November 2, 2017

Dear Dave, dear reviewers,

Thank you very much for your insightful comments on our manuscript. We have now revised the manuscript on the basis of your comments and hereby resubmit the paper for consideration in the *Journal of Semantics*.

In the remainder of this letter we provide details on how we have addressed the comments: the original comments are in black and our responses are in blue.

To make the data and the R code of the experiments available to you all, we have included the link to the github repository (and therefore are resubmitting the paper without anonymizing it).

Kind regards,

Judith, Judith and David

Associate Editor:

I have also reread your manuscript, and find myself in agreement with both the very positive assessments of the reviewers, as well as their points of concern. I agree with Reviewer 1, for example, that the notion of "robustness" could be sharper in places (sometimes it seems like something else, like likelihood, was intended).

As discussed in more detail in response to reviewer 1, we have removed all mentions of "robustness" and clarified our take on projection variability, including a reformulation of the Projection Principle as the Gradient Projection Principle.

Also, I agree with Reviewer 2 that differences between lexical content, propositional content, world knowledge, etc. are difficult to tease apart. It's unclear whether differences you call lexical are truly lexical. For example, the likelihood of flying to the moon is captured as a lexical effect in the following passage, but to me seems related to world knowledge (which could be quite easily shifted in the context of a NASA discussion).

"Our experiments also take into consideration that lexical content may influence projection: a speaker might, for instance, be more likely to be taken to be committed to the content that Alexander flew to New York than to the content that Alexander flew to the moon, simply because people are more likely to fly to New York than the moon."

We have expanded the discussion of the relation between lexical content, world knowledge, and prior event probabilities in the Introduction. See the response to reviewer 2 for details.

Finally, Reviewer 3 makes a variety of important points, including questions concerning the strength of past claims in the literature, and whether they deny variability to the degree portrayed here.

We now discuss Kadmon's claims about projection variability in both sections 1 and 2, engage with her discussion of conventional versus non-conventional triggering of presuppositions in section 4, and also clarify that the projective contents explored in Exps. 1 include both conventional implicatures and presuppositions. See the responses to reviewer 3 for details.

In addition to asking that you respond to these helpful comments, I want to add several comments of my own.

First, I would ask that before each experiment you lay out the logic of including two sub-experiments. What differs between Experiments A and B, why was this difference introduced, and what is its theoretical significance. I can see the difference in items and their nature, but more explicit structure is needed in the paper to make the significance clear.

Thank you for this comment. We have now layed out the logic of the two experiments. At the beginning of section 2 (which now is the section in which experiments 1a and 1b are presented), we mention that we collected projectivity and at-issueness judgments for 19 different projective contents that were instantiated by different lexical contents, that each participant rated both the projectivity and the at-issueness for a given item, that this allowed us to test whether at-issueness predicts projectivity while controlling for between-participant variability, and that Exps. 1a and 1b differed only in the set of projective contents tested.

And at the beginning of section 3 (now the section with experiments 2a and 2b), we mention that we collected at-issueness judgments with this second diagnostic for the same 19 projective contents as in Exps. 1, that we related them to the projectivity judgments collected in Exps. 1, and that Exps. 2a and 2b differed only in the set of projective contents tested, parallel to Exps. 1a and 1b.

Relatedly, the null effect in Experiment 2b needs to be more directly addressed, and its significance explained. This is currently done somewhat obliquely.

We are now more explicit about the fact that Exp. 2b yielded a null result and have also added prose in section 3.3 to remind the reader that Exps. 1 allowed for analysis at individual participants' data points (1,890 and 2,820 data points) whereas Exps. 2 only allowed for analysis at the item-wise means (43 and 240 data points). We therefore consider the findings of Exps. 1 to carry more weight than those of Exps. 2.

Second, it strikes me taht the content in the final 3 pages of the Introduction is more procedural than conceptual in nature, and belongs, I believe, in the Method section, not the Introduction to the paper. What the reader needs to know at this point in the paper is that the authors planned to include a variety of items that varied along the mentioned dimensions, but not the remaining details. I recommend that you delete the last three pages of the introduction, and (1) replace them with the summary you currently provide at the end of that section, adding that details will be presented in the method section, and (2) that in the method section provide only an overview of the variability, pointing readers to an Appendix for details.

