Effects in amygdala

Given prior research linking amygdala activity and emotional responding, we also considered effects within the amygdala, defined by the Harvard-Oxford atlas (thresholded at 25% probability). We did not observe a clear main effect of trial type in the amygdala, but did see a cluster within right amygdala (32, 4, -26) showing a strategy by trial type interaction (SVC $p<.05$), such that activity was modulated for the emotion-focused strategy but not for the persuasion-focused strategy. We also did not see that activity within bilateral amygdala clearly predicted ratings of negative affect, perceived effectiveness, or self-relevance (69%, 90% and 83% of the posterior densities for these predictive relationships were above zero). Finally, we did not see that amygdala activity mediated cognitive regulation effects on negative affect, perceived effectiveness, or self-relevance (40%, 41% and 46% of the posterior densities for these mediation effects were above zero).

Effects in a region of vmPFC implicated in positive reappraisal

Because of the conceptual overlap being finding persuasive value in an arousing anti-binge drinking message and finding positive meaning in a negative situation more generally, we also considered effects within a vmPFC region implicated in positive reappraisal, a cognitive strategy that entails thinking more positively about negative situations (e.g., imagining that getting in a minor car accident while under the influence of alcohol could lead a person to reduce their drinking and ultimately be happier and healthier). Specifically, we used a cluster within vmPFC that was identified in a previous study as more responsive to positive reappraisal than to minimizing reappraisal (a 184 voxel cluster centered at [3, 21, −9]; Doré et al., 2016). We observed a main effect of

trial type (up- versus down-regulate) within this region, b=.02, 95%CI[.01, .03]. Further, activity in this region was predictive of in-scanner ratings of negative affect, b=.04, 95%CI[.01, .06], and perceived effectiveness, b=.06, 95%CI[.03, .09], as well as post-scan ratings of self-relevance, b=.05, 95%CI[.02, .09]. Further, there was also mediation of the effects of trial type (up- versus down-regulate) by activity within this vmPFC region for negative affect, indirect path = .0012, 95%CI[.0001, .0028], perceived effectiveness, indirect path = .0022, 95%CI[.0005, .0044] , and post-scan self-relevance, indirect path = .0009, 95%CI[.0001, .0024].

Mediation paths did not show clear differences by strategy type

We saw similar mediation effects for the emotion-focused strategy and the persuasion-focused strategy, with no interactions indicating clear differences between the strategy types. In one case, we found a marginal interaction suggesting weak evidence that expression of the negative emotion pattern could be more predictive of post-scan self-relevance when participants are applying the emotion-focused strategy versus the persuasion focused strategy, b=.08, 95%CI[-.01, .16] (i.e., 91% of the posterior density for the interaction coefficient was above zero).

Follow-up analyses assessing estimated out-of-sample model accuracies in predicting negative affect, perceived effectiveness, and message self-relevance

We conducted follow-up analyses in which we used Bayesian leave-one-out (LOO) cross-validation to estimate the expected out-of-sample accuracy of models predicting persuasion outcomes. To estimate the out-of-sample accuracy of our models in predicting outcomes, we ran Bayesian LOO cross-validation using Pareto-smoothed importance sampling (LOO; Vehtari et al., 2016). Instead of model re-fitting, as in exact

cross-validation, the LOO procedure draws samples from posterior distributions of the model parameters in order to estimate expected log-likelihood for new data and thus adjust for over-optimism (bias) inherent to within-sample measures of model fit. From this procedure, we derived LOO-adjusted deviance values (LOOIC) that can be used to compare models in terms of their expected out-of-sample predictive error on the model deviance scale (a lower number indicates higher expected out-of-sample accuracy). This is conceptually similar to comparing AIC (Akaike information criterion), DIC (deviance information criterion), or WAIC (widely applicable information criterion) scores, which approximate out-of-sample error under a more restrictive set of conditions (see Gelman et al., 2014).

