SUMMARY:   
In this study, the authors explore the ability of goals or attentional focus to modulate choices in a food task and in an altruism task. In the tasks, subjects chose or rejected different foods or different monetary offers. Instructions to choose healthy or to make generous choices modulated the behavior in both tasks and also affected neural activity patterns measured using fMRI. Regulatory success - or the ability to use instructions to change default choice behavior - was correlated with DLPFC activity.   
  
ESSENTIAL REVISIONS:   
**Editor Comment 1. “One reviewer, not someone doing fMRI work, had significant difficulty simply following how the concepts of value versus regulation were operationalized in the task and analysis. What precisely were the conditions compared to isolate regulation for example from value? Describing this in more concrete terms would be helpful.”**

**Response E1:** We thank the editor for the opportunity to resubmit our paper, and for the thoughtful comments and suggestions. We agree with Reviewer 1 that we were not as clear as we could have been in the previous instantiation of the manuscript about the focus of our paper, and on the distinctions between value and regulation. We have substantially revised the introduction (as described in greater detail below, see Response 1.1) to make this clearer. In brief, we have more clearly specified that we are focused on disentangling different hypotheses about *where* regulation acts to modify value computations, have explicitly spelled out different possibilities along with their predicted consequences, and have taken greater care to orient the reader to how different analyses address different questions.

**Editor Comment 2: “This was related to the overarching concern of the other reviewers as well,, which was the distinction between the authors conceptualization versus a more trivial interpretation - namely that when participants pay attention to a particular type of information (by instruction), this information is easier to decode from their brain activity. As noted below: "This effect should be seen as the outcome (and not the process) of cognitive regulation. How cognitive regulation is implemented, and how better decoding translates into biased choices, still need to be explained. " Basically what is needed is to provide a more mechanistic account tor what is going on.”**

**Response E2:** We agree that in a simple world it should be “trivial” to observe changes in information decoded from brain activity that matches changes at the behavioral level. Nevertheless, as we make clear in the revised introduction and discussion sections, extant research has actually NOT consistently found such patterns despite clear changes in behavior, raising an important puzzle to be explained. Our results are the first study to identify a clear representation of ALL choice-relevant attributes across multiple contexts, and to show clear modulation of ALL choice-relevant attributes in a manner consistent with regulatory goals. Respectfully, we believe that observing changes in attribute encoding is not trivial or uninformative as to the mechanism of action. As we make clearer in the new version of the paper in the Introduction, Results, and Discussion, identifying *where* attribute encoding changes occur may allow us to distinguish between different theories about computational mechanism. Specifically, we make clearer that regulation could occur via two mechanisms: modulation at the level of specific attribute representations in non-overlapping, attribute-specific regions, or at the level of weighting in the value integration process, and that how regulation operates could depend on where and how different attributes are computed. We highlight how these different mechanisms of action predict non-trivial differences in patterns of regulatory success. And we now make clearer in the Results and Discussion how our results help to distinguish these mechanisms, suggesting that a hybrid account in which some regulation occurs only at the level of value-integration mechanisms, while other regulation occurs at the level of specific attribute representations. For this reason alone, we believe our results shed more light on mechanistic questions of theoretical importance than perhaps was obvious given our previous results. We hope the reviewers and editor will agree!

**Editor Comment 3: “Suggestions are given in reviews below. For example " it would be important to show that the decoded value (i.e., the decoder output) correlates with the behavioral weights. The alternative would be that changes in decoding accuracy correspond to changes in precision (i.e., signal-to-noise ratio) and not changes in the signal itself. If correct, this would mean that a region downstream to the DLPFC could just read this value, add it to other values corresponding to other attributes, and feed the aggregate value to a selection process that makes the decision. Perhaps functional connectivity could be used to test for such a transfer of information. Thus, the neural model would parallel the behavioral model. " Another suggestion was that some sort of functional connectivity analysis could clarify \*how\* DLPFC (and potentially rTPJ) enables successful regulation mechanistically, possibly using additional ROI analyses on VMPFC to ensure this is indeed the case.**