We agree and have followed your advice to move the content of the final 3 pages of what previously was section 2 into the relevant experimental sections (which are now sections 2 and 3) and appendices.

Finally, I would ask you to consider sharing your data, code, and materials on a publicly available archive. This does not need to happen now, but I encourage you to consider doing so eventually.

Footnote 7 includes a link to the repository that includes data, analysis scripts, experiment files, and documentation of pilot experiments.

Reviewer 1

I found this a very interesting paper: it presents valuable new data that considerably enhance our understanding of the relationship between measures of at-issueness and projection, extending coverage to a wider range of triggers, and offers a nuanced and thoughtful discussion of the theoretical consequences of these results. Nevertheless, I felt that there were a couple of issues that could usefully be clarified before the paper is published, and in some places it seemed to me that the exposition made use of some terminology that remained carefully undefined.

The main concern I have is that the interpretation of Beaver et al.'s Projection Principle seems a little unstable. In the formulation quoted here, the PP holds that "C projects if and only if C is not at-issue". The authors are careful to clarify that, following Beaver et al., they are agnostic as to the causal relationship involved, and are only interested in testing whether the claimed relationship exists. However, they construe the PP variously as predicting that "projectivity is a function of at-issueness" (p.31), predicting projection variability (p.3), and for instance predicting that "if the content of NRRCs is more robustly not at-issue than the content of the complement of 'discover', the former projects more robustly than the latter" (p.3).

We agree that we may have overinterpreted the Projection Principle provided in Simons et al 2010 and Beaver et al 2017. We have clarified that we take projectivity to be a gradient property (as suggested by previous experimental results and long-standing intuitions about differences in how projective different projective contents are) and have also formulated a revised version of the Projection Principle, which we call the Gradient Projection Principle:

> **Gradient Projection Principle:** If content C is expressed by a constituent embedded under an entailment-canceling operator, then C projects to the extent that it is not at-issue.

It is this revised principle that the experiments were designed to test. Thank you for encouraging us to be explicit.

I may be worrying unduly about this, but the definition of at-issueness appealed to (at least initially) here appears to be one that is entirely crisp and exhibits no gradience: something is either at-issue or not at-issue. This fits with the PP: content that is not at-issue projects, content that is at-issue does not. So although it's true to say that the PP predicts that "projectivity is a function of at-issueness", it really does so in a trivial way, because there are only two values that at-issueness can take, just as there are only two values that projectivity can take. The experiments reported in this paper also fit with this view of the phenomena, in that participants are asked for clear-cut binary judgements.

We are now more explicit in section 1 about the fact that we take at-issueness, like projectivity, to be a gradient property. In line with our assumption that both of these properties are gradient, our experiments elicited gradient responses on a sliding scale.

With the exception of a few occasions on which the authors use the term "more at-issue", they seem to be adopting a view in which at-issueness really is a yes/no matter, and use the word "robust" (many times) to characterise the dimension along which at-issueness can vary. But I'm never quite sure what the authors mean by "robustly projective" or "robustly (not) at-issue". I surmise that most of the time they mean that content is taken to be projective or (not) at-issue either by most people who encounter it, or in most contexts in which it surfaces, or with a high probability for a given hearer in a given context. (There's an exception to this on p.11, where the

observation is made that "at least half of the participants took it to be robustly projective": here "robustly projective" seems just to mean "projective", where the point is that to answer a "certain that" question in the affirmative one has to be confident in the projectivity of the relevant content.)

Thank you for pointing out that we used the terms “robust” and “robustly” ambiguously. We have removed all mentions of “robustly projective” and “robustly not-at-issue” and replaced it with language that is in line with our conception of projectivity and at-issueness as gradient properties: for instance, “this projective content robustly projects” became “this projective content is highly projective”.