First, we used this LOO procedure to compare the fit of the forward multilevel mediation models described in the main manuscript (cognitive regulation leads to change in brain activity which in turn leads to a change in self-report ratings) to models that reverse the order of the mediators and outcomes (cognitive regulation leads to a change in self-report ratings which in turn leads to a change in brain activity). A forward mediation model performed slightly better than a reverse mediation model for in-scanner ratings of negative affect ($LOOIC_{fwd\_neg}$ = 54112; $LOOIC_{rev\_neg}$ = 54143), for in-scanner ratings of ad effectiveness ($LOOIC_{fwd\_neg}$ = 54815; $LOOIC_{rev\_neg}$ = 54823), and for delayed ratings of ad self-relevance ($LOOIC_{fwd\_rel}$ = 28744; $LOOIC_{rev\_rel}$ = 28749). Importantly, mediation modelling (including comparisons of different models) cannot provide clear evidence for causality in the absence of experimental manipulations. However, it can usefully assess the compatibility of the data with the hypothesized

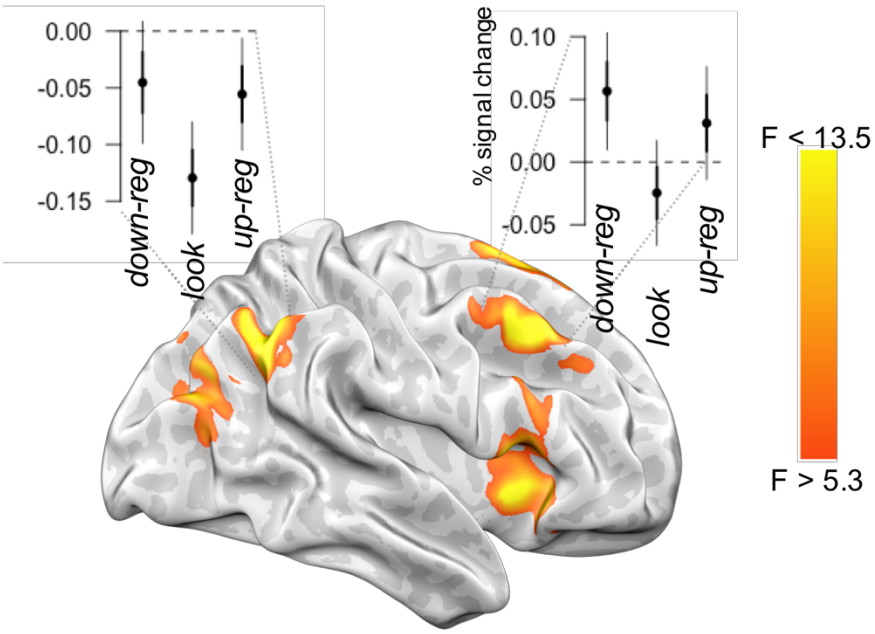
causal models and in this manner provide direction for future work applying experimental manipulations (e.g., direct physical manipulation of brain activity).

Next we compared the expected out-of-sample predictive accuracy of models predicting self-report ratings from groups of predictors that capture i) stimulus characteristics, ii) experimentally instructed cognitive goals, and iii) brain activity related to emotion and value. For immediate ratings of negative affect, a full model including negative emotion pattern expression, valuation pattern expression, vmPFC activity, regulatory goal (up- versus down-regulate), task (emotion regulation versus persuasion regulation), normative message negativity, and normative message persuasiveness, showed substantially better predictive fit ($LOOIC_{M3neg}$ = 11634) than a reduced model predicting negative affect from regulatory goal, task, and normative ratings but not brain activity ($LOOIC_{M2neg}$ = 11709), or a further reduced model predicting negative affect from only normative ratings ($LOOIC_{M1neg}$ = 12069). Similarly, analogous full models were preferred over analogous reduced models for immediate ratings of ad effectiveness ($LOOIC_{M3eff}$ = 12343; $LOOIC_{M2eff}$ = 12460; $LOOIC_{M1eff}$ = 12894) and for ratings of ad self-relevance made at a one-hour delay ($LOOIC_{M3rel}$ = 6715; $LOOIC_{M2rel}$ = 6726; $LOOIC_{M1rel}$ = 6733). Overall, consistent with our multilevel mediation analyses, this pattern of results indicates that predictive models including brain indices of affect- and valuation-related processes were generally higher in expected out-of-sample accuracy than reduced models including only experimentally manipulated regulatory goals and normative message characteristics (for in-scanner ratings of negative affect and persuasiveness, and re-exposure ratings of self-relevance).

