complete validation	multivariate	0.63	0.76	0.69 [0.67, 0.71]
	univariate	0.62	0.60	0.61 [0.59, 0.63]

Note. Chance accuracy is 0.5.
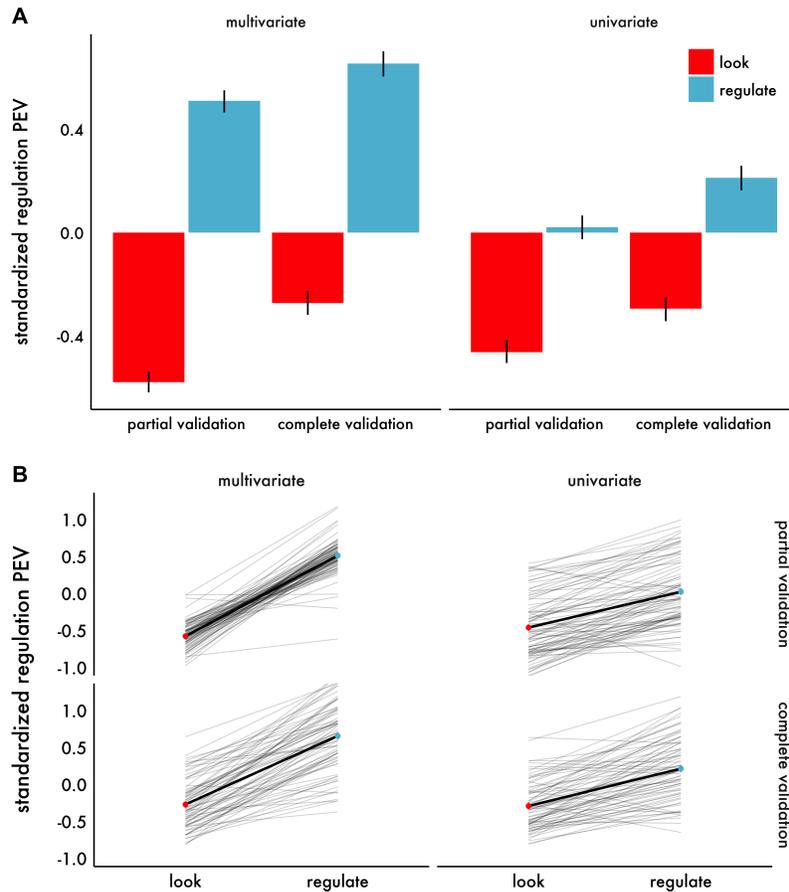
Fig. 5. A) Mean difference in standardized regulation pattern expression values as a function of instruction (look or regulate), signature type (multivariate classifier or univariate contrast), and sample (partial validation or complete validation sample). B) Group means are overlaid on individual participant means; each thin line represents a single participant. Higher positive values represent relatively higher evidence for regulation, whereas lower negative values represent relatively higher evidence for viewing; zero is the decision boundary between conditions. Error bars are 95% confidence intervals across trials. Pattern expression values are standardized within participant and signature type. PEV = pattern expression value.

ate $b = 0.12$, 95% CI [0.06, 0.18]; univariate $b = 0.07$, 95% CI [0.00, 0.14]; see Table S5 for all model parameter estimates).

The specification curves revealed that neural signatures accounted for additional variance in trial-level craving ratings beyond the base model (instruction only) across both samples (Figure 7). Of these 5 better fitting models, 100% included the multivariate signature and 80% included the univariate signature. Eighty percent included both, suggesting that the signatures account for unique variance. Consistent with this, the neural signatures were moderately correlated (Table 5). With respect to the interaction between instruction and neural signature expression, 40% included the multivariate signature interaction and 20% included the univariate signature interaction.

Table 5

Repeated Measures Correlations in the Regulation of Craving Task Between Neural Signatures as a Function of Sample

Sample	df	r [95% CI]
partial validation	3084	0.59 [0.57, 0.62]
complete validation	2736	0.59 [0.56, 0.61]

