AUTHORS' NOTE TO EDITOR/REVIEWERS:

We have addressed reviewers’ comments as detailed below. Amended sections of the text are denoted in bold type. We have also made some additional edits to the paper that were not requested by reviewers.

Firstly, since we submitted the paper, a debate arose in Health Psychology Review around habit measures. Within this debate, it was argued that the self-report habit measure that we used in this study may be inadequate (Hagger et al, 2015), and alternative measures – namely, assessments of the context-dependent frequency of behaviour, which represents the likelihood of habit forming (Labrecque & Wood, 2015) – may be superior. In light of this debate, we now discuss the limitations of our habit measure more fully on lines 436-446.

Secondly, we have changed the title, from ‘Do habits consistently override intentions?’ to ‘Do habits always override intentions?’ Our original title used the word ‘consistently’ because we wanted to make the point that, when our paper is considered alongside other tests of habit-intention interactions, our results will be seen as inconsistent with them. The data in this paper do not however reveal the consistency with which habits override intentions, and so cannot address this question. Hence, we have changed the word ‘consistently’ to ‘always’, because, through this one single negative case in which habits do not override intentions, our data are sufficient to address this question.

RESPONSES TO REVIEWER 1:

Minor essential revisions:

Line 167: Sample sizes in Table 2 indicate that there were more missing data. Please mention these here as well (for example by stating that sample size differs across analyses because of missing data in the other variables).

AUTHORS' RESPONSE:
Line 167 -168 (now lines 172-173) have now been amended to clarify that the sample of 239 was simply the sample used for the overall T1 and T2 analysis. A footnote has been added to what was Table 2 (now Table 4) to highlight that sample sizes differ across analyses due to missing data on demographic variables, which were controlled for in the main analysis but not the sensitivity analyses.

Lines 171-175: It is not necessary that you state these numbers here as all numbers are reported in table 1.

AUTHORS' RESPONSE:
We have amended this section (now lines 176-179) so that we report participants' typical characteristics ('typically female, White British, employed' etc). We have retained the numbers where describing mean age and BMI, but a description of how many participants were in each weight category (healthy weight, overweight, obese) has been removed, given that these figures are provided in Table 1.

Line 174: Although most people probably know what BMI means, abbreviations should be defined when they are first mentioned and, thus, it should read “body mass index (BMI)”. Please report units when you report descriptive statistics, in this case “kg/m²”.

AUTHORS' RESPONSE:
This abbreviation has been spelt out in full on first usage (lines 178-179) and included in the list of abbreviations. The unit ‘kg/m²’ has also been added to all BMI statistics in the text and in Table 1.

Lines 196-198: Why was BMI used as a categorical variable? I understand that categories are reported for sample description, but I don’t think that it makes sense in the regression analyses. Are results different when you use BMI as continuous predictor in the regression analyses?

AUTHORS' RESPONSE:
Apologies for this error (now on lines 202-203). BMI was indeed used as a continuous variable in the regression analyses. BMI categorisation was only used for sample characterisation, which we now clarify on lines 202-203:

“BMI was categorised into healthy weight (BMI #18.5 < 25 kg/m²), overweight (BMI #25<30 kg/m²) and obese (BMI#30 kg/m²) for sample characterisation purposes only.”

Line 216: As intention was measured with only three items, please also present the other two items here.
AUTHORS' RESPONSE:
All three items are now provided (now lines 221-224):

“Intention was measured using three items (‘I intend to avoid eating unhealthy snacks over the next two weeks’, ‘I want to avoid eating unhealthy snacks over the next two weeks’, ‘I expect to avoid eating unhealthy snacks over the next two weeks’; #=.90).”

Lines 226-240: Please insert a table, which includes the 21 snack foods and which of these were classified as unhealthy.

AUTHORS' RESPONSE:
We have added a table (now Table 2) which lists each of the 21 foods and whether they were deemed unhealthy or not. We have also amended the text (line 244), as this comment made us realise that it is not strictly true to say that foods were classified as ‘healthy’ or ‘unhealthy’. Rather, the aim was to indicate which food were *unhealthy*, not whether the remaining foods were healthy. So, we now refer to whether “snack foods were categorised as unhealthy or not”, rather than “healthy or unhealthy”.

Lines 258-259: This sounds like you mean-centered the interaction term, but I guess that you mean-centered the two continuous predictors and then calculated the interaction term with these mean-centered variables, right? (as this would be the correct way to do it)

AUTHORS' RESPONSE:
Yes, we meant that we mean-centred the habit and intention variables before calculating the interaction term. We have now clarified as follows (lines 264-265):
‘…and an interaction term composed of means-centred habit x intention scores at the third step’.

