Author’s response to reviews

Title: After a Pair of Self-Control-Intensive Tasks,Sucrose Swishing Improves Subsequent Working Memory Performance

Authors:

Evan C Carter (evan.c.carter@gmail.com)
Michael E McCullough (mikem@miami.edu)

Version: 2 Date: 17 August 2013

Author’s response to reviews:

August 17, 2013

Dear Professor Laws:

Thank you very much for your letter dated July 18, 2013, in which you invited us to revise and resubmit our paper (MS# 1586490809937252), “After a Pair of Self-Control-Intensive Tasks, Sucrose Swishing Improves Subsequent Working Memory Performance” for BMC Psychology. We have revised the paper according to the round of comments we received. We have copied your requests and the reviewers’ comments below (in bold) and added our responses following each of them.

Editor, point 1.

Please acknowledge anyone who contributed towards the article by making substantial contributions to conception, design, acquisition of data, or analysis and interpretation of data, or who was involved in drafting the manuscript or revising it critically for important intellectual content, but who does not meet the criteria for authorship. Please also include the source(s) of funding for each author, and for the manuscript preparation. Authors must describe the role of the funding body, if any, in design, in the collection, analysis, and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication. Please also acknowledge anyone who contributed materials essential for the study. If a language editor has made significant revision of the manuscript, we recommend that you acknowledge the editor by name, where possible.

We have added an acknowledgment section, per BMC Psychology guidelines. It is reproduced here for convenience:

“We gratefully acknowledge Lilly Kofler for running experimental sessions, Brandon Schmeichel for providing us with experimental materials, and, as our funding source, the John Templeton Foundation.”

Editor, point 2.
Manuscripts should include a 'Competing interests' section. This should be placed after the Conclusions/Abbreviations.

We have added a section on competing interests, per BMC Psychology guidelines.

Reviewer 1: Angela Duckworth

Reviewer 1, major point 1.

How can the authors be sure than none of the beverage swished before the working memory task was consumed? If participants inadvertently swallowed some of the beverage, might this account for the observed moderation of prior task difficulty on subsequent working memory performance?

Unfortunately, we are unable to determine for sure whether participants ingested any of the drink they were given. However, in our view, the possibility of accidental ingestion of glucose does not help to account for our results. The following has been added to the discussion on page 17 to address this issue:

“One important limitation of the current study is that we did not measure blood glucose, so we cannot be certain that swishing the glucose sweetened drink did not affect blood glucose; that is, it is possible that some participants swallowed some of the glucose that they were asked to swish. However, given the results of previous work that suggests that swishing procedures that are almost identical to those we used here do not affect blood glucose levels (Molden et al., 2012, Experiment 4), it seems likely that our procedures also did not increase blood glucose. Furthermore, even if participants did ingest some portion of the drinks they were given, our major findings still present problems for the limited strength model because we found no evidence for a decrease in self-control performance following the completion of tasks that required self-control. Consequently, our tentative explanation for the results we did obtain, which rely on the concept of learned industriousness, would still hold (i.e., the presence of glucose in the mouth should function as a reward, rather than as the replenishment of a resource, just as its ingestion should, though perhaps with weaker effect). Nevertheless, future experimenters might consider measuring blood glucose to better arbitrate between the effects of sensing glucose in the mouth rather than in the digestive system.”

Reviewer 1, major point 2.

If the ego depletion model is not correct, what model of self-control do the authors believe to be true? Can more be said about what the mechanisms at work are, in addition to what they seem not to be?

We believe that the mechanism underlying self-control failure is most likely motivational in nature. We discuss how our results are consistent with the variety of motivation-based accounts on page 14. However, based on our data, it is difficult to say more than this. This addition, which has been copied below, should help to address some of the comments of Reviewers 2 and 3 about how
the discussion was too cursory.