Note. 95% confidence intervals are bracketed. Repeated measures correlations include all trial-level data and adjust for trials nested within participant using multilevel modeling.
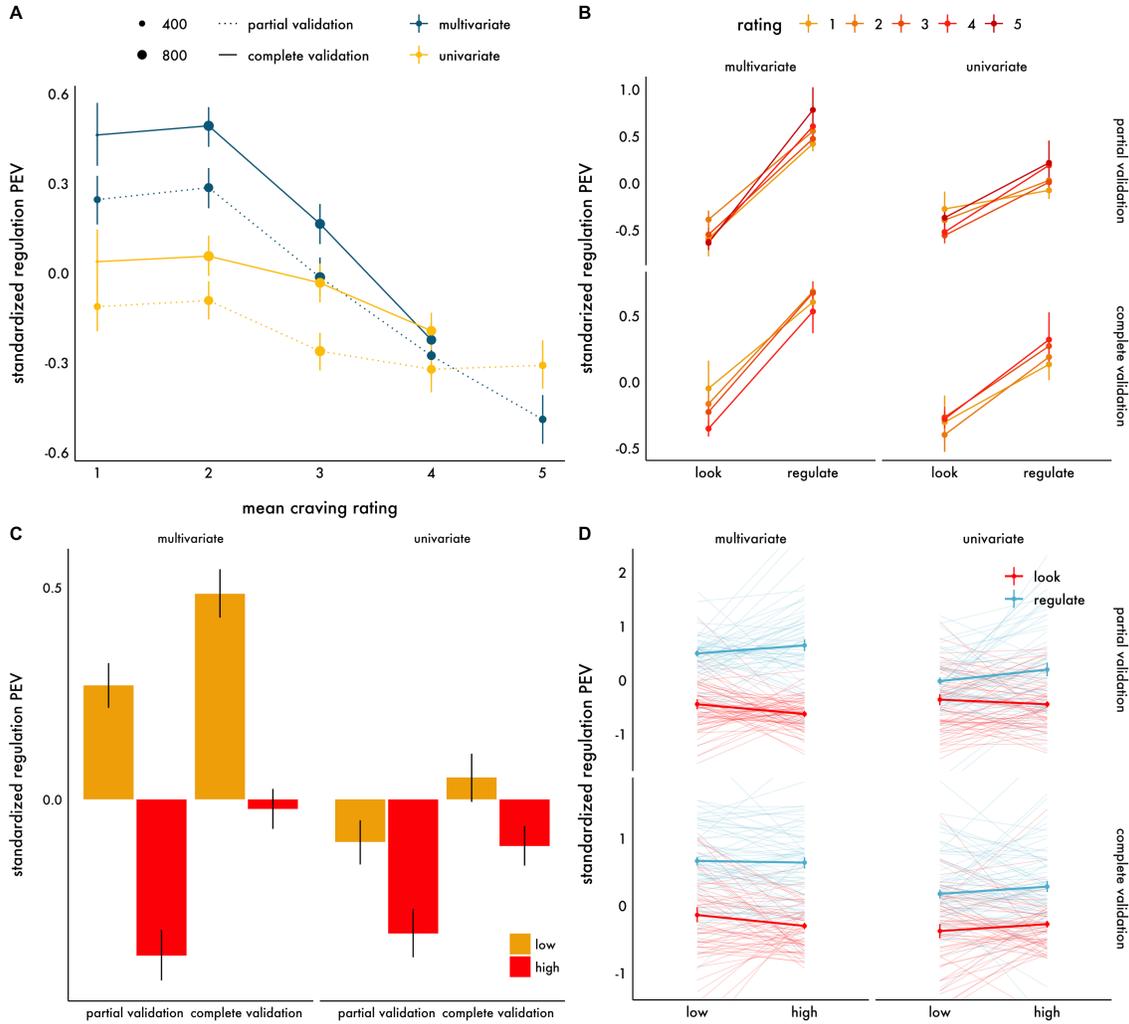
Fig. 6. The relationship between craving ratings and mean standardized regulation pattern expression values as a function of signature type (multivariate classifier or univariate contrast) and sample (partial validation or complete validation sample). Top panel: Continuous craving ratings A) collapsed across instruction (look or regulate), scaled by the number of observations in each rating category, represented by the size of the point, and B) as a function of instruction. Bottom panel: Mean dichotomized craving ratings (low = lower than scale midpoint, high = higher than scale midpoint) C) collapsed across instruction, and D) as a function of instruction. In D, group means are overlaid on individual participant means; each thin line represents a participant. Error bars are 95% confidence intervals across all trials. The partial validation sample used a 1-5 craving rating scale, whereas the complete validation sample used a 1-4 craving rating scale. PEV = pattern expression value.
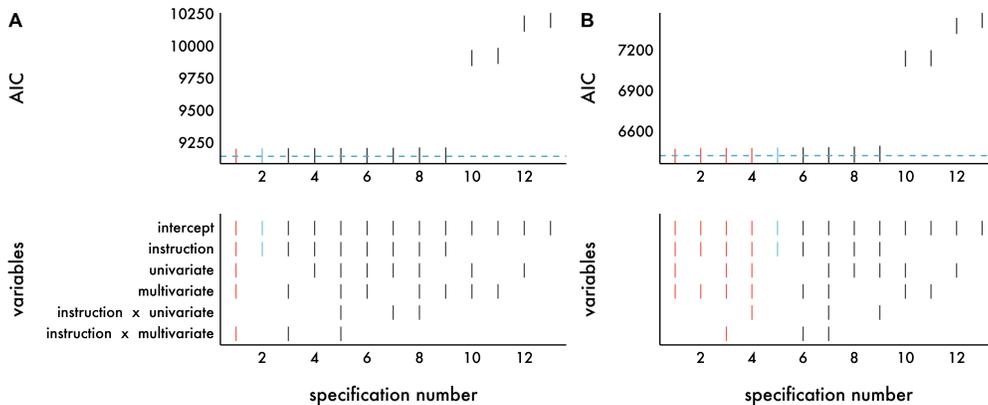
Fig. 7. Specification curves of 13 unique models regressing trial-level craving ratings on predictors ordered based on model fit (AIC) for A) the partial validation sample, and B) the complete validation sample. Each column corresponds to a single model specification. The AIC value for each model specification is plotted in top panels and the variables included in each model are visualized in the bottom panels. Because each panel is ordered based on AIC, specification numbers do not necessarily correspond to the same model specifications in each panel. The base model, which included instruction (look or regulate) as the only predictor, is highlighted in blue and the dotted blue line represents the AIC for this model. Models with AIC values lower than the base model are highlighted in red. Potential variables in each model included: intercept, instruction (look or regulate), standardized pattern expression values for the multivariate and univariate signatures, and the interaction between instruction and each signature type.

Table 6

Descriptive Statistics for Bids in the Food Valuation Task

Instruction	partial validation sample			complete validation sample		
	n	M	SD	n	M	SD
healthy	2623	0.96	0.65	2314	0.69	0.54
unhealthy	2627	0.65	0.63	2314	0.57	0.53

Note. n = number of trials. The bid range for the partial validation sample was $0-2$ and $0-1.50$ for the complete validation sample.