Lines 271-280: I think it would be nice if you include a table showing all correlations between variables. For example, is there actually no correlation between BMI and all the snack-related measures?

AUTHORS' RESPONSE:
We have added a correlation matrix of pertinent variables (now Table 3). There was no association between BMI and snacking score in our data, which echoes current literature suggesting that snacking is not consistently related to weight status.

Constructing this correlation matrix made us realise that we had made an error in our analyses, by including education (measured categorically) as a continuous variable in correlation and regression analyses. This has been corrected, by
including education as a dichotomous variable (scored as 1 = below university degree level, 2 = degree level or higher). (We also ran a version of all analyses using education as a series of dummy variables [education1 = postgraduate vs all other, education2 = undergraduate degree vs all other, etc], and results remained as in the manuscript. We have not reported the multiple-dummy-variable versions of the analysis for fear of overcomplicating or overemphasising what is essentially a peripheral variable.)

Using the recoded education variable in the analyses has not affected any results in any meaningful way, though the coefficients stated in the paper have now changed very slightly. We have scanned our paper and can confirm that no more errors exist in our analyses.

Line 432: typo

AUTHORS' RESPONSE:
This typo (spelling of ‘deviation’) has been amended (now line 506).

Table 1: It makes little sense when the headings of the last two columns are n & % when sometimes M, SD, and range are reported. As you specify M, SD, and range in the second column, may be you can do it like this for n and % also (e.g., “male (n, %)

AUTHORS' RESPONSE:
We understand the reviewer’s concern, but rather than use ‘n, %’ on almost every row of Table 1, we have instead now clearly divided the table up into subsections (one section per variable) and indicated in the first row of each subsection whether the statistics in that subsection refer to n and % or means and SDs.

RESPONSES TO REVIEWER 2

Major essential revisions:
1. While I agree with the authors statement in the introduction that unhealthy eating is “a setting in which habits (for unhealthy snacking) could reasonably be expected to be incongruent with intentions (to avoid eating unhealthy snacks)” (Line 124-126), I am not confident that this expectation is well supported by the data presented. Indeed the statement in the discussion “The significant, albeit small, negative correlation between intention and habit suggests that at least some participants had both snacking habits and intentions to avoid snacking.” (Line 369 – 371) appears to recognise a major issue here (i.e. that although it is assumed that habits and intention conflict within this sample this assumption is not empirically tested). In the absence of evidence of intention-habit conflict the critique of past research made in the introduction: “Forecasts of behaviour where habit is strong and intentions are weak, and vice versa, thus lack ecological validity” could equally apply to this area of research.

While this assumption is certainly more justified in this context than in some
others where intention-habit interactions have previously been studied, given that a major rationale for this study was the need to investigate conflicting habits and intention, I believe that this issue warrants further consideration within this manuscript (either in the analyses or on the discussion).

AUTHORS' RESPONSE:

We recognize the reviewer's concern that the small negative correlation indicates that we may have failed to adequately capture situations in which habits and intentions conflict. However, we do not see that this compromises the validity of our data or conclusions in the same way that strongly positive habit-intentions compromise the ecological validity of observed interactions. If habit correlates strongly and positively with intention, then participants are highly unlikely to hold both strong habits and weak intentions, and so predictions of behaviour in these settings will lack validity. As we now outline in our revised Discussion, a weak negative correlation, or indeed no meaningful correlation, between habit and intention suggests that it is possible for participants to have strong habits and weak intentions (and vice versa), or that intention strength may vary independently of habit. In such instances, modeling behaviour where habit is strong and intention weak can be valid. Measuring habit and intention in the same direction also risks compromising the validity of self-reports, because respondents may fail to recognize the distinction between what they see as very similar measures. Specifying habit and intention in opposing directions reduces this risk.

We have added a paragraph outlining these arguments to the limitations section of the Discussion (lines 415-434):

“Limitations must be acknowledged. We modelled relationships between intentions and counterintentional habits in relation to diet, a domain in which we expected many conflicting intentions and habits. Yet, the small negative intention-habit correlation indicated that many participants did not hold directly opposed habits and intentions. It is possible that a true interaction may have emerged had habits and intentions more strongly conflicted. Future work might explore the role of counterintentional habits in the intention-behaviour relation more reliably by purposefully recruiting samples most likely to hold incongruent habits and intentions (e.g. new dieters), or examining behaviours likely to invite such conflict (e.g. habitual speeding versus intending to adhere to the speed limit). However, the lack of strong habit-intention conflict need not invalidate our findings. Where habit and intention correlate strongly and positively, predictions of action where habit is strong and intention weak can lack validity. A negative correlation, or no meaningful correlation, is most likely to indicate that participants either have opposing habits and intentions, or that intention strength varies independently of habit, in which case no such threat to validity is posed. Furthermore, measuring incongruent habits and intentions reduces the likelihood that participants will give similar answers to both sets of questions due to not recognising the distinction between them, which may render results more reliable (Ogden, 2003).”
2. In a number of places within the manuscript that authors appear to imply a three-way interaction between intention, habit, and self-control:

Line 344-347 “These findings undermine the hypothesis that habits will always moderate the intention-behaviour relationship by showing that, where intention is accompanied by self-control, habitual action can be prevented.”