I think it would be useful to clarify how this notion of "robust" projection/at-issueness relates to the hypotheses under test. My (perhaps naive) reading of the PP would be that the content projects iff the hearer judges it not at-issue, and the possibility of variation stems purely from the fact that the hearer's confidence about whether the content is truly not at-issue might vary depending on a range of factors (such as those discussed later in the paper). But from this point of view, the approach taken to the analysis of the data is a little surprising. Certainly the PP predicts that, if content C1 is considered not at-issue more than content C2, it should be considered to project more often.

Formulating the relevant principle to be tested as the Gradient Projection Principle has helped us be more explicit about the fact that we treat projectivity and at-issueness as gradient properties. Specifically, we are now clearer about the claim that the more not-at-issue a content is, the more projective it is. In section 1, we now also offer two possible ways of interpreting the claim that projectivity is a gradient property, along the following lines: On a first interpretation, a listener’s (or reader’s) judgment that a content is projective to a certain extent means that the listener takes the speaker (or writer) to be committed to the content to that extent. On this interpretation, projectivity being a gradient property is a consequence of speaker commitment being a gradient property. On a second interpretation, a listener’s judgment that a content is projective to a certain extent reflects the probability with which they believe the speaker to be committed to the content. On this interpretation, speaker commitment may be a binary, categorical property and projection variability arises from the listener’s uncertainty about the whether the speaker is committed. In this paper, we remain agnostic about the underlying interpretation of this gradience, though our discussion of projection variability will be in line with the first interpretation.

But more specifically, the PP predicts that if content C is considered not at-issue in X% of cases it will be judged to project in X% of cases; and even more specifically, these are supposed to be exactly the same cases. Given that experiment 1 appears to collect judgements about both projection and at-issueness, surely the best test of the PP is exactly how often participants give the same judgements on the same materials, when asked the two separate questions? (Granted, there are issues with the test for at-issueness presented here, so the participants' consistency may be slightly exaggerated, but it would be worth hearing about.)

Thank you very much for this comment! Because each participant in Exps 1a and 1b provided both an at-issueness and a projectivity judgment for each item that the participant saw, these experiments indeed allow us to test whether the Gradient Projection Principle holds within participants and items. In addition to the mixed effects model, which already takes this fact into consideration, we have now also included relevant visualizations for Exps. 1a (in the paper) and 1b (in the github repository), and explicitly state (for Exp. 1a) that our findings show that at-issueness predicts projectivity not only at the level of the projective contents (i.e., collapsing over participants and lexical contents), but also at the levels of the individual participants and items

(projective content/lexical content pairings). That is, because participants rated both the at-issueness and the projectivity of each of their items, we are able to show that the at-issueness rating a participant gave to an item predicts their projectivity rating of the item. In Fig. 4, we visualize each participants' projectivity rating against their at-issueness rating for two items, one with a large amount of by-participant variability and one with a small amount of by-participant variability. The full set of item-level ratings is provided in Appendix B.

A few minor points:

p.2-3 - The point about conventionalist approaches made here has already been made in almost the same terms earlier on.

We have removed the duplicate sentence and kept the earlier one.

p.3 - "but does not address the challenge that this variability poses for conventionalist approaches to projection". I guess this depends on whether one could gloss semi-factives as conventionally encoding some specification about how the relevant content relates to the interlocutors' common ground that is weaker than the specification applicable to factives - but that doesn't immediately seem like an outrageous suggestion. I think the key point here may be whether it is possible to sort projective contents into some small finite number of distinct "pigeonholes" which could each be given a conventionalist treatment, or whether the variability is so thoroughgoing that any such enterprise is doomed.

Thank you for this comment. To address it (and related comments by reviewer 3), we now more explicitly discuss in section 4 (the new discussion section) what we take the challenges of the observed projection variability to be for conventionalist approaches to projection. While our experiments show that the expression that the projective content is associated is predictive of the projectivity of the content, these results are compatible with analyses by which the expressions lexically code projectivity but also with analyses by which projectivity is derived from other parts of the conventional meanings of the expressions. We argue that, under the first type of analysis, the lexical specification of projectivity would need to be very fine-grained in order for conventionalist approaches to account for the observed variability. Whether such very fine-grained lexical specifications are possible is a question for future research.

p.4 - "lexical content may influence projection: a speaker might, for instance, be more likely to be taken to be committed to the content that Alexander flew to New York than to the content that Alexander flew to the moon, simply because people are more likely to fly to New York than to the moon." I think the implications of this might need to be spelled out more clearly; I'm not immediately convinced that the speaker's likelihood of commitment to a content necessarily influences its projection. Take an example like

A: You should put a shirt on.
B: Sorry, I didn't realise that the Queen was coming.