Discussion. These results collectively establish the initial construct validity for the multivariate and univariate signatures. Both signatures clearly differentiated craving regulation from viewing and classified conditions on a trial-by-trial basis across both samples. These signatures also generalized to new individuals, confirming the viability of cross-subject classification of craving regulation. Overall, the multivariate signature was more reliable, indicating that models based on multivariate patterns might contain additional information beyond univariate activation models. Though the signatures clearly differentiated regulation from viewing and tracked with craving ratings across instruction conditions, they were not strongly associated with ratings in the regulation condition and were modestly negatively associated in the look condition. It may be useful to develop neural signatures that incorporate cravings (e.g., with support vector regression) to maximize differentiation between ratings, and indeed, this is the approach taken by Chang et al., (2015) and Wager et al., (2013). It is notable that these signatures were created using average effects for each participant, so trial-level accuracy may be improved by developing machine learning signatures using all available trials rather than averages to increase variability during training.

Aim 2: Generalization to the food valuation task and assessment of spontaneous craving regulation.

Participants bid lower on unhealthy foods than healthy foods (Table 6), which is expected given participants were enrolled in healthy eating programs. We tested whether participants spontaneously regulated their cravings during this task by examining regulation pattern expression during valuation.

In contrast to the distinct differences in regulation pattern expression for instruction in the CR task, pattern expression differences for healthy and unhealthy foods during valuation were smaller. Though small, they were reliable across samples for the multivariate signature, with higher expression on unhealthy than healthy foods (partial validation M_{diff} = 0.15, 95% CI [0.11, 0.20]; complete validation M_{diff} = 0.13, 95% CI [0.08, 0.18]). This pattern was observed for the univariate signature, but only within the complete validation sample (partial validation M_{diff} = -.02, 95% CI [-0.08, 0.03]; complete validation M_{diff} = 0.09, 95% CI [0.04, 0.14]). These effects are visualized in Figure 8A-B; descriptive statistics are reported in Supplementary material.

In the complete validation sample, which contained both liked and relatively disliked foods, these effects were not moderated by palatability. Unhealthy foods were associated with stronger regulation pattern expression for both liked and relatively disliked foods (Figure 8C-D). We expected that liked foods would elicit greater pattern expression than relatively disliked foods, but only observed this for the univariate signature (multivariate M_{diff} = -0.00, 95% CI [-0.05, 0.05]; univariate M_{diff} = 0.14, 95% CI = [0.09, 0.20]).

We next turned to the bids. We anticipated stronger regulation pattern expression would be associated with lower bids to the extent that participants spontaneously regulate their cravings. Because relatively disliked foods (included in the complete validation sample only) are unlikely to produce regulatory goal conflicts, we expected to observe greater regulation pattern expression for liked foods only in the complete validation sample. This was the case for the multivariate signature. Stronger regulation pattern expression was indeed associated with lower bids (Figure 9; Table S10) across trials in the partial validation sample (b = -0.19, 95% CI [-0.22, -0.15]) and for liked food trials in the complete validation sample (b = -0.07, 95% CI [-0.15, 0.01]). Surprisingly, the opposite pattern was observed for the univariate signature; stronger regulation pattern expression was associated with higher bids in both samples (partial validation b = 0.08, 95% CI [0.03, 0.12], complete validation liked trials b = 0.16, 95% CI [0.06, 0.25]).

Food type moderated the relationship between bid value and pattern expression, but only in the complete validation sample (Figure 9B,D). As expected, for liked foods, the difference between expression for high and low bids was stronger for unhealthy foods for the multivariate signature (b = 0.07, 95% CI [-0.10, 0.24]). The opposite was observed for the univariate signature. This difference in bids was stronger for healthy foods rather than unhealthy foods (b = -0.13, 95% CI [-0.32, 0.06]; Table S11 for all results).

The specification curves for the food valuation task (Figure 10) mirrored the results for the craving regulation task. The neural signatures accounted for additional variance in trial-level bids beyond the base model (food type only) across samples. Of these 14 better fitting models, 79% included the multivariate signature, 86% included the univariate signature, and 64% included both. As with the CR task, this suggests they account for unique variance, which is supported by the moderate correlations between signatures (Table 7). With respect to the interaction between food type and neural signature expression, 36% included the multivariate signature interaction, 36% included the univariate signature interaction, and 14% included both.