"[Our findings are] generally consistent with all motivation-based accounts of performance, in which high self-control performance is theorized as being due to a strategic increase in effort designed to achieve tasks that have been deemed important or that result in the receipt of reward (e.g., Eisenberger, 1992; Beedie & Lane, 2012; Baumeister & Vohs, 2007; Molden et al., 2012; Inzlicht & Schmeichel, 2012; Kurzban, Duckworth, Kable, & Myers, in press). Importantly, however, the lack of evidence for the depletion effect (or, in terms of the Bayesian analysis, the evidence for the null model), makes our results difficult to reconcile with all models that predict decreased self-control performance as a function of previous self-control (e.g., Beedie & Lane, 2012; Baumeister & Vohs, 2007; Molden et al., 2012; Inzlicht & Schmeichel, 2012; Kurzban, et al., in press). Instead, our data seem most consistent with an interpretation based on the secondary reward theory of industriousness (Eisenberger, 1992), which does not predict that initial high effort will necessarily lead to subsequent low effort. For example, the taste of sucrose after the initial effort required in the high-effort condition may have reinforced high mental effort so that on the subsequent working memory task, participants worked harder and performed better, whereas for participants in the low-effort condition, the taste of sucrose encouraged continued low-levels of effort."

Reviewer 1, major point 3.

The Vohs et al. (2012) article cited in the introduction compared a condition in which only one or two vs. three or four initial tasks were used to produce the depletion effect. In the current investigation, only two tasks were used. Can the authors rule out the possibility that their study design would have yielded different results if a longer sequence of initial tasks had been used?

This is an excellent point and we have added some text to the discussion (pages 16 and 17) to address it. What we have added here should also satisfy reviewer 2’s major point 5. Our point is that it is impossible to rule out the possibility that we would have obtained different results if more manipulation tasks (or more outcome tasks, per Reviewer 2’s point) were included in our experiment. However, we believe that this issue is a critical one for the limited strength model, because it highlights the fact that the flexibility of resource-based (and some motivation-based accounts) allows one to explain any pattern of data—that is, one can always argue that, had the experiment just included one more task, the depletion effect would have been observed. Based on this argument, we recommend that future work focuses on identifying a measurable mechanism that underlies self-control failure. What we have added to the text has been copied here:

“Notably, the lack of a method for directly measuring the resource on which self-control relies means that resource-based explanations can be made consistent with the pattern of data we report here: For example, one might propose that the depletion effect would have been observed in the present experiment if participants had been required to complete a third initial task (i.e.,
our participants were simply not fully depleted; Vohs et al., 2012). One might also argue that participants who performed well on the OSPAN used their remaining resources to do so, and their depleted state would have been revealed had we included one more dependent variable. It will only be possible to rule such speculations out after the resource underlying self-control has been identified and a method for measuring it developed. Of course, a similar criticism can be leveled at any motivation-based explanation for self-control failure that is not sufficiently specific about the relationship between motivation and self-control. Thus, future work by theorists interested in resource-based and motivation-based explanations of self-control failure, such as the limited strength model, should focus on identifying and directly measuring the resource in question, or the process by which motivation changes (e.g., as proposed by Kurzban et al. [in press], the motivation to perform on a task is exactly a function of opportunity cost: The greater the potential rewards the participant forgoes by putting effort into the task, the lower the participant’s motivation to perform the task).”

Reviewer 2: Stuart Ritchie

Reviewer 2, major point 1.

The citation given for the first task completed by participants (watching a video of a woman being interviewed) is Baumeister et al. (1998), but no such task appears in that study. The task does, however, appear in the Schmeichel (2007) reference, which is given as the citation for the second, vacation essay-writing task – but this task itself does not appear in Schmeichel (2007)! The authors should ensure that the correct citations are given for each measure.

We have fixed this citation oversight with regard to the first depletion task. The vacation essay-writing task was indeed taken from the second experiment reported by Schmeichel (2007). This reference was not changed.

Reviewer 2, major point 2

More importantly, it’s difficult to compare the measures used here with those in previous studies. First, only one ego-depletion-type study—and it’s not one that addressed the glucose question per se—has used the OSPAN, to my knowledge (Shmeichel, 2007), raising the question of why the authors picked a measure that precludes easy comparison to previous glucose-swishing studies.