**Response E3:** Although we believe that functional connectivity analyses are somewhat complicated (see our more detailed response to Reviewer #3), we have also taken to heart the reviewers’ suggestions to dig deeper into our data. Thus, we have now conducted exploratory functional connectivity analyses (see Response R3.xxx for details) to test whether and how changes in patterns of functional connectivity between regions predicts regulatory success. As can be seen, we find hints that connectivity patterns between VMPFC and both the DLPFC and Precuneus correlate in a theoretically consistent manner with changes in behavioral weighting of these attributes. Given that representations of these attributes do NOT change within the VMPFC, we believe this suggests that VMPFC may actually be modulating computations within DLPFC and Precuneus to amplify attribute values in these areas. Although this is not the modal view of the function of the VMPFC, it is actually consistent with some early work suggesting that this area may be involved in modulating affective representations in lower-level areas such as the amygdala in a goal-consistent way (e.g. Etkin et al, 2006). We now highlight this literature in the introduction, and include a discussion of functional connectivity results in the Supplementary Materials. However, given the exploratory nature of these results, and the relatively low statistical thresholds used to identify them, we have opted to mention them only briefly in the main paper. We can of course include a more detailed discussion of them in the paper if the editor or reviewers deem it appropriate.

These are the essentially revisions or problems identified in our discussion - clarify the approach and provide a more mechanistic account of the proposed interaction to rule out trivial explanations of the findings.   
  
  
MINOR POINTS:   
See reviews.....   
  
-----   
  
Reviewer #1 (General assessment and major comments (Required)):   
  
**Reviewer 1.1: “I think the general question is of interest and the authors approach is quite novel to me. They are basically testing how changes in goals affect behavior. This is similar to the use of devaluation in animal learning theory tasks, but here they are using simple instructions in the course of training with humans. I think this is a creative way to integrate economic decision making, which focuses on a unitary utility as guiding choices, with work from experimental psychology and computational neuroscience, which typically distinguishes different sources of information by its associative basis or computational basis. This is excellent. “**

**Response 1.1:** We thank the reviewer for their positive comments and enthusiasm for the approach!

**Reviewer 1.2. “However, beyond that I had difficulty following the authors framing, predictions and understanding the outcomes of the neural activity analysis. The DLPFC is the key "regulatory" area. But what does that mean? It seems to me that I would expect some areas to represent value independent of goals and some areas to represent that value only when goal relevant. Is this what is meant by regulatory? Does the DLPFC represents value relevant to goals and some other area does not - VMPFC?”**  
**Response 1.2:** The reviewer has the basic theory exactly correct. We predicted that there might be some areas that represent attributes independently of goals, and other areas that might represent them in a goal-directed manner (i.e., they would represent these attribute values more strongly when participants were motivated to *use* those attributes, and less strongly otherwise). A key question then is 1) whether regions like the VMPFC and/or DLPFC represent attribute values regardless of the choice domain (i.e. social or non-social choice) and regardless of the specific attribute (i.e. Self, Other, Taste, Health, etc.); and 2) whether and where attribute value representations are modified by goal. We hypothesized two possibilities. The *attribute-level hypothesis* suggests that regulatory goals operate by altering computations in specific attribute areas (such that there might be specific regions that represent Taste and Health, or Self and Other, and that each of these specific regions would be independently modulated by goal). This view suggests that we should see alterations in *both* attribute specific areas, as well as downstream in more domain-general value integration areas like VMPFC and DLPFC. On the other hand, the *integration-level hypothesis* suggests that regulatory goals could operate not by altering specific attribute representations, but by altering weighting of those attributes in an area (or areas) that represents multiple attributes in a more domain general way. Thus, while attribute-specific areas represent attributes regardless of regulation, value integration regions might shift to flexibly represent only those attribute relevant to the current goal.