Discussion. These results suggest it is possible to use these signatures to index spontaneous regulation in an ecologically valid food valuation task absent explicit regulation instructions. Critically, this enables health neuroscientists to assess regulation in contexts that more closely resemble the real world. Since this task is substantially different from the CR task (e.g., the former requires complex decision making), these results support the generalizability of the neural signature approach. Overall, the multivariate signature was
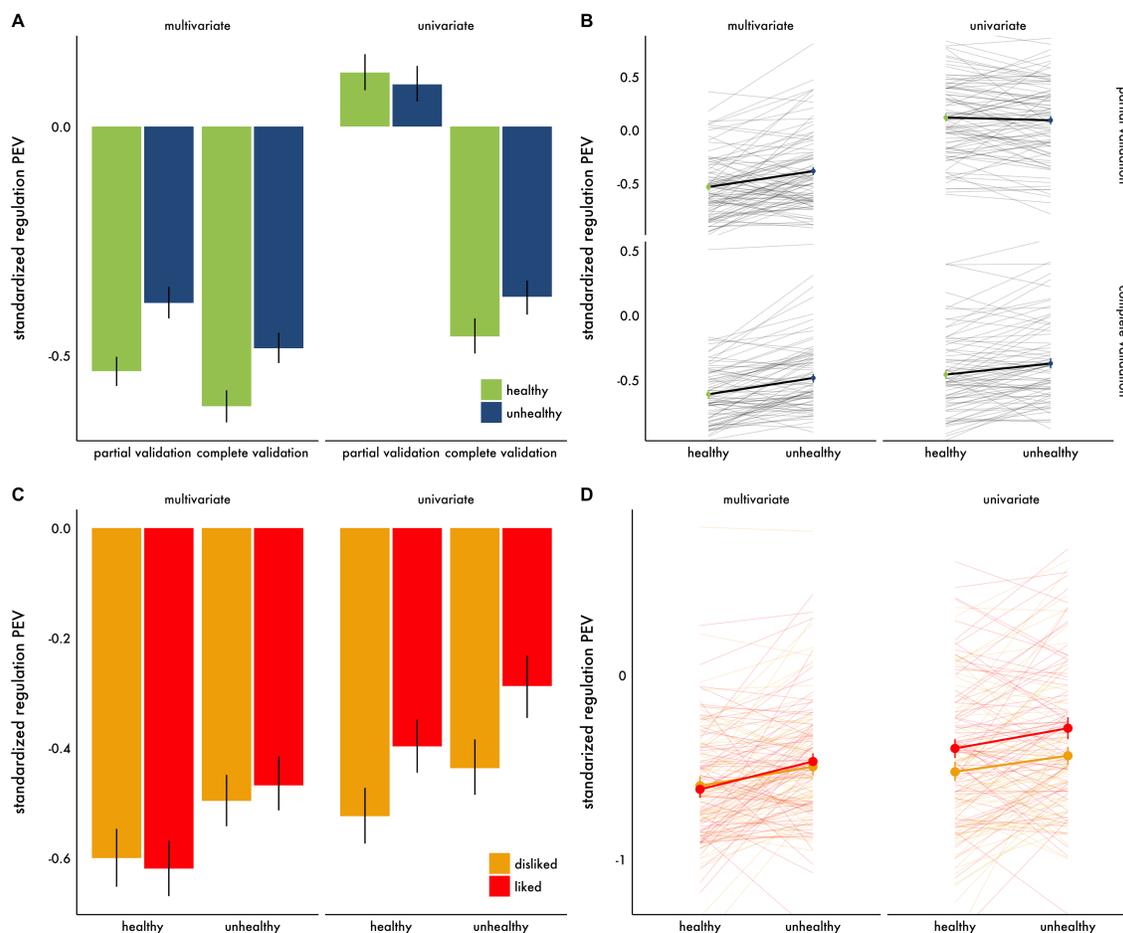
Fig. 8. Mean standardized regulation pattern expression values as a function of food type (healthy or unhealthy), signature type (multivariate classifier or univariate contrast), and sample (partial validation or complete validation sample). Top Panel A shows group-level means, whereas panel B shows group-level means overlaid on individual participant means; each thin line represents a participant. Panel C and D visualize this relationship within the complete validation sample only as a function of pre-session palatability ratings (relatively disliked = ratings 1-2, liked = ratings 3-4). Higher positive values represent relatively higher evidence for regulation, whereas lower negative values represent relatively higher evidence for viewing; zero is the decision boundary between conditions. Error bars are 95% confidence intervals across trials. Pattern expression values are standardized within participant and signature type. PEV = pattern expression value.

Table 7

Repeated Measures Correlations in the Food Valuation Task Between Neural Signatures as a Function of Sample

Sample	df	r [95% CI]
partial validation	5006	0.51 [0.49, 0.53]
complete validation	4336	0.54 [0.52, 0.56]

Note. 95% confidence intervals are bracketed. Repeated measures correlations include all trial-level data and adjust for trials nested within participant using multilevel modeling.

more reliable across samples and seemed to index craving regulation in the expected ways. Within the complete validation sample, the univariate signature differentiated food type, palatability, and scaled with bid value, but in the opposite direction–greater expression was associated with *higher* not lower bids. One possibility is that this signature may have indexed engagement in unsuccessful craving regulation. Combined with the specification curve analysis showing that the signatures account for unique variance in bids, results suggest that they might best be used together to improve prediction. Additional research is needed to disentangle precisely what

information is being represented by each signature. More fine-grained analyses with a similar task design to that used in the complete validation sample; future studies might use this trial-by-trial approach to test neurocognitive theories of self-control within trials that require self-control for dieters (e.g., specifically for healthy disliked and unhealthy liked foods). These efforts would be bolstered by using strongly disliked (rather than relatively disliked) foods and explicitly asking about task goals (e.g., avoiding unhealthy foods, choosing healthy foods, avoiding all foods) to better control for individual differences in goal conflicts.

Aim 3: Individual differences within each task.

