

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 below, we have removed all mentions of robustness and clarified our take on projection variability, including an explicit 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 greatly expanded the discussion of the relation between lexical content, world knowledge, and prior event probabilities in the Introduction. See below for detailed responses.

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.

XXX past literature

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.

XXX lay out logic JT: We already did this!

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

XXX null effect Exp 2b

Second, it strikes me that 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 (previously section 2) into the relevant experimental sections 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.

What was previously footnote 7 already indicated that we intended to share data and code. We have now included a link to the repository that includes data, analysis scripts, experiment files, and documentation of pilot experiments.

Referee(s)' Comments to Author:

Referee: 1

Comments to the Author

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 with the reviewer that we may have overinterpreted the Projection Principle as provided in Simons et al 2010 and Beaver et al 2017. We have therefore clarified that we take projectivity to be a gradient property and have formulated a revised version of the principle, the Gradient Projection Principle. It is this revised principle that the experiments were designed to test, so we thank the reviewer 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 now also clarify in Section 1 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.)

We thank the reviewer for pointing out that we have been using the term "robustly" ambiguously. We have removed all the "robustly projective/at-issue" language 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" becomes "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.

Our revision of the Projection Principle into the Gradient Projection Principle has helped us be much more explicit about the fact that we treat projectivity and at-issueness as gradient properties. 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:

*"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, projection variability 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.)

Because Exps 1a and 1b ask each participant for both an at-issueness and a projectivity judgment for an item, 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 visualizations. For Exp. 1a:

*"We emphasize 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."*

For Exp. 1b we have also included the full visualization in the github repository.

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.

XXX add note?

JT suggestion: As discussed in section 2.3., the findings of Exps. 1a and 1b do not provide empirical support for a categorical distinction between factive and semi-factive predicates.

#### ADD DISCUSSION IN SECTION 4 OF KADMON 2001 WHO CONSIDERS ACCOUNTING FOR DIFFERENCES IN PROJECTIVITY THROUGH CONVENTIONAL VERSUS CONVERSATIONAL/PRAGMATIC ANALYSES

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.

XXX ?

JT suggestion: We have clarified in section 1 that our experiments take into consideration the influence of world knowledge on projection in light of the fact that the various lexical contents that instantiate the projective contents in our experiments describe events with different prior probabilities. The findings of experiments 1a and 2a thus can be taken to show that participants' projectivity ratings are influenced by world knowledge. The hypothesis that the likelihood that a speaker wants (or is taken to want) to express a content influences the extent to which they are committed to this content is interesting, but not one that the experiments we report on here can be used to evaluate. The example mentioned in this comment shows that the discourse context plays a role in whether the speaker is taken to be committed to a content (here, the content of the complement of 'realise'). The experiments reported on in our paper were designed to explore the influence of at-issueness and of expressions associated with projective content on the projectivity of content. We have also conducted experiments to explore the role of context on projectivity, but they are reported on in a different manuscript.

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. 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.

XXX ??

JT suggestion: We mention Simons et al's 2010 and Beaver et al's 2017 definitions of at-issueness, which references the QUD, in section 1 in order to introduce the Projection Principle. However, as we discuss in section 3.3, there is no consensus in the literature about how to define at-issueness: definitions/characterizations other than that of BRST have been proposed and it remains an open question how the diagnostics for at-issueness that are currently used in the literature are related to the various definitions of at-issueness used in the literature. In order for the current paper to be maximally impactful, we remain agnostic about how to define at-issueness and investigate the Gradient Projection Principle using diagnostics that are based on assumptions made in the literature about how at-issue and not-at-issue content differ. We have updated section 1 of the paper to better foreshadow the assumptions that are made in the literature about how at-issue and not-at-issue content differ, the two diagnostics we use that are based on these assumptions, and our discussion in section 3.3 about open questions concerning at-issueness. Addressing those open questions, including how the diagnostics relate to the QUD-based definition of at-issueness, is outside the scope of this paper.

Referee: 2

Comments to the Author

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.

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 p. 30 – 31, there are many definitions of and diagnostics for at-issueness currently used in the literature. We have clarified this and have more clearly foreshadowed the employed at-issueness diagnostics and the assumptions from the literature they rely on at the end of section 1. Because the QUD-based definition of at-issueness is only one of several definitions of at-issueness, we only mention this definition in the comparison of at-issueness diagnostics.

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 (p. 29) – “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 discussion more prominent by more clearly foreshadowing it early in the paper (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.

XXX get inspired for title on walk?

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 clarify in two places in the paper why we use the terms “lexical content” and “projective content” – lexical contents instantiate projective contents, and the distinction is an important one. We previously first clarified the difference in what used to be Footnote 5, and have now moved the clarification into the main body of the text (p. 4). We also retained the second clarification in the Materials section of section 3.1.1.