One possible analysis of B's utterance is that the presupposition isn't intended to project, because it's mutually evident to A and B that it's false. Another analysis is that it is intended to project, despite being false. I would be inclined to favour the latter, because it better explains the intuition that B's utterance is intended to be ironic (and we need this in order to explain why it's a coherent response to A).

My point in essence is that the likelihood of a speaker being committed to a content - in the sense

of holding that it is true - isn't necessarily a good proxy for the likelihood that they want to express that content, which would require projection in a case such as this. My impression is that the likelihood of wanting to express the content is the more relevant consideration here, although I would welcome an argument to the contrary.

This example is very interesting but, unfortunately, a discussion of the interaction between irony and projection is beyond the scope of this paper.

p.10 - Both the "certain that" and "asking whether" appear to be crisp yes/no diagnostics, which implies an answer to what is meant by "robustness" of projectivity, although the former in particular seems to be getting into the area of eliciting how easy it is for the participant to imagine a set of circumstances under which this utterance could be produced by a speaker who was uncertain about the candidate projective content.

As mentioned above, we now clarify in section 1 that we take at-issueness and projectivity to be gradient properties of content. In line with our assumption that both of these properties are gradient, our experiments elicited gradient responses on a sliding scale.

p.10 - I have slight reservations around the "Is she asking...?" question as a diagnostic for at-issueness, apart from its close conceptual relationship to the projection diagnostic (as discussed on p.21-22). At-issueness is defined earlier as "the ability of content to address the Question Under Discussion", and the examples considered earlier are all responses to explicit QUDs. The relevant notion of at-issueness here seems to be more like "the ability of content to contribute to the Question Under Discussion", and it's not immediately clear that this is the same thing. At least, I think we would need to be shown that what is at-issue in an explicit question is no more or less than what the questioner is asking. (This is not wholly straightforward, because we can certainly prefix a question with material that is at-issue but not being asked about: "This is the gun that killed the victim - have you seen it before?") The discussion on p.30-31 takes this up, but at the cost of rather undermining the definition of at-issueness that has gone before; I wonder whether it would be possible to do a little more at the beginning of the paper to clarify the matter.

Since our goal in this paper is to test the Gradient Projection Principle, we wanted to test it based on two diagnostics for at-issueness that are based on assumptions made in the literature about how at-issue and not-at-issue content differ. By not grounding our experimental investigation in the specific definition of at-issueness assumed in Simons et al 2010 or Beaver et al 2017 (which, as we discuss in the paper, is not the only way to define at-issueness), our experimental findings are more easily generalizable to however at-issueness is ultimately defined. You are right, of course, that it needs to be clarified how the two diagnostics we have used (and other diagnostics used in the field) relate to a theoretical definition of at-issueness, as we discuss in section 3, but we cannot do this in the current paper. To make clearer that we intend to remain neutral here about how at-issueness is defined, we only mention Simons et al's (2010) characterization in section 1 to introduce their Projection Principle, but have removed other mentions of the QUD until we discuss, in section 3, the two main ways in which at-issueness has been defined.

Reviewer 2

The paper provides experimental evidence for variability of projection and correlation between at-issueness and projection for a wide range of presupposition triggers using a fairly straightforward experimental methodology. The paper is clearly written, the results are clear and convincing and constitute an important contribution to the study of presupposition, at-issueness and related concept and should be published as quickly as possible. For this reason, I suggest that

the paper is published with minor revisions.

Overall, I suggest that the paper is shortened wherever possible.

We have tried to keep the length of the paper at bay by moving portions of the materials sections into appendices. Even though we expanded the discussion section, the main body of the paper (without the abstract, appendices or references) remains under 20,000 words.