The third aim of this study was to assess the degree to which the craving regulation signatures differentiated individuals. We calculated mean regulation pattern expression and craving rating/bid value for each participant and task condition (Figure 11 and Figure 12) and difference scores between conditions. Correlations between mean regulation pattern expression and craving ratings (Figure S4) or bid val-
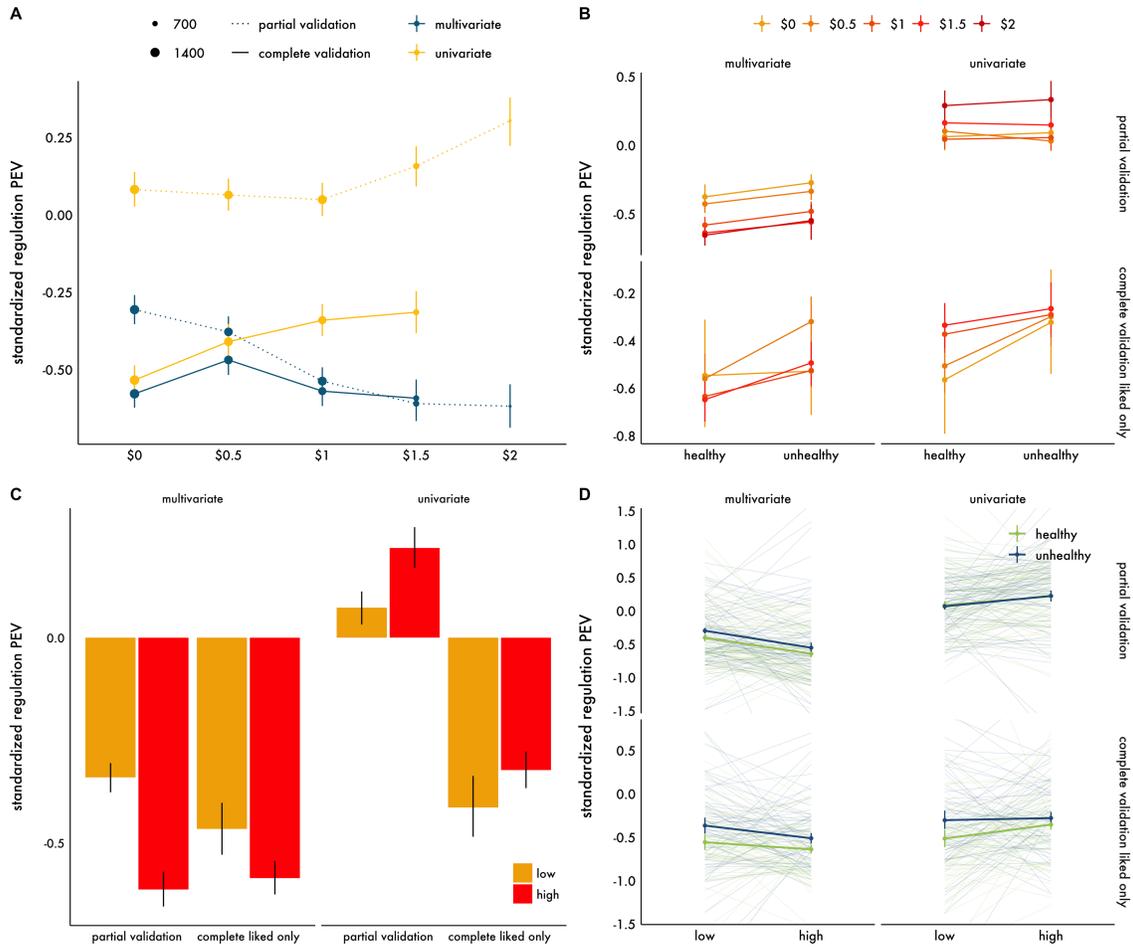
Fig. 9. The relationship between bid values and mean standardized regulation pattern expression values as a function of signature type (multivariate classifier or univariate contrast) and sample (partial validation or complete validation sample). Top panel: Continuous bid values A) collapsed across food type (healthy or unhealthy), scaled by the number of observations in each rating category, represented by the size of the point, and B) as a function of food type. Bottom panel: Mean dichotomized bid (low = lower than scale midpoint, high = higher than scale midpoint) C) collapsed across food type, and D) as a function of food type. The complete validation sample in A includes both liked and relatively disliked foods, whereas in panels B-D, it includes liked foods only (i.e., pre-session palatability ratings > 2 on a 1-4 scale). In D, group means are overlaid on individual participant means; each thin line represents a single participant. Error bars are 95% confidence intervals across all trials. The partial validation sample used a $ 0-$ 2 bid value scale, whereas the complete validation sample used a $ 0-$ 1.5 bid value scale. Pattern expression values are standardized within participant and signature type. PEV = pattern expression value.
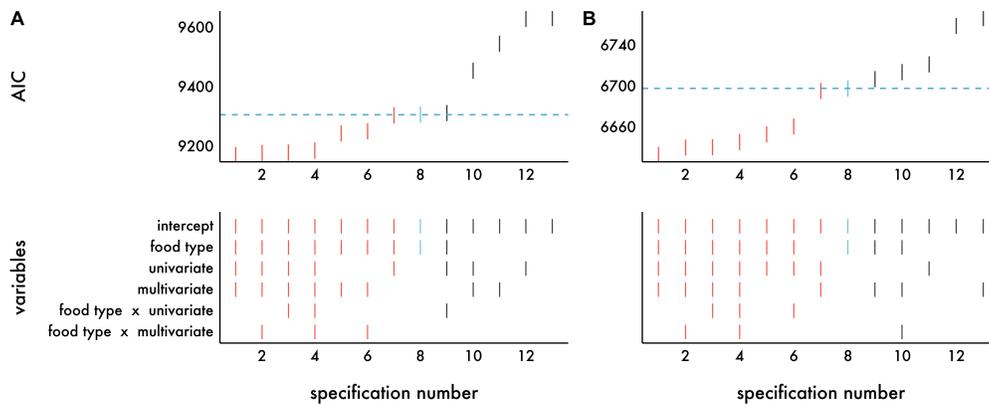
Fig. 10. Specification curves of 13 unique models regressing trial-level bid values on predictors ordered based on model fit (AIC) for A) the partial validation sample, and B) the complete validation sample. Each column corresponds to a single model specification. The AIC value for each model specification is plotted in top panels and the variables included in each model are visualized in the bottom panels. Because each panel is ordered based on AIC, specification numbers do not necessarily correspond to the same model specifications in each panel. The base model, which included food type (healthy or unhealthy) as the only predictor, is highlighted in blue and the dotted blue line represents the AIC for this model. Models with AIC values lower than the base model are highlighted in red. Potential variables in each model included: intercept, food type (healthy or unhealthy), standardized pattern expression values for the univariate contrast and the multivariate classifier, and the interaction between food type and each signature type.
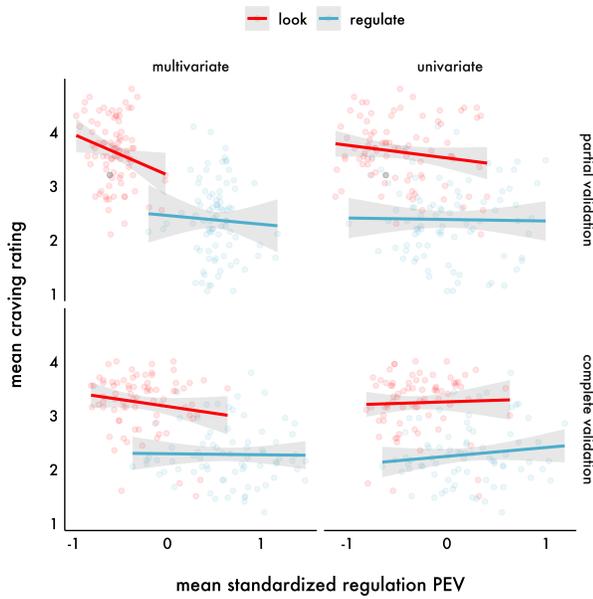
Fig. 11. The relationship between participant mean standardized regulation pattern expression values and mean craving ratings during the craving regulation task as a function of instruction (look or regulate), signature type (multivariate classifier or univariate contrast) and sample (partial validation or complete validation sample). Condition outliers are visualized as grey dots, but were excluded when computing linear effects. PEV = pattern expression value.