We have discussed the implications of this issue (page 16) and provided our justification for using the OSPAN in an endnote. Both of these have been copied here:

“…perhaps OSPAN performance does not decrease when participants are depleted, but performance on other outcome tasks, such as persistence at difficult tasks, does (e.g., Baumeister, et al., 1998). It is noteworthy that the OSPAN is not especially widely used in the literature on the limited strength model (Hagger et al., 2010)c. However, according to the limited strength model, performance on any task that is thought to require self-control, such as the OSPAN, should suffer as a function of previous acts of self-control, so if it is true
that the depletion effect is moderated by task type, the limited strength model will require revision on the basis of the results we have reported here.”

“c We choose to use the OSPAN in our experiment because it has been argued that working memory performance indexes something fundamental to self-control (e.g., Hofmann, Friese, Schmeichel, & Baddeley, 2011).”

Reviewer 2, major point 3

Furthermore, Schmeichel (2007) used the OSPAN task as the dependent variable, but did not extract one factor as the authors have here. The authors don’t give the raw scores on each of the OSPAN outcomes (just on the overall factor), making it even harder to compare across studies – the authors should perhaps add these to their supplementary table, or elsewhere. I also think it’s important the authors give a fuller description of the OSPAN task in the paper – it isn’t enough to just refer to another paper. For instance, what does “number of words in full sets” (p. 9) mean?

We have added the raw scores to the supplementary materials (supplementary table S1), and we have extended our description of the OSPAN on pages 9 and 10. The description has been copied here:

“Next, participants completed a version of a working memory task called the operation span (OSPAN), in which participants were presented with sets of words to remember. Participants were presented with 15 sets of words, containing between two and five words. In each set, words were presented one at a time, and the presentation of a word was followed by the presentation of a mathematical equality, such as (9 x 3) – 1 = 2. Participants were instructed to remember each word until the end of a set, at which point they were asked to recall as many words as possible from only the set they had just completed. Additionally, participants were instructed that when they saw an equality, they were to respond either “yes” or “no” to indicate whether they believed the equality was true. The rate of presentation of word/equality pairs was controlled by the participant. The OSPAN provides four possible measures of working memory performance: the total number of full sets of words remembered (maximum 15 sets), the total number of words remembered across all sets (maximum 48 words), the longest set of words remembered (maximum five words), and the number of words in fully recalled sets only (maximum 48 words). Schmeichel (2007) reported that OSPAN performance, as measured by each of the above variables, generally decreased for participants who had previously exercised self-control (i.e., the depletion effect).”

Reviewer 2, major point 4

Incidentally, one wonders why the authors used PCA, and not factor analysis, to extract the overall OSPAN score. I’m fairly sure there wouldn’t be much difference between the score from PCA or FA, but the authors should justify their choice of method (and should be consistent – in the graph, the score is referred to as a “factor”, not a “component”).
We have justified our use of PCA in an endnote and made our language more consistent, as suggested. The endnote has been copied below:

“Principal component analysis, or PCA, is a mathematical transformation that allows researchers to reduce a set of variables to one or more so-called components. PCA is technically different from factor analysis methods, which are based on the common factor model and should be used when the researcher wishes to explore how unobserved latent variables (i.e., factors) underlie the correlations between measured variables. Factor analysis, therefore, is based on a specific model and tests a specific hypothesis. PCA, on the other hand, should be used when a researcher simply wishes to reduce a set of measures down to a set of independent components that account for as much of the variance between the observed variables as possible (Fabrigar, Wegener, MacCallum, and Strahan, 1999). We choose to use PCA because we were concerned with how our experimental manipulations would affect variance in OSPAN performance scores, rather than any hypothesis about the latent variables underlying OSPAN performance.”