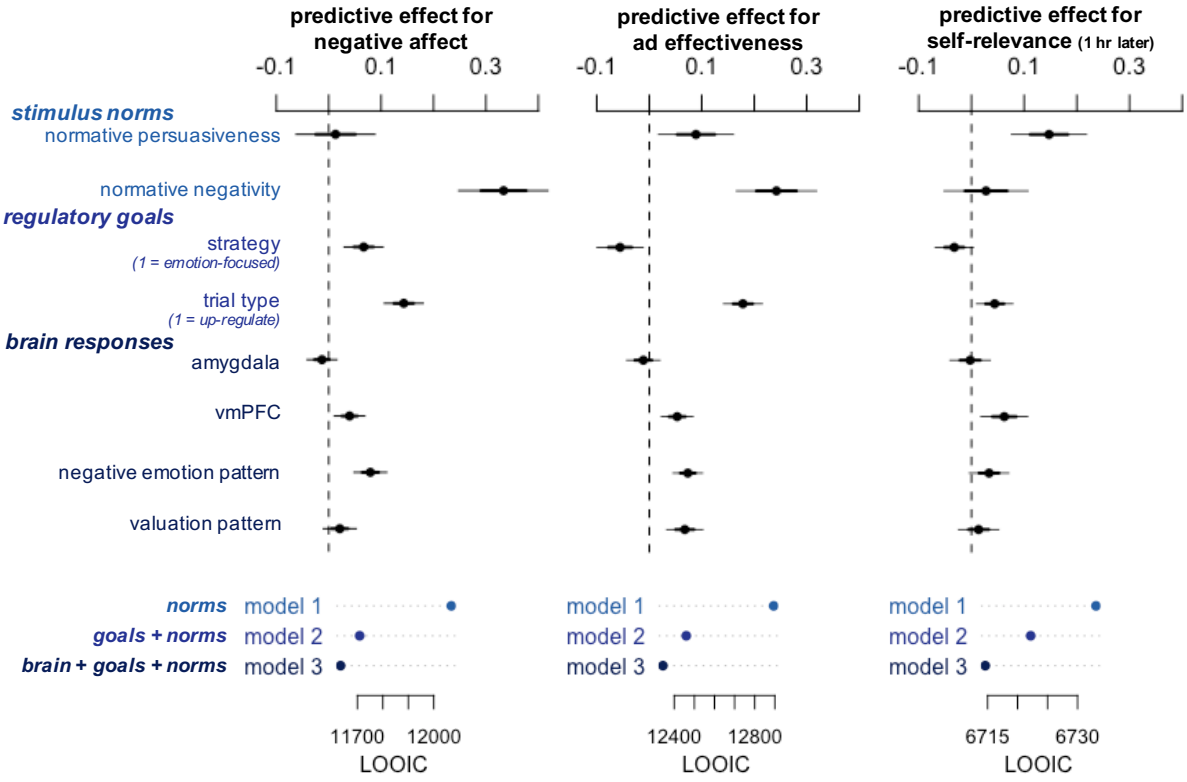
Supplementary References

Doré, B. P., Boccagno, C., Burr, D., Hubbard, A., Long, K., Weber, J., Stern, Y, & Ochsner, K. N. (2017). Finding positive meaning in negative experiences engages ventral striatal and ventromedial prefrontal regions associated with reward valuation. Journal of cognitive neuroscience, 29(2), 235-244.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. Statistics and computing, 24(6), 997-1016.



Supplementary Figure S1. Lateral PFC and posterior parietal clusters identified by a whole-brain search for regions showing an omnibus effect of trial type. These regions were engaged when participants were instructed to deliberately up- or down-regulate their responses to the ads.



Supplementary Figure S2. Brain variables enhanced accuracy in predicting subjective experience in response to anti-binge drinking ads beyond regulatory goal and stimulus variables. (Top) Coefficient plots for models predicting immediate ratings of negative affect, immediate ratings of ad effectiveness, and delayed ratings of ad self-relevance, from normative ratings of stimulus negativity and persuasiveness, experimentally instructed cognitive regulation goals, and brain responses associated with emotion and valuation. Estimates reflect posterior means (with 50% and 95% credibility intervals) from models with predictors entered simultaneously. **(Bottom)** Predictive fit of full models (model 3) including brain, regulatory goal, and stimulus variables as well as reduced models including regulatory goal and stimulus variables (model 2) or only stimulus variables (model 1) is summarized by LOOIC (model deviance adjusted by leave-one-out cross-validation; a lower number indicates better fit).