Also, I list some minor issues in what follows. I will specifically refrain from enumerating typos and comments on style.

1) The connection between the diagnostics for at-issueness and the definition of at-issueness should be made more transparent (partly by putting the bits of pieces addressing this to one location in the paper). It is not entirely clear whether the notion of at-issueness as defined in the literature cited is the same as the one used to in this paper; The paper is fairly aware of this on page 30-31, but not before or after. Crucially, the notion of QUD is not even mentioned when explaining why the diagnostics are supposed to measure at-issueness, instead something like "at issue content being . Also, the fact that the correlation between the two diagnostics is not particularly strong (r = 0.62 or r= 31 is a rather moderate correlation), is an indicator of the fact, that caution is needed.

As discussed in what was previously pp. 30-31, several definitions of and diagnostics for at-issueness are currently used in the literature. We now clarify this in section 1 already and have also more clearly foreshadowed the employed at-issueness diagnostics and the assumptions from the literature they rely on. Because the QUD-based definition of at-issueness is only one of several definitions of at-issueness, we only mention this definition when the Simons et al's (2010) and Beaver et al's (2017) Projection Principle is introduced in section 1, and in the comparison of at-issueness diagnostics (in section 3).

We have softened our tone about the similarity between the two diagnostics by adding the following sentence at the end of the comparison of at-issueness diagnostics: "In short, these findings are compatible with the two diagnostics both measuring at-issueness, though the imperfect correlation suggests that other factors are also contributing to participants' ratings."

I stress that I would not mind at all, if the claim that the diagnostics measure at-issueness is dropped altogether and the paper merely states that the diagnostics measure some notion of prominence which inversely correlated to projection and probably intimately connected to at-issueness. Anyway, the discussion on page 30-31 should be made more prominent.

We have made the relevant discussion more prominent by foreshadowing it in section 1.

2) The title seems somewhat uninspired and too programmatic. I suggest a less programmatic title, especially because of what I mentioned under 1.

Thank you for this comment – we agree and have changed the title of the paper to "On the heterogeneity of projective content: Gradient projectivity and its relation to gradient at-issueness"

3) I find the distinction between "lexical content" and "projective content" highly confusing and I also don't understand why it is needed. Why not just use "propositional content" and specify any caveats at the relevant points in the paper.

We distinguish “projective content” and “lexical content” because the projective content associated with an expression (e.g., ‘the content of the complement of *discover*’) can be instantiated by different lexical contents (e.g., ‘Richie is a stuntman’ or ‘Martha has a new BMW’). To better help the reader understand the distinction, we have moved what used to be footnote 5 into the main body of the text in section 1. We also remind the reader of the distinction in the Materials section of Exp 1a (in what is now section 2).

4) The discussion of projection in the introduction is too simplistic. People - especially the ones cited - have always known that context, including syntactic context, plays a role in projection and that projection cannot simply be defined as a relation between a trigger and the common ground. Notions such as "local" or "global" accommodation, have often been used to deal with cases in which projection somehow failed to arise. One cannot entirely ignore this.

We now acknowledge in the introduction that local accommodation can be used to account for content not projecting under conventionalist approaches, while also highlighting that this process is not understood well enough to capture systematic differences between projective contents associated with different expressions.

5) On page 4 the discussion of the impact of "lexical content" on projection should be phrased in terms of prior probabilities and relevant work should be cited, e.g. in the domain of Game Theory, Rational Speech Act Models etc.

We have expanded the discussion of lexical content and clarified its relation to world knowledge as captured in varying prior event probabilities. Specifically, we say that our experiments also take into consideration that world knowledge may influence projection: a speaker might, for instance, be more likely to be taken to be committed to a content describing an event of Alexander flying to New York than to a content describing an event of Alexander flying to the moon, simply because people are more likely to fly to New York than the moon. Thus, whether projective content projects may depend on the prior probability of the event described by lexical content such that content that describes more a priori likely event may be more likely to project. If this is right, then, for example, the content of the clausal complement of *Did Bill discover that Alexander flew to New York?* should be more likely to project than that of *Did Bill discover that Alexander flew to the moon?*. We then go on to clarify that, in this paper, we do not systematically manipulate the prior probabilities of events, but we do introduce event-type variability by including lexical contents that describe a wide variety of events. In our experiments, these lexical contents instantiate the projective contents associated with expressions like discover.