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:

*"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?."*

*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 the remainder of the paper, the term 'lexical content' refers to the description of a particular event and the term 'projective content' refers to an abstract characterization of the projective content associated with an expression."*

In a footnote, we now mention the relation to the probabilistic pragmatics literature:

*"World knowledge as encoded in prior probabilities of different parts of the meaning space has been shown to affect interpretation in ways captured by models of language use that treat interpretation as Bayesian reasoning about an observed utterance (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 (10) into one line. As for (9), we think it's important to give the reader an example of each case, given that these expressions are much more heterogeneous and have thus left it as is. We also moved what was previously (15) – the list of lexical contents for Exp 1a – to Appendix A and now illustrate the three lexical contents that the projective content of "stop" is instantiated by. Analogously for Exp 1b, where we moved what was previously (18) to Appendix B.



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

We have now removed what used to be footnote 8. We originally included the footnote to acknowledge that piloting was conducted to establish the usefulness of different diagnostics. However, we agree that the amount of information included in the footnote was not useful. For reasons of space we can't elaborate on the details in the paper, so we have instead included a supplementary report containing details about the conducted pilot studies in the github repository.

For the same reason, we removed footnote 12.

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

We have clarified the contents of Figure 10 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...

This was an unfortunate typo from an earlier version. We have updated the title of the section as follows: "*Confirming the role of at-issueness in projectivity*"

Referee: 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.

We now introduce pragmatic approaches more clearly on p. XXX and have included a more thorough discussion of these approaches in section 5.

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?

XXX address?

JT proposes: These are great questions and it is our expectation that our finding about projection variability and the effect of at-issueness on projectivity will contribute to identifying empirically adequate analyses of projectivity for different classes of projective content. As we mention at the beginning of the paper, we intend to stay agnostic here about whether projectivity is triggered conventionally or derived conversationally/pragmatically. We have, however, UPDATED SECTION 4 TO ENGAGE MORE EXPLICITLY WITH THE QUESTIONS OF WHETHER CONTENT IS CONVENTIONALLY TRIGGERED OR PRAGMATICALLY/CONVERSATIONALLY DERIVED.

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?

XXX ?

JT proposes: RELATION EXISTS BECAUSE OF OUR DEFINITION OF AT-ISSUENESS. SO TO CLARIFY THIS WE NEED TO GIVE OUR DEFINITION OF AT-ISSUENESS. IF WE DON'T WANT TO GIVE THAT DEFINITION (SEE COMMENTS ABOVE ABOUT ONLY WANTING TO WORK WITH DIAGNOSTICS USED IN THE FIELD AND STAYING AGNOSTIC ABOUT RELATION OF THE DIAGNOSTICS TO DEFINITIONS), THEN WE CAN EITHER TRY TO JUST SAY, AS WE CURRENTLY DO, THAT OUR DEFINITION PREDICTS THAT OLE- = NOT-AT-ISSUE; OR WE CAN ALSO REMOVE EVERY MENTION OF OLE AND ALSO THE FINDING THAT OLE EXPRESSIONS ARE MORE LIKELY TO PROJECT.

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.

XXX ah good, include

JT: I DON'T HAVE KADMON HERE BUT WHAT I CAN SEE THROUGH AMAZON LOOKS VERY RELEVANT AND WE SHOULD HAVE A BRIEF DISCUSSION OF THE CLASSES SHE PROPOSES AND OF HER CONSIDERATIONS FOR CONVENTIONAL VS CONVERSATIONAL DERIVATION

OF PROJECTION IN OUR PAPER. AMARAL & CUMMINS ARE DISCUSSED IN THE PAPER BUT THEY ARE NOT RELEVANT FOR PROJECTION BUT FOR AT-ISSUENESS (FOR WHICH WE DISCUSS THEM); WE SHOULD ADD INFO THAT NOT ONLY DO WE TAKE THEIR RESULTS TO SUGGEST THAT AI IS GRADIENT PROPERTY, BUT ALSO ACKNOWLEDGE THAT THEY ALREADY SAY THAT TOO.

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.

XXX hm – too far beyond scope?

JT: I THINK WE SHOULD CLARIFY EARLY ON IN THE PAPER THAT WE CONSIDER PROJECTIVE CONTENT AS A WHOLE CLASS BECAUSE THE GOAL IS TO IDENTIFY WHY SUCH CONTENT PROJECTS. TO SEE WHETHER A UNIFIED STORY IS POSSIBLE, WE NEED TO SEE WHETHER PROJECTIVITY IS A UNIFIED PHENOMENON. THEN, IN SECTION 2.3 OR 4, WE SHOULD ENGAGE IN A DISCUSSION OF WHAT OUR FINDINGS SHOW REGARDING THE QUESTION OF WHETHER A UNIFIED ANALYSIS IS EMPIRICALLY ADEQUATE. PERHAPS SECTION 4, OR WHEREVER WE TALK ABOUT KADMON, IS THE BEST PLACE TO DO THAT.

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.

XXX be explicit

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.

XXX be clear in expressing uncertainty about diagnostics, as always in experiments

Typos:

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

Fixed.

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

Fixed.