To address the points raised by the reviewer, and to clarify both our goals as well as the distinction between regulation and value, we made a number of substantial changes to the manuscript. First, we now try to make clearer that in this paper our intention is to address how *value representations* change as a function of regulation, not to identify specifically the mechanisms that *implement* this regulation. We spell out different hypotheses about how this value modulation might take place in considerably more detail in the introduction (see pgs. XX-XX), and highlight several important questions about this process. Second, we now highlight in an overview section at the beginning of the Results section how we aim to test these specific hypotheses, with specific predictions about patterns of results that should be observed in different cases. We also now separate out a discussion of regulatory effects on VMPFC and DLPFC to make it clearer that prior literature is actually ambiguous about whether the VMPFC and DLPFC are *regulators* or are instead *regulated by* CRDM (pgs. XXX). We then clarify in the Discussion section that in this context, we view the DLPFC as being the key to successful domain-general regulation because it is *regulated by* CRDM to flexibly represent different attributes (pg. XXX), and that our results are actually less consistent with the idea that regulation happens at the level of a single value integration mechanism than that it happens within both domain-specific and domain-general attribute representation areas. We also now discuss in more detail the surprising finding of VMPFC consistency, and discuss its implications more thoroughly in light of additional analyses we performed at the suggestion of the reviewers (see our responses to the other Reviewer comments below). We hope that these changes help to clarify the framing, predictions and results, as well as how we view the roles of both the VMPFC and DLPFC.

**Reviewer 1.3: “Generally I think the question is very interesting and the approach is attractive, but I simply could not follow how the authors framed and then conducted their analysis. I will be interested to see what the other reviewers say. I might grasp things better if it were more clear how the theoretical concepts were operationalized for the analyses - precisely what is "regulatory" for example and how is it distinguished from non-regulatory versus just not involved by the task and then in the data analysis.”**

**Response 1.3**.Again, we thank the reviewer for their positive assessment of the importance and interest value of this question. Please see Response 1.2 for a more detailed discussion of the ways in which we have changed the framing, as well as discussion of the hypotheses and results, to make it clearer what we predicted and what we found, as well as how this relates questions of regulatory vs. non-regulatory regions.  
  
Reviewer #1 (Minor Comments):   
  
None   
  
Reviewer #1 (Additional data files and statistical comments):   
  
No specific criticisms, but difficult to follow analysis generally.   
  
  
**Reviewer #2 (General assessment and major comments (Required)):**  
  
**Reviewer 2.1**. **“Tusche and Hutcherson present a thought-provoking and methodologically impressive study on cognitive regulation of dietary and altruistic choices, a topic of broad interest. The analyses, which combine a drift diffusion computational model of attribute-weighted choice and trial-by-trial MVPA decoding, are sophisticated, appropriate, and comprehensive. They show that while attribute values across choice and regulatory goal contexts can be decoded in VMPFC, they do not appear to be modulated by regulatory goal. Conversely, attribute values in DLPFC are modulated consistently with model-derived behavioral weights across regulatory goals that emphasize either healthiness or tastiness during dietary choice, and personal gain in altruistic choice contexts. Regulatory goals that emphasize another's benefits (e.g. feelings), however, could instead be decoded from right TPJ and precuneus, but not DLPFC, suggesting representations of others' wellbeing is modulated according to the prosocial regulatory goal but only when it requires theory of mind. These findings speak to both the domain-generality and domain-selectivity of cognitive regulation of decision making and importantly advance our understanding of the neural systems important for cognitive control and decision making more generally. In general I am supportive of this paper but think some outstanding issues could be better addressed to confirm some of their interpretations and rule out others.”**

**Response 2.1.** We thank the reviewer for their thoughtful comments here and below, as well as for their positivity about the approach!

**Reviewer 2.2: “It is notable that DLPFC flexibly encoded values of tastiness, healthiness, and $Self but not $Other or $Fairness on the one hand but strongly predicted altruistic choices on the other hand. Is the DLPFC prediction of altruistic choices then mediated especially by the change in $Self during altruistic choices, but not the change to $Other or $Fairness? “**

@Anita

**Reviewer 2.3: “A related question: Do the rTPJ and precuneus group effects predict individual differences in regulatory success during altruistic choices when goals depended on another's thoughts or feelings? Such evidence would tie together their argument that these latter regions "assume responsibility" for regulatory success when DLPFC does not because of the component process required to meet that goal.“**

**Response 2.3:** We thank the reviewer for this important suggestion for analysis. Indeed, exactly as one might expect, we find that we can specifically predict regulatory success for changes in weight on $Other via response patterns in the precuneus, and to a lesser extent, TPJ.