In a footnote 6, we now mention the relation to the probabilistic pragmatics literature: “World knowledge, i.e., subjective prior probabilities assigned to events, has been shown to affect interpretation in ways captured by models of language use that treat interpretation as Bayesian reasoning about an observed utterance (see, e.g., Franke & Jaeger 2016, Goodman & Frank 2016).”

6) The presentation of the experimental method is way too long and cumbersome. A lot of material belongs into the appendix. For example (10) can be condensed to 2-3 lines instead of 12! Same goes for 9, where most of the examples are either superfluous or belong into the appendix. Similarly, (15) is a waste of space. I think the presentation of the experiments can be shortened by at least 3 pages.

We have now condensed (what used to be) (10) into one line. As for (what used to be) (9), we think it's important to give the reader an example of each target expression and corresponding projective content, given that these expressions are more heterogeneous and some, like *be stupid to* are less-discussed in the literature. We moved what was previously (15) – the list of lexical contents for Exp 1a – to Appendix A and instead only illustrate the three lexical contents that the projective content of "stop" is instantiated with. We have done the same for Exp. 1b, where we moved what was previously (18) into Appendix C.

7) Footnote 8 is confusing. What were the alternative diagnostics tested? What exactly is the purpose of footnote 8. Either elaborate or delete.

The footnote was originally included to acknowledge that piloting was conducted to establish the usefulness of different diagnostics. But we agree that the amount of information provided in the footnote was not useful. We have therefore removed the footnote and have instead included a supplementary report containing details about the conducted pilot studies in the GitHub repository. Footnote 12, which had also alluded to the pilot studies, was removed for the same reason.

8) Is there a reason why a correlation coefficient was not provided for at-issueness against projectivity? I understand that the data show that at-issueness is a predictor of projectivity, but the natural measure expected is a correlation analysis. If there are problems with it, what are those problems? This is all the more surprising as the correlation between the at two at-issueness diagnostics were given.

We now include correlation coefficients (collapsed and not collapsed across lexical contents) in addition to the mixed effects analyses for all four experiments.

Also, I did not understand the plot in Figure 10. A bit of explanation would be useful, even just stating what the X and what the Y axis represent and what the colors code.

We have clarified the contents of (what is now) Figure 11 further in the caption.

9) I totally failed to understand the title of section 4. If anything the title should be something like: "Another test for at-issueness". But even so, I am not sure a new section is justified. One could have just as well group all experiments together and present all the results as one big experiment 1a, 1b, 1c...

Sorry, that section title was an unfortunate typo from an earlier version. We have updated the title of the section (now section 3) as follows: "*Confirming the role of at-issueness in projectivity*". We have also removed all other mentions of "information structure".

Reviewer 3

This paper has two goals: (i) to provide empirical evidence for the variability of projective content, and (ii) to test the hypothesis that projection variability correlates with the at-issueness of the projective content (the "Projection Principle"). The authors report on two pairs of experiments that provide support both for projection variability and for the Projection Principle. The content of the paper is both theoretically and methodologically sound, relevant for current discussions in the field, and appropriate for this journal. The paper is well organized and makes important connections with the existing literature on projection. However, I think certain aspects need to be fleshed out and further explained, as detailed below.

Recommendation: Accept with revisions

1. In the beginning of the paper the authors mention both conventionalist and pragmatic approaches of projection but only explain the conventionalist ones. Authors should briefly define and give references of what they consider to be pragmatic approaches.

At the beginning of the paper, we only mention conventionalist approaches to projection because they give the most successful and widely adopted answer to the question of why content projects and also because they are particularly challenged by projection variability. However, we now also acknowledge the existence of non-conventionalist approaches to projection in footnote 3 of section 1 and provide references to such approaches there. We now also discuss in section 4 (the new discussion section) how non-conventionalist approaches to projection might deal with projection variability.