Training module script

Welcome to the training module for the task that you will complete in the scanner today. Please listen carefully and let us know if you have any questions.
In our daily lives, we sometimes let our thoughts and feelings come and go naturally. At other times, however, we actively try to change the way we think and feel about the things we encounter.
Before each ad you are going to see an instruction that tells you what to do while the ad is on the screen.
—
The LOOK instruction
One instruction you will see is the LOOK instruction. When you see the instruction to LOOK, we want you to look naturally at the ad, and have whatever thoughts and feelings you would normally have.
—
The DECREASE NEGATIVE instruction
Another kind of instruction you will see is the DECREASE NEGATIVE instruction. When you see the DECREASE NEGATIVE instruction, we want you to try to make yourself feel less negative (bad) about the ad by thinking about it in ways that make your response to it less negative and more unemotional. One way you could do this is by imagining that the picture is fake or not as bad as it initially appears.
For example, if the ad depicts someone who is injured or people who are fighting, you could imagine that the scene has been staged or edited. For another example, if the ad provides information about the risks of binge-drinking, you could think that those risks are not likely or are not as bad as they appear. The key is to change the way you think about the ad so you feel as unemotional as you can about it.
Does that make sense?
—
The INCREASE NEGATIVE instruction
Another kind of instruction you will see is the INCREASE NEGATIVE instruction. When you see the INCREASE NEGATIVE instruction, we want you to make yourself feel more negative (bad) about the ad by thinking about it in ways that make your response to it more negative and more emotional. One way you could do this is by imagining that the picture reflects a real-life situation that is just as bad as it appears or worse.
For example, if the ad depicts someone who is injured or people who are fighting, you could think about the serious consequences of such an injury or fight. For another example, if the ad provides information about the risks of binge-drinking, you could think about how bad it would be if you or someone close to you had to experience those risks. The key is to change the way you think about the ad so you feel as negative as you can about it. Does that make sense?
—
The WHY PERSUASIVE? instruction
Another kind of instruction you will see is the WHY PERSUASIVE? instruction. When you see the WHY PERSUASIVE? instruction, we want you to think about a reason why this ad is persuasive (i.e., effective in getting a point across). That is, we want you to identify a strength of the ad.
For example, you could focus on something convincing about the argument that is presented, or a reason why the visual imagery is particularly effective. Does that make sense?
—
The WHY NOT PERSUASIVE? instruction
The final kind of instruction you will see is the WHY NOT PERSUASIVE? instruction. When you see the WHY NOT PERSUASIVE? instruction, we want you to think about a reason why this ad

is NOT persuasive (i.e., why it is ineffective in getting a point across). That is, we want you to identify a weakness of the ad.
For example, you could focus on something unconvincing about the argument that is presented, or a reason why the visual imagery is not effective.
Does that make sense?
—
In your own words, what are you supposed to do when you see:
'LOOK'
'INCREASE NEGATIVE'
'DECREASE NEGATIVE '
'WHY PERSUASIVE?'
'WHY NOT PERSUASIVE?'
—
Rating NEGATIVE FEELINGS and AD EFFECTIVENESS
After each ad, you will be asked to rate your current negative feelings, and the effectiveness of the ad.
You will rate your negative feelings on a 1-2-3-4-5 scale from 1 (not at all negative) to 5 (very negative).
You will also rate the effectiveness of the ad on a 1-2-3-4-5 scale from 1 (not at all effective) to 5 (very effective).
You will only have a few seconds to make these ratings, so make sure to keep your fingers on the button pad at all times. Questions?'
—
Great. Now we will have you do a few practice trials of the task. The task will advance automatically. Follow the instructions while paying attention to the image, and then make ratings using the 1,2,3,4,5 number keys.
[practice trials]
—
You have completed the training!
The task in the scanner will be the same as what you have just practiced. However, there are a few more things we need to tell you about the scanner environment. First, when the scanner is running it can be very loud. We will give you earplugs to dampen the noise and protect your hearing. Second, when you are in the scanner, it is critical that you do your best to move your head as LITTLE as possible. Even very slight movements can disrupt the measurements we need for this study. Try your best to not move any part of your body, as moving other parts of your body can also move your head. Finally, please let us know if you feel uncomfortable at any time. You will be in the scanner for about 50 minutes. We will check in every 10 minutes or so over the intercom. Any questions?'