Reviewer 2, major point 5

I am quite sympathetic to the authors when they say they don’t want to engage in too much post-hoc speculation about the results. However, I still think the discussion section is rather too short, and needs some expansion. Are the results here completely incompatible with the ego-depletion-as-glucose theory? What might have happened, for instance, if a fourth task had been included at some point after the OSPAN task? Given that OSPAN performance is—on the author’s interpretation, at least—to some extent a reaction to the reward of the swishing, might it be that a fourth task would show decreased performance for those who had previously had higher scores? This would be because, when swishing the glucose drink, their systems had been ‘fooled’ into thinking there was the prospect of plentiful glucose in future, and thus using up the resources they had at the time, which helped them get a higher score. This wouldn’t explain the main effect of effort found here – it still would be difficult to understand why low effort led to later low performance, even when rewarded. But I still think the authors should reflect on what would have happened on that hypothetical fourth test (and ideally, someone should do such a study!).

We have addressed this point while simultaneously addressing Reviewer 1’s major point 3 (pages 16 and 17). For convenience, we’ve copied the text that has been added below:

“Notably, the lack of a method for directly measuring the resource on which self-control relies means that resource-based explanations can be made consistent with the pattern of data we report here: For example, one might propose that the depletion effect would have been observed in the present experiment if participants had been required to complete a third initial task (i.e., our participants were simply not fully depleted; Vohs et al., 2012). One might also argue that participants who performed well on the OSPAN used their remaining resources to do so, and their depleted state would have been revealed had we
included one more dependent variable. It will only be possible to rule such speculations out after the resource underlying self-control has been identified and a method for measuring it developed. Of course, a similar criticism can be leveled at any motivation-based explanation for self-control failure that is not sufficiently specific about the relationship between motivation and self-control. Thus, future work by theorists interested in resource-based and motivation-based explanations of self-control failure, such as the limited strength model, should focus on identifying and directly measuring the resource in question, or the process by which motivation changes (e.g., as proposed by Kurzban et al. [in press], the motivation to perform on a task is exactly a function of opportunity cost: The greater the potential rewards the participant forgoes by putting effort into the task, the lower the participant’s motivation to perform the task).”

Reviewer 2, major point 6

Given the authors talk about industriousness, it might be interesting to discuss how their results might fit with the ‘resource allocation’/motivational theory of Beedie and Lane (2011; Personality and Social Psychology Review).

We appreciate this suggestion and have added a discussion about how our results are consistent with motivation-based explanations of self-control failure, such as the explanation provided by Beedie and Lane (page 14). We also addressed this point in part through our discussion of Reviewer 1’s point 3 and Reviewer 2’s previous point. The text we have added is copied here, as well:

“[Our findings are] generally consistent with all motivation-based accounts of performance, in which high self-control performance is theorized as being due to a strategic increase in effort designed to achieve tasks that have been deemed important or that result in the receipt of reward (e.g., Eisenberger, 1992; Beedie & Lane, 2012; Baumeister & Vohs, 2007; Molden et al., 2012; Inzlicht & Schmeichel, 2012; Kurzban, Duckworth, Kable, & Myers, in press). Importantly, however, the lack of evidence for the depletion effect (or, in terms of the Bayesian analysis, the evidence for the null model), makes our results difficult to reconcile with all models that predict decreased self-control performance as a function of previous self-control (e.g., Beedie & Lane, 2012; Baumeister & Vohs, 2007; Molden et al., 2012; Inzlicht & Schmeichel, 2012; Kurzban, et al., in press). Instead, our data seem most consistent with an interpretation based on the secondary reward theory of industriousness (Eisenberger, 1992), which does not predict that initial high effort will necessarily lead to subsequent low effort. For example, the taste of sucrose after the initial effort required in the high-effort condition may have reinforced high mental effort so that on the subsequent working memory task, participants worked harder and performed better, whereas for participants in the low-effort condition, the taste of sucrose encouraged continued low-levels of effort.”