ues (Figure S5) for each condition separately were relatively weak and tended to be higher for difference scores. Individuals with relatively greater regulation pattern expression when regulating versus looking in the craving regulation task had larger decreases in craving ratings when regulating, but this effect was only stable for the multivariate signature ($r_{partial}$ = .24, 95% CI [0.03, 0.43], $r_{complete}$ = .33, 95% CI [0.11, 0.52]). This pattern was also observed in the food valuation task for the multivariate signature: relatively greater regulation pattern expression for unhealthy versus healthy foods was associated with higher bids on healthy foods relative to unhealthy foods ($r_{partial}$ = .46, 95% CI [0.28, 0.61]; $r_{complete}$ = .24, 95% CI [0.01, 0.44]; $r_{complete–liked}$ = .28, 95% CI [0.06, 0.48]). These effects were unrelated for the univariate signature.

Discussion. The data presented in this section indicate there is individual variability in craving regulation signature expression and that it relates to individual differences in regulation success and relative value of healthy over unhealthy foods. These moderate correlations were most reliable for the multivariate signature and were stronger for difference measures than condition measures, perhaps reflecting individual differences in flexible behavioral modulation between conditions. Although we explored the relationship between signature expression and individual differences in task behavior (and age, sex, and BMI in Supplementary material), future research should extend these findings to assess potential relationships with individual factors such as dieting goals, motivation, and self-control.
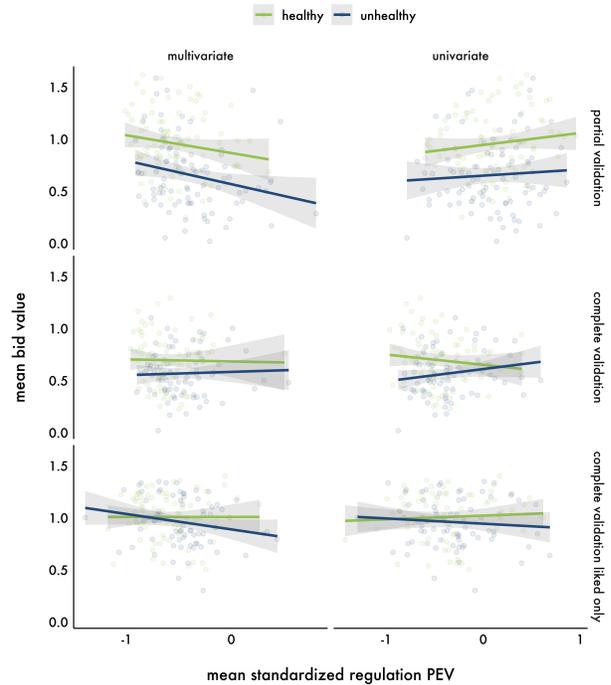

Fig. 12. The relationship between participant mean standardized regulation pattern expression values and mean bids values during the food valuation task as a function of food type (healthy or unhealthy), signature type (multivariate classifier or univariate contrast) and sample (partial validation or complete validation sample). The bottom panel shows correlations within the complete validation sample for liked snack foods items only (i.e., they were rated as 3 or 4 during the pre-session palatability rating task on a 1-4 scale). No liking ratings were collected for the partial validation sample. PEV = pattern expression value.

Conclusions, limitations, future directions

This study establishes the construct validity for the multivariate and univariate signatures as measures of food craving regulation, demonstrates their generalizability to out-of-sample individuals, and identifies conditions when they provide divergent information. The neural signatures tracked with food valuation in an untrained bidding task, suggesting that they might also index spontaneous regulation. The multivariate signature was overall the most reliable across both tasks and thus has the greatest promise for further development. Interestingly, this signature did not scale with ratings during explicit regulation. This result may be an artifact of the modeling approach or it might indicate that variation within cognitive reappraisal may not be reflected in the same neural pattern as its mere presence vs. absence. Future studies can compare signatures trained using classification and regression incorporating ratings to better understand this process.

The application of machine learning tools to multivariate patterns of brain activation can add reliability and predictive validity beyond univariate contrasts. The relatively lower accuracy of neural signatures developed using average condition effects for each suggests that trial-level training may increase accuracy by incorporating greater variability during training. Variability might also be increased in the sample itself. We used data from four studies collected on two scanners, from adolescents and adults with varying body mass

indices. The variance in the development sample and therefore the generalizability of the results could have been increased even further by accounting for demographics (e.g., race/ethnicity) and other individual factors.

We focused on neural signatures of craving regulation, but our general approach has clear implications for health neuroscience more broadly. First and foremost, using whole-brain multivariate signatures may improve the ability to predict health-relevant outcomes using functional neuroimaging data. We did not assess incremental validity, for example compared to ROIs, but others have shown that they outperform ROIs (Chang et al., 2015; Doré et al., 2017). Future research might identify cases where multivariate representations and ROIs can collectively predict outcomes to a greater degree than either on its own. The ability to model individual differences in spontaneous craving regulation should allow future studies to test whether deficits in this capacity increases risk for future weight gain and poorer response to obesity prevention and treatment interventions. A similar approach might be used to assess individual differences in other appetitive disorders, such as modulation of cravings for alcohol or drugs, and test of whether deficits in spontaneous craving for such substances predict outcomes for these health problems.