377-381 “People with snacking habits and intentions to avoid snacking may be better able to inhibit their habitual tendencies on occasions where their intentions are particularly salient and self-regulatory capacity is strong, and less able where self-regulatory capacity is diminished (e.g. availability of attention and memory resources) or other goals are prioritised.”

The authors should explicitly mention this possibility and make reference to other models that include all three variables (e.g. the Temporal Self-Regulation Theory) as well as associated empirical literature (some of which has investigated snacking behaviour) in order to provide more context to this part of the discussion.

AUTHORS’ RESPONSE:

The three-way interaction that was previously implicit in our argument has now been made explicit, and we refer to Temporal Self-Regulation Theory in making this argument. We also cite evidence that self-control is associated with dietary behaviour (de Ridder et al, 2012; Wills et al, 2007). We did not wish to make further additions to this section because the point about self-control regulating habitual tendencies, while proposed by habit theory (e.g. Gardner, 2015) and supported by a recent study of habit inhibition under situations of reduced self-control (Neal et al, 2013; both cited in the paper), does not arise directly from our data.

The revised section of the Discussion (lines 340-356) reads:

“Habitual tendencies can therefore be inhibited (Gardner, 2015b). This may be facilitated by self-control: a wealth of research suggests that people with greater self-control are less likely to engage in unhealthy behaviours, such as eating a high-fat diet (de Ridder, Lensvelt-Mulders, Fineknauer, Stok, & Baumeister, 2012; Wills, Isasi, Mendoza, & Ainette, 2007). Temporal Self-Regulation Theory proposes that ‘prepotent’ default responses, such as those generated by habit, take precedence over alternative responses (e.g. intended responses) unless they are wilfully and effortfully resisted (Hall & Fong, 2007). This predicts a three-way interaction between self-control, habit strength and intention, such that habit strength will overrule intentions only where self-control is weak, but where self-control is strong, the intention-behaviour relationship will be reinstated because prepotent habitual actions are consciously restrained. We did not measure dietary self-control in this study and so could not test this hypothesis. However, one study showed that, under conditions of high self-control, unwanted habits could be inhibited, but where self-regulatory capacity was diminished, people were less able to block their unwanted habits (Neal et al., 2013).”
Minor essential revisions:

1. The authors should be cautious in interpreting main effects within regression models once an interaction term has been added. The interpretation of significant direct effects within regression models that include interaction terms has been the subject of substantial concern within statistical literature. For example, in evaluating the use of interaction models, Brambor et. al. (2006) note that the interpretation of constitutive terms (in this case intention and habit) as unconditional marginal effects is a major error within the interpretation of regression models that include interaction terms. They write: “Scholars should refrain from interpreting the constitutive elements of interaction terms as unconditional or average effects—they are not... the coefficient on the constitutive term X must not be interpreted as the average effect of a change in X on Y as it can in a linear-additive regression model [emphasis in original]” (p.71).

In discussing the manner in which these effects can be interpreted Barbour et. al. note that “absent any knowledge about the distribution of condition Z, the only clear way to gauge the average effect of X on Y is to run an unconditional model in which X is not included in a multiplicative interaction term.” (Brambor, et al., 2006, p. 72). As such, researchers who require an estimate of the average effect of intention and habit on behaviour should ensure that they interpret the coefficients from the step of the regression model that does not include the interaction term (i.e. Step 2). In light of these issues I would suggest that the authors remove the in-text reporting of beta-weights for intention and habit within Step 3 (of course this data should remain in Table 2).

AUTHORS’ RESPONSE:

We thank the reviewer for bringing this argument to our attention, which we will bear in mind for future studies. We have amended the manuscript accordingly by removing description of beta weights for intention and habit at step 3 of the regression models both in the main analysis and sensitivity analyses (lines 282-287 and 291-300).

2. Line 316 – 317 “making estimates of behaviour where habit is strong and intention lack validity (Gardner, 2014).” – missing word?

AUTHORS’ COMMENTS:

We have corrected this typo (‘...making estimates of behaviour where habit is strong and intention WEAK lack validity’; lines 316-317).