Reviewer 2, major point 7

The authors state that the end of data collection was not contingent on any particular pattern of results. Obviously this is good, but how did they know that one semester’s worth of participants (surely it would have been better setting a
number, rather than a time limit) would give them adequate power to test the
theory? If they ran a power analysis, it isn’t reported here, but I think it should be
(though this is another issue with conceptual replications; if you don’t use the
same procedure as previous studies, it’s hard to estimate the size of effect you’re
expecting).

We agree. We have added a discussion of this important issue on pages 17 and
18. We provide a detailed discussion of the possibility that our tests were
underpowered, and we try to put this in the context of the overall literature on ego
depletion and what is known about the effect sizes one might realistically expect.
We also point out that our sample sizes were favorably comparable to almost all
of the other published studies on the ego depletion phenomenon. What we have
added to the discussion is copied here:

“A second limitation of the current work is the possibility that our null findings
were the result of inadequate power. We did not conduct an a priori power
analysis for our tests of the depletion effect (as mentioned, our data collection
plan was to collect as much as possible in one semester). A priori power
analyses are difficult to conduct for conceptual replications in any case, however,
because it is not known if the parameter estimates provided by previous work
generalize to the procedures that constitute the conceptual replication.
Nevertheless, assuming the alternative hypothesis is true (i.e., the depletion
effect is non-zero) for participants in the sucralose-sweetened and unsweetened
rinse conditions, then our test of the depletion effect would have had 80% power
for effect sizes of $d = 0.47$ or greater. According to Hagger et al. (2010), who
provided a variety of meta-analytic estimates of the depletion effect for
subsamples of experiments that were methodologically similar to ours, the
depletion effect is at least this large. However, if the depletion effect is
considerably smaller than $d = 0.47$, as our re-analyses of Hagger et al.’s (2010)
meta-analytic data suggest it is (Carter & McCullough, 2013; in press), then the
tests we conducted here are underpowered. For example, if the depletion effect
is actually, say, $d = 0.25$, then any test that comprises fewer than 252
participants per group will have less than 80% power. Note, however, that 188 of
the 198 experiments reviewed by Hagger et al. (2010) had a total sample size of
$N = 100$ or less, and the two largest experiments had total sample sizes of $N =
284$ and 501). Thus, to the extent that our study was underpowered, the majority
of the published literature on the depletion effect is similarly hobbled by this
limitation. Therefore, as we have recommended elsewhere (Carter &
McCullough, 2013; in press), it is critical in the future that researchers conduct
large-scale (i.e., experiments with total samples of at least $N = 504$) direct
replications of the classic tests of the depletion effect (e.g., Baumeister et al.,
1998).”

Reviewer 2, major point 8

While the clarity of the writing in the paper is generally good, I thought the
abstract was relatively poorly written. For instance, I think the “Results” part
needs to be far clearer and less vague (“…some evidence…”). The “Conclusion”
part mentions “null” results, but the results were significant – they were just in the
opposite direction to what you’d predict on the Baumeister theory. This should be clarified.

We have tried to improve the abstract per Reviewer 2’s suggestions.

Reviewer 2, minor point 1
On p.3, the authors state that swishing sucrose in the mouth “does not increase blood glucose”. It’s not immediately clear which of the references they cite backs this point up, so I think this statement—critical as it is—needs to have the appropriate reference directly after it.

We have rearranged the citations so that this problem is solved.

Reviewer 2, minor point 2
Saying the participants ‘found the task less difficult’, for instance on p. 9, is confusing - it could sound as if this is the objective outcome. For that reason, I think the authors should refer to this measure as ‘subjective difficulty’, to differentiate it from the actual scores, which come afterwards.

We have modified our language to help make this point.

Reviewer 2, minor point 3
What’s the justification for using SEs as error bars in the figure, instead of the usual 95% CIs? And would a bracket with one asterisk not be clearer than two asterisks?

The design of Figure one was based simply on taste. However, we have modified it as suggested by Reviewer 2 and do feel that the figure has been improved.

Reviewer 2, minor point 4
I don’t like the statement about the results being worth “more than a perfunctory glance” (p. 12) – I’d prefer the words in the discussion were used to make more substantive points, and the authors should give themselves some more credit!