2. In "conventionalist" approaches, projective content projects because it is conventional, i.e. lexically specified. When we question this approach, a question arises: if we say that at-issueness predicts projection variability, then is projective content lexically specified but can sometimes be overridden? To what extent is it "part of the grammar"? Or is it not lexically specified but entirely context-dependent? The possible lexical specification of projective content is "lurking" in the text but is never really addressed: is it a lexical property or not? And if yes, can content be lexically specified to be projective to varying degrees?

These are all important questions that we now engage with more explicitly in section 4. Specifically, we point out there that, following Kadmon 2001, it is possible that some differences in projectivity are due to the projectivity of some projective content being conventionally specified and the projectivity of other projective content being non-conventionally derived. We also point out, following Abrusán, that differences in projectivity may be derivable from a conventional specification of projectivity and other differences between triggers (such as syntactic status, focus, anaphoricity, etc). By and large, however, these questions remain open and we hope that our experimental findings about projection variability and the role of at-issueness, lexical content and the expression that the projective content is associated with contribute to our understanding of the empirical patterns and an empirically adequate analysis for accounting for those patterns.

3. What is the relation between the obligatory local effect and projectivity? This is alluded to on page 7 but not really explained. Also, the examples in (12) are somehow artificial: are the acceptability judgments always as clear with other triggers? Have the authors looked at naturally-occurring examples for similar items?

Projective content that does not have Obligatory Local Effect is predicted on Beaver et al's (2017) definition of at-issueness to be highly not-at-issue content and hence is expected, by the Gradient Projection Principle, to be highly projective. Because we want to remain agnostic in this paper about the formal characterization of at-issueness (see responses above), we have relegated the discussion of Obligatory Local Effect and the aforementioned prediction into footnote 11. As a consequence, the examples in (what used to be) (12) have been removed.

4. Page 20: the authors state that the literature assumes that distinctions in projectivity are categorical. Really? This does not make justice to certain accounts: cf. Kadmon (2001: 222), on her chapter on types of presuppositions, talks of a continuum (even if she then goes on to propose categories within that continuum). She says: "But how robust a ps seems to be is a matter of degree. There is a whole continuum of pss of various degrees of robustness" (223). Amaral and

Cummins (2015: 168) mention "gradience within classes of triggers", and also in their previous work talk about a "continuum" of expressions. It would be good to include these accounts in the paper, and to relate the current paper to this conversation.

Thank you very much for reminding us of Kadmon's (2001) discussion of projection variability. We now mention her discussion in section 1, when projection variability is first introduced to the reader, as well as in section 2, when we mention that the projective contents explored in Exp. 1 allow us to test whether classifications like the distinction between "soft" and "hard triggers" align with what we observe in the experiment or whether, as Kadmon (2001:223) suggested, "[t]here is a whole continuum of [presuppositions] of various degrees of robustness, a continuum on which no point of qualitative difference in robustness can be found" (we included in this quote in the paper). We also reference her work in section 4, in our discussion of the implications of our findings for conventionalist versus non-conventionalist approaches to projection.

As for Amaral & Cummins 2015, we already cited this work in the first version of the paper but since this work explores at-issueness, not projectivity, it was mentioned in our discussion of variable at-issueness among projective contents in section 1. The work by these authors on variable at-issueness among projective contents associated with English expressions is also discussed in (what is now) section 3.3.