Second, there are cases when multivariate neural signatures have greater sensitivity than univariate signatures to detect within and between person differences as a function of time, health status, or intervention. Greater levels of sensitivity can be achieved by training classifiers on increasingly specific process-level distinctions. As such, multivariate signatures have the potential to deliver stronger and more precise tests of mechanistic models of change. Multivariate approaches can benefit health neuroscience by enabling researchers to test theoretical predictions about mental processes that are believed to be targeted by interventions. That is, the ability to model individual differences in craving regulation signatures and change in these signatures may be a useful measure of target engagement for craving modulation interventions (e.g, cognitive reappraisal training for high-calorie foods, alcohol, or drugs). This is a particular advantage over univariate approaches, which are not typically optimized to allow inferences about mental processes from neural data (Poldrack, 2010).

Third, the evidence presented here regarding spontaneous regulation enables researchers to assess evidence for engagement in target psychological processes without their explicit elicitation (Doré et al., 2019). The ability to measure a process in this way affords the use of more naturalistic paradigms in the scanner. Several neural signatures can be deployed simultaneously to index multiple processes that are each expected to be engaged. For example, assessing expression of indices of craving, craving regulation, and self-relevance together might capture more variance in health-relevant choices or receptivity to health messages than any single process alone. This combination approach might enable elucidation of the relative contributions of various psychological constructs in behavior.

It is important to emphasize that all of these implications require that the multivariate neural signatures are valid, sensitive, and specific indicators of relevant psychological constructs. Rigorously establishing validity across heterogeneous, representative samples is a data-intensive process that can be greatly facilitated by open practices and data sharing (Woo et al., 2017). Convergent and divergent validity can be tested across labs by making neural signatures publicly available online. Through such collaborative efforts, we can improve our ability to predict outcomes, rigorously test psychological theory, and further our understanding of the relationships between the brain and health behavior.

Summary. Researchers in social, cognitive, and affective neuroscience have increasingly combined multivariate and machine learning approaches to generate new insights into a range of psychological phenomena. We argue that health neuroscience, too, can benefit from these innovative approaches. Neural signatures are complementary to existing tools used in health neuroscience, adding predictive utility to our models and uncovering new information in existing data. Health neuroscience will benefit to an even greater extent in the future as researchers develop additional uses of multivariate machine learning and find new ways to apply them to health-relevant data.

References

Berkman, E. T., & Falk, E. B. (2013). Beyond Brain Mapping: Using Neural Measures to Predict Real-World Outcomes. *Current Directions in Psychological Science*, *22*(1), 45–50. https://doi.org/10.1177/0963721412469394

Boswell, R. G., & Kober, H. (2016). Food cue reactivity and craving predict eating and weight gain: A meta-analytic review: Food cue reactivity and craving meta-analysis. *Obesity Reviews*, *17*(2), 159–177. https://doi.org/10.1111/obr.12354

Boswell, R. G., Sun, W., Suzuki, S., & Kober, H. (2018). Training in cognitive strategies reduces eating and improves food choice. *Proceedings of the National Academy of Sciences*, *115*(48), E11238–E11247. https://doi.org/10.1073/pnas.1717092115

Buhle, J. T., Silvers, J. A., Wager, T. D., Lopez, R., Onyemekwu, C., Kober, H., . . . Ochsner, K. N. (2014). Cognitive Reappraisal of Emotion: A Meta-Analysis of Human Neuroimaging Studies. *Cerebral Cortex (New York, NY)*, *24*(11), 2981–2990. https://doi.org/10.1093/cercor/bht154

Burger, K. S., & Stice, E. (2012). Frequent ice cream consumption is associated with reduced striatal response to receipt of an ice cream–based milkshake. *The American Journal of Clinical Nutrition*, *95*(4), 810–817. https://doi.org/10.3945/ajcn.111.027003

Bzdok, D., & Ioannidis, J. P. A. (2019). Exploration, Inference, and Prediction in Neuroscience and Biomedicine. *Trends in Neurosciences*, *42*(4), 251–262. https://doi.org/10.1016/j.tins.2019.02.001

Chang, L. J., Gianaros, P. J., Manuck, S. B., Krishnan, A., & Wager, T. D. (2015). A Sensitive and Specific Neural Signature for Picture-Induced Negative Affect. *PLOS Biol*, *13*(6), e1002180. https://doi.org/10.1371/journal.pbio.1002180

Corbin, N., Todd, N., Friston, K. J., & Callaghan, M. F. (2018). Accurate modeling of temporal correlations in rapidly sampled fMRI time series. *Human Brain Mapping*, *39*(10), 3884–3897. https://doi.org/10.1002/hbm.24218

Cosme, D., Flournoy, J. C., & Vijayakumar, N. (2018). auto-motion-fmriprep: A tool for automated assessment of motion artifacts. https://doi.org/10.5281/zenodo.1412131

Cosme D., Ludwig R. M., & Berkman E. T. (in press). Comparing two neurocognitive models of self-control during dietary decisions. *Social Cognitive and Affective Neuroscience*.