**Reviewer 2.4: “Given the prior literature on this topic, and in order to rule out a model whereby VMPFC value representations are modulated in an attribute-specific manner that depends on regulatory goals, it would be informative to see an ROI-based analysis of VMPFC. I may have missed something but I could not find one. The ROI from the conjunction presented in Figure 3 could be used or prior literature could be used.“**  
**Response 2.4:** @Anita

**Reviewer 2.5: “One question which is unclear from the conjunction analyses is whether it is the same neural code (e.g. in DLPFC) that is used across task contexts, or whether the code is distinct (context-dependent) but found in the same brain region. What do the authors find if the SVR is trained on one attribute (e.g. tastiness) and tested on the other two (e.g. healthiness and $self)? Can the overlap between representations be better visualized? “**

**Response 2.5:** @Anita

**Reviewer 2.6: “It would be informative to visualize the feature weights for the key areas (e.g. VMPFC, DLPFC) across voxel space in every subject. This procedure should help to assess to what extent any decoding effects are due to hard anatomical boundaries between subareas (e.g. dorsal and ventral aspects of DLPFC) or to distributed patterns within areas.”**

**Response 2.6:** @Anita

**Reviewer 2.7: “How do the authors interpret the altered representations in some primary sensory and motor areas, e.g. less strong decoding of tastiness values for NC than HC and TC? Have the authors considered a model whereby coupling between DLPFC and distinct regions of sensory cortex is modulated according to the regulatory goal? The DLPFC must get its attribute representations from someplace.”**

**Response R2.7**. We speculate that part of the altered representations in visual and motor cortex may be due to differences in patterns of eye movements, visual attention, and motor arousal. While these effects are not uninteresting, we suspect they are a corollary of changes happening in other areas, rather than drivers of regulation per se.

We wholeheartedly agree with the reviewer that, in theory, information transfer between the DLPFC and other areas could account for changes in attribute representations, as we now explicitly note on pg. XX of the revised manuscript. However, it is not entirely clear what the appropriate method is to test for this coupling, given that attribute encoding effects in the DLPFC appeared only in multivariate response patterns, not univariate response, while current functional connectivity methods assume that connectivity will manifest in the coupling of univariate responses in one region to univariate responses in another. We suspect that functionally coupling actually occurs at the level of distributed representations within and across regions, which complicates the interpretation of effects in univariate connectivity analyses.

Nevertheless, and with that caveat in mind, we performed a complete set of connectivity analyses for four regions of interest (ROIs) identified in our dataset: 1) the VMPFC area showing a conjunction across all attribute representations, 2) the DLPFC area showing a conjunction between regulatory effects on Taste, Health, and Self attributes, 3) & 4) the TPJ and precuneus regions showing an effect of attribute encoding for $Other. For each of these regions, we performed a beta series functional connectivity analysis (Rissman, Gazzaley, & D’Esposito, 2004), chosen because this method has more power to detect context-specific changes in functional connectivity for event-related designs (Cisler, Bush, & Steele, 2013). We hypothesized two possibilities about changes in functional connectivity. On the one had, DLPFC and/or TPJ might be more connected with visual or motor areas during regulation, either on average, or as a function of regulatory success. Alternatively, or in addition, we hypothesized that since the VMPFC represented all attributes independent of regulatory goal, that TPJ and DLPFC might show differential connectivity with the VMPFC during regulation, either overall or as a function of regulatory success. Our results provide some evidence for the latter hypothesis. Using the VMPFC as a seed region, we observed the following patterns:

1. During the Food Choice Task, changes in coupling between the VMPFC and the DLPFC during Health vs. Natural trials correlated significantly with the extent to which participants were able to increase the weight of health in Health vs. Natural trials (P < .005 uncorrected, see Supplementary Methods and Results for more detail).
2. During the Altruism Task, changes in coupling between the VMPFC and DLPFC during (Partner + Ethics) vs. Natural trials *decreased* to the extent that the weight on $Self decreased in (Partner + Ethics) vs. Natural trials.
3. At the same time, increased coupling between the VMPFC and Precuneus on Partner vs Ethics vs. Natural trials also correlated with greater weightin of $Other in Partner vs. Ethics trials (P < .005 uncorrected).