Reviewer 2’s point is well taken and we have modified our language.

Reviewer 2, minor point 5
p. 8: “Bonferoni” should be “Bonferroni”.

We have corrected this misspelling.

Reviewer 2, minor point 6
There are a few problems with the references. The authors need to check that all their citations are mentioned in the reference section. On p. 11, "Wagenmaker" should be "Wagenmakers", and this paper is not included in the reference section. The Sanders et al. (2012) paper, mentioned on pp. 3 and 5, is not included in the reference section. On p. 15, "Negativeland" should be "Negativland". Also, the authors should make sure that all the journal titles are italicised in the reference section.
We have gone through and double checked all of our citations. They should now be error-free and correctly formatted.

Reviewer 2, discretionary revision 1
Does the Bayesian analysis actually add anything? I've never seen the point of this kind of analysis, especially since in this case the results seem to be exactly the same in both sets of analyses. Maybe the authors can save space by leaving this out.

We firmly believe that the Bayesian analysis adds to the paper, as we explain in the main text on page 12. We think that including this analysis is absolutely worth the space that it takes up (which, in truth, is not much).

Reviewer 3: Andrew Dunn

Reviewer 3, major point 1
I appreciate that this is in some ways an exploratory endeavour, however there are some very clear and obvious hypotheses that could and should be made in the introduction, that stem from the effort depletion literature in particular. Possible direction of effects might be hinted at but they are not clearly stated. This may sound old fashioned, but by providing the hypotheses you will achieve two important things (a) make clear and more manageable your results section (and the findings therein) & (b) emphasize your null effect findings.

This is an excellent suggestion, but given the exploratory nature of our experiment, somewhat difficult to comply with. Instead, on pages 6 and 7 we have added a description of the possible predictions one might make based on the different models that have been proposed (copied below). We hope this addition achieves what Reviewer 3 was suggesting, though we remain open to considering other options, too, if deemed prudent.

“Based on the limited strength model (Baumeister, et al., 1998; Vohs, et al., 2012), one would predict that participants who complete two initial self-control tasks should perform worse on a third self-control task compared to participants who complete two initial tasks that are relatively less self-control-intensive (i.e., the depletion effect). Based on the work by Gailliot et al. (2007), one would also predict that the depletion effect would not be observed for participants who have ingested glucose. However, based on more recent work (Molden, et al., 2012; Sanders, Shirk, Burgin, & Martin, 2012; Hagger & Chatzisarantis, 2013), one would predict that the depletion effect should also be reduced for participants who merely rinsed their mouths with a drink containing glucose, not necessarily only those that ingested glucose.

In contrast, based on experiments inspired by learned industriousness, one should predict that completing two self-control-intensive tasks (i.e., expending a relatively higher amount of effort) should actually increase subsequent self-control performance (e.g., Eisenberger, Masterson, & McDermitt, 1982; Eisenberger, 1992; Converse & DeShon, 2009). The mechanism thought to
underlie findings in the learned industriousness work is described by the secondary reward theory of industriousness: Rewarding high effort results in continued high effort because the sensation of effort is learned as a predictor of reward (Eisenberger, 1992). Based on the secondary reward theory of industriousness, therefore, one would predict that increased self-control performance on a third task will only follow the completion of self-control-intensive tasks if participants are subsequently rewarded in some way. In the current experiment, the sweet taste of either sucrose- or sucralose-sweetened drinks may be rewarding, so based on the secondary reward theory of industriousness, one would predict that participants who swish either sucrose- or sucralose-sweetened drinks following high effort (i.e., completing two self-control tasks, rather than two relatively less self-control-intensive tasks) will perform better on the third self-control task. Note, however, that Converse and DeShon (2009) found that, for participants who completed multiple, unrewarded self-control tasks, subsequent self-control performance was improved relative to the performance of participants who completed multiple, unrewarded tasks that required relatively less self-control. Based on these findings, one would predict that completing multiple initial self-control-intensive tasks should increase subsequent performance, regardless of whether completion of these tasks was rewarded (i.e., regardless of the type of drink given to participants)."