Cumming, G. (2014). The New Statistics: Why and How. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Demos, K. E., Heatherton, T. F., & Kelley, W. M. (2012). Individual Differences in Nucleus Accumbens Activity to Food and Sexual Images Predict Weight Gain and Sexual Behavior. *Journal of Neuroscience*, *32*(16), 5549–5552. https://doi.org/10.1523/JNEUROSCI.5958-11.2012

Doré, B. P., Tompson, S. H., O'Donnell, M. B., An, L. C., Strecher, V., & Falk, E. B. (2019). Neural Mechanisms of Emotion Regulation Moderate the Predictive Value of Affective and Value-Related Brain Responses to Persuasive Messages. *Journal of Neuroscience*, *39*(7), 1293–1300. https://doi.org/10.1523/JNEUROSCI.1651-18.2018

Doré, B. P., Weber, J., & Ochsner, K. N. (2017). Neural Predictors of Decisions to Cognitively Control Emotion. *The Journal of Neuroscience*, *37*(10), 2580–2588. https://doi.org/10.1523/JNEUROSCI.2526-16.2016

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., . . . Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111. https://doi.org/10.1038/s41592-018-0235-4

Falk, E. B., Berkman, E. T., Whalen, D., & Lieberman, M. D. (2011). Neural activity during health messaging predicts reductions in smoking above and beyond self-report. *Health Psychology*, *30*(2), 177. https://doi.org/10.1037/a0022259

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Giuliani, N. R., Calcott, R. D., & Berkman, E. T. (2013). Piece of cake. Cognitive reappraisal of food craving. *Appetite*, *64*, 56–61. https://doi.org/10.1016/j.appet.2012.12.020

Giuliani, N. R., Mann, T., Tomiyama, A. J., & Berkman, E. T. (2014). Neural Systems Underlying the Reappraisal of Personally Craved Foods. *Journal of Cognitive Neuroscience*, *26*(7), 1390–1402. https://doi.org/10.1162/jocn_a_00563

Giuliani, N. R., & Pfeifer, J. H. (2015). Age-related changes in reappraisal of appetitive cravings during adolescence. *NeuroImage*, *108*, 173–181. https://doi.org/10.1016/j.neuroimage.2014.12.037

Giuliani, N. R., Tomiyama, A. J., Mann, T., & Berkman, E. T. (2015). Prediction of daily food intake as a function of measurement modality and restriction status. *Psychosomatic Medicine*, *77*(5), 583–590. https://doi.org/10.1097/PSY.0000000000000187

Gross, J. J. (1998). Antecedent-and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology*, *74*(1), 224.

Hall, P. A., Bickel, W. K., Erickson, K. I., & Wagner, D. D. (2018). Neuroimaging, neuromodulation, and population health: The neuroscience of chronic disease prevention. *Annals of the New York Academy of Sciences*, *1428*(1), 240–256. https://doi.org/10.1111/nyas.13868

Hutcherson, C. A., Plassmann, H., Gross, J. J., & Rangel, A. (2012). Cognitive Regulation during Decision Making Shifts Behavioral Control between Ventromedial and Dorsolateral Prefrontal Value Systems. *The Journal of Neuroscience*, *32*(39), 13543–13554. https://doi.org/10.1523/JNEUROSCI.6387-11.2012

Kober, H., Mende-Siedlecki, P., Kross, E. F., Weber, J., Mischel, W., Hart, C. L., & Ochsner, K. N. (2010). Prefrontal–striatal pathway underlies cognitive craving regulation. *Proceedings of the National Academy of Sciences*, *107*(33), 14811–14816. https://doi.org/10.1073/pnas.1007779107

Lopez, R. B., Hofmann, W., Wagner, D. D., Kelley, W. M., & Heatherton, T. F. (2014). Neural predictors of giving in to temptation in daily life. *Psychological Science*, *25*(7), 1337–1344.

Poldrack, R. A. (2010). Mapping mental function to brain structure: How can cognitive neuroimaging succeed? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *5*(5), 753–761. https://doi.org/10.1177/1745691610388777

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, *72*(5), 692–697. https://doi.org/10.1016/j.neuron.2011.11.001

Rapuano, K. M., Huckins, J. F., Sargent, J. D., Heatherton, T. F., & Kelley, W. M. (2016). Individual Differences in Reward and Somatosensory-Motor Brain Regions Correlate with Adiposity in Adolescents. *Cerebral Cortex*, *26*(6), 2602–2611. https://doi.org/10.1093/cercor/bhv097

Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage*, *23*(2), 752–763. https://doi.org/10.1016/j.neuroimage.2004.06.035

Shahane, A. D., Lopez, R. B., & Denny, B. T. (2018). Implicit reappraisal as an emotional buffer: Reappraisal-related neural activity moderates the relationship between inattention and perceived stress during exposure to negative stimuli. *Cognitive, Affective, & Behavioral Neuroscience*. https://doi.org/10.3758/s13415-018-00676-x

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Specification Curve: Descriptive and Inferential Statistics on All Reasonable Specifications. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2694998

Stice, E., Yokum, S., Burger, K., Rohde, P., Shaw, H., & Gau, J. M. (2015). A Pilot Randomized Trial of a Cognitive Reappraisal Obesity Prevention Program. *Physiology & Behavior*, *0*, 124–132. https://doi.org/10.1016/j.physbeh.2014.10.022

Wager, T. D., Atlas, L. Y., Lindquist, M. A., Roy, M., Woo, C.-W., & Kross, E. (2013). An fMRI-Based Neurologic Signature of Physical Pain. *New England Journal of Medicine*, *368*(15), 1388–1397. https://doi.org/10.1056/NEJMoa1204471

Westfall, J., Nichols, T., & Yarkoni, T. (2016). *Fixing the stimulus-as-fixed-effect fallacy in task fMRI* (No. biorxiv;077131v1). Retrieved from http://biorxiv.org/lookup/doi/10.1101/077131

Woo, C.-W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: Brain models in translational neuroimaging. *Nature Neuroscience*, *20*(3), 365–377. https://doi.org/10.1038/nn.4478

Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Yokum, S., & Stice, E. (2013). Cognitive regulation of food craving: Effects of three cognitive reappraisal strategies on neural response to palatable foods. *International Journal of Obesity*, *37*(12), 1565–1570. https://doi.org/10.1038/ijo.2013.39