Taken together, these results support a somewhat different model than has currently been advanced in the literature.

Reviewer Comment 2.8. “The one comparison in which a match with the behavioral analyses failed to emerge was for ethical considerations compared to normal or personal considerations. I do not view this null finding as problematic, but out of curiosity do the authors have any thoughts as to what is going on when regulatory success depended on changes to fairness due to the goal of complying with social norms? This null finding should be addressed in the Discussion.” 

Response 2.8: @Cendri

Reviewer #2 (Minor Comments):   
  
Results page 13-14: the explanation of the rather sophisticated and unusual SVR could be clearer. A reader unfamiliar with the methodological approach is unlikely to understand how the analysis works and what it is testing.   
  
  
Reviewer #3 (General assessment and major comments (Required)):   
  
In this manuscript, Tusche and Hutcherson report an fMRI study on how cognitive regulation affects the neural representation of choice-relevant attributes. They look for mechanisms that may generalize across two types of choices, one involving conflict between healthiness and tastiness of food items, the other involving conflict between self-interests and altruistic concerns. In different conditions, participants are asked to focus on one or the other attribute, which regulates the weights assigned to the targeted attributes, as shown via computational modeling of choice data. The key findings are the links between these changes in attribute weights and the decoding accuracy obtained for these attributes using multivariate pattern analysis (MVPA) in various cortical regions. There results are quite convincing, with successful decoding across tasks and individuals. There is no clear conclusion about whether the regulation is centralized or distributed though, since changes in decoding accuracy are observed in the DLPFC for most of the attributes but not all.   
  
Reviewer Comment 3.2: The role of cognitive control in economic choice is poorly understood and this study brings valuable insights by applying MVPA to standard choice paradigms. My main concern is the absence of a mechanistic account linking brain activity to behavioral output. In a sense, the results seem a bit trivial: when participants pay attention to a particular type of information (by instruction), this information is easier to decode from their brain activity. This effect should be seen as the outcome (and not the process) of cognitive regulation. How cognitive regulation is implemented, and how better decoding translates into biased choices, still need to be explained.

Response 3.2  
  
For the latter point, it would be important to show that the decoded value (i.e., the decoder output) correlates with the behavioral weights. The alternative would be that changes in decoding accuracy correspond to changes in precision (i.e., signal-to-noise ratio) and not changes in the signal itself. If correct, this would mean that a region downstream to the DLPFC could just read this value, add it to other values corresponding to other attributes, and feed the aggregate value to a selection process that makes the decision. Perhaps functional connectivity could be used to test for such a transfer of information. Thus, the neural model would parallel the behavioral model.   
  
Other points:   
  
- The correlation across individuals could reflect compliance to the instructions rather than self-regulation capacity. The arguments taken from subjective report and from body-mass index are quite weak. For subjective report it could be that the rating scale is not reflecting the propensity to comply with the instructions. For body-mass index the opposite correlation could be expected: those who regulates food intake in real life should not need instructions in the lab.   
  
- The observation that all attributes are represented in the VMPFC but inaccessible to cognitive regulation is super interesting (and novel, to my knowledge). The dissociation with DLPFC should be more emphasized and discussed. Would this mean that VMPFC representations are closer to stimuli and DLPFC to responses?   
  
- To compare the pattern of attribute weights and the pattern of decoding accuracy across conditions, the authors intend to reproduce significance of pair-wise comparisons. As they know this approach heavily depends on the statistical threshold, which may be matter of debate when multiple comparisons are made. I would favor a straight regression of decoding accuracy against weight (across conditions).