Reviewer 3, major point 2

I found the opening two paragraphs (particularly the second) of the “OSPAN ratings and performance” subsection of the results really hard to follow. This needs to be much clearer and more accessible to your readers. I have a reasonable understanding of component analysis and it’s many virtues but it is not clear what you have done and what it is you are using in your analysis (at least not to me). This is important because this feeds in to the Bayesian analysis that follows (which I understand but only if I guess at the component analysis). This makes me nervous because although I broadly understand the Bayesian analysis and I am convinced that what you claim is happening (or not) is true I understand the specific because I feel as if I am guessing at the component analysis.

We have modified the language in the paragraph in question (page 11) so that it is clearer. The edits we made to address Reviewer 2’s point 4 were also made to help address this point.

Supplemental: are you sure that the bonferroni corrections aren’t just too conservative given the main effect?

Given the exploratory nature of our experiment and the sheer amount of hypothesis testing we conducted, we do not believe our corrections to be too conservative, particularly because they were only applied based on subsets of tests (see page 10), rather than based on all tests. However, we have reported our results with all p values, so that readers can tell what would be statistically significant with different adjustments to the alpha levels of the tests.
Reviewer 3, minor point 1
Rinse initial task and mood ratings. This section left me with an odd feeling of not knowing what could be drawn from them and I guess that’s perhaps one reason for the later analyses. It might be helpful to be more explicit about what is happening. You may want to draw from the supplementary methods for this.

We have modified this section and believe it is now clearer (see page 11).

Reviewer 3, minor point 2
I would like to see the Discussion fleshed out with regards to pertinent literature and the implications the findings have here. At present it feels more like a brief.

We have complied with this suggestion, though we remain open to new considerations that should be incorporated (see pages 13 through 18).

Reviewer 3, discretionary revision 1
End of paragraph 1, pg3: What is a “conceptual replication”?  
We have added a definition and small discussion of the term “conceptual replication” using an end note. That endnote is copied here:

“a Following Schmidt (2009), we use the term conceptual replication to refer to any attempt at replicating a previous test via different methods. Conceptual replication can be contrasted with direct replication, which is a repetition of a previous test via identical methods (Schmidt, 2009). For a test of the depletion effect, most experiments are conceptual replications because different combinations of self-control tasks are used in the sequential task paradigm.”

Reviewer 3, discretionary revision 2
Second new paragraph, page 4: “Nevertheless, based on their continued conviction…”. Unintentional though this may be, it could be interpreted as being highly pejorative. You may want to soften this; you might not.

We have found a less pejorative way to communicate what we have in mind here (page 4).

Reviewer 3, discretionary revision 3
Last sentence of page 4 that runs over to page 5. Is this missing a word?  
Reads: “These patterns obtained, and were interpreted as evidence…”  
Suggestion: “These patterns [were] obtained, and interpreted as evidence…”

“Obtain” can be used as an intransitive verb, as we have done here. We would prefer to leave as is.

Reviewer 3, discretionary revision 4
Page 7 second line: “… condition the end on of data collection…”. I don’t understand this.
We have modified the wording so that it is clearer.

SUMMARY

We have worked hard to provide a thoughtful and thorough response to each comment that you and your reviewers raised. First, we believe we have improved the sections that the reviewers found unclear. Second, we have provided more detailed justification for our choice in experimental and statistical methodology. Third, we have provided more background and discussion to address possible alternate findings and the implications of what we ultimately observed. Finally, we fixed all of the technical and grammatical points that your reviewers raised. We look forward to the possibility that this version of the manuscript might be acceptable for publication in BMC Psychology. If you have additional concerns that you would like for us to address, feel free to contact me directly and we can deal with them immediately.

Thank you very much for your work on our behalf thus far.

Sincerely,

Michael E. McCullough, PhD
Professor