5. Something that is strikingly absent in most of the paper is the use of the term "presupposition", that only appears in the Discussion section at the end. By using "projective content" instead, with just one very large category, I wonder if the findings do not become just a confirmation of what we already knew. True, the authors convincingly provide experimental evidence showing that projective content varies in projectivity (research question 1), and this is an important contribution, but it is also a consequence of the fact that the authors are glossing over a number of distinctions between types of meaning and lumping them in one category, by using the label "projective content". One example of this: on page 20 the authors mention the case of *only* and how the results defy previous (mostly binary) classifications. I agree that most classifications of presupposition triggers rely on binary distinctions, like "soft" and " hard", but the expressions used in the critical items include other types of meanings, not just presuppositions. In talking about "projective content" in general the authors are casting a very wide net that captures several different types of implications. While it's true that the authors also show variability within categories supposed to be homogeneous (e.g. soft and hard ps triggers), they do not address the behavior of specific classes of implications that they consider in their heterogeneous set. Are there any systematic differences between them or only unsystematic variability? One could argue that variability among projective contents is what linguists have tried to capture through different classifications, e.g. presupposition, conventional implicature, assertorically inert implications. The authors should discuss the relation between this broad category of "projective content" and other categories in view of their findings. This does not make their findings less interesting and instead may show that there is experimental evidence for the terminological differences previously made in the literature. They address some of these points in the Discussion session, but they almost exclusively discuss Abrusan's proposals and not others. For example, they do not mention what would be expected of CIs vs presuppositions.

Thank you very much for this comment. We now clarify in the 'Materials' sections of Exp. 1a and 1b that the projective contents we explore are heterogeneous, and include both conventional implicatures as well as presuppositions, and, among the presupposition triggers, include both 'soft' and 'hard' ones, as well as both 'factive' and 'semi-factive' predicates. We do so to prepare the reader for the upcoming discussion in section 2.3 of whether the assumed classifications align with the observed differences in projection variability. And we also pick this up again in section

4, where we now also return to the distinction between conventional implicatures and presuppositions in our discussion of whether projectivity is conventionally coded or non-conventionally derived.

6. Related to this point, it is unclear what the authors mean when they say that one of the possible sources of variability among projective content is "expression associated with projective content" (page 33). Another way of saying this, but with a terminology that the authors explicitly avoid, would be "depending on the trigger". What do the authors mean here: is it a lexical difference, is it degree of encoding of the projective content, is it a new version of hard vs soft triggers? This is too vague as is and needs to be fleshed out.

We agree that we did not spell out what this experimental finding could mean. As we now discuss in section 4, the finding that the expression that the projective content is associated with plays a role in the projectivity of the projective content could mean (at least) two things. First, it could mean that projectivity is coded in the meaning of the expression, as in conventionalist analyses of projection like Heim 1983 or van der Sandt 1992. Second, it could mean that some part of the conventional meaning of the expression predicts the projectivity of the projective content, as in non-conventionalist approaches, or a combination of the two, e.g., along the lines of Abrusán's 2011 analysis.

7. I understand the rationale behind the "Are you sure?" diagnostic but I wonder if the little dialogues that are intended to apply it, like (23), do not license other interpretations from non-linguist native speakers. These small artificial dialogues seem to trigger a contrastive meaning between the two contents (the content of the main clause and that of the complement clause) as if the speaker in the third turn had the intention of making a point about contrasting the content of the two clauses, and that contrast was indeed what the question in the 2nd turn was about. This relates to the results mentioned on page 28: the 'asking whether' diagnostic received higher ratings than the *Are you sure?* diagnostic and there was greater variability in the not-at-issueness ratings with the *Are you sure?* diagnostic (page 29). The " asking whether" diagnostic is less open to inferencing from the part of the participant than the question about certainty (what exactly is it diagnosing?). I think this is an indication that indeed this diagnostic is much more open to other interpretations that are not controlled for in the minimal contexts used in the experiments.

This is an interesting intuition about which factors, besides at-issueness, may have influenced participants' responses in the *Are you sure?* diagnostic. As we now explicitly state at the end of section 3, while both diagnostics seem to measure at-issueness, the imperfect correlation between the two diagnostics suggests that other factors are also contributing to participants' ratings. We also make clearer that the two diagnostics for at-issueness that we used in this paper to explore the Gradient Projection Principle are just two of many diagnostics one could possibly use and we have included in the GitHub repository documentation of pilots we ran on different operationalizations of assumptions about at-issueness. As we note, future research should not only attempt to verify the Gradient Projection Principle using other diagnostics of at-issueness but also clarify how the diagnostics relate to a formal characterization of at-issueness.

Typos:

-Page 26, line 21: of the clausal complements *of* the predicates (add "of")

Fixed.

-page 30, line 1: heterogen*e*ous

Fixed.