

What is at-issueness? An experimental comparison of diagnostics

Abstract At-issueness is a key concept in theoretical semantics/pragmatics, but there is no consensus about how it is defined or diagnosed (e.g., [Tonhauser 2012](#); [Tonhauser et al. 2018](#); [Koev 2018](#)). We present experimental data investigating whether four widely used diagnostics for at-issueness yield consistent results. Our findings reveal significant differences across diagnostics, indicating they are not interchangeable. Since the diagnostics target distinct theoretical conceptions of at-issueness, these differences offer insight into their comparability.

1 Introduction

At-issueness is a key concept in theoretical semantics and pragmatics, distinguishing between at-issue propositions conveyed by an utterance, those contributing to its main point, and those that do not (e.g., [Karttunen & Peters 1979](#); [Horton & Hirst 1988](#); [Abbott 2000](#); [Faller 2003](#); [Potts 2005](#); [Tonhauser 2012](#)). Despite its importance, the concept lacks a unified definition. Instead, various theoretical notions ([Koev 2018](#); [Tonhauser et al. 2018](#)) and empirical diagnostics (e.g., [Tonhauser 2012](#)) have been proposed. This paper addresses the question whether four widely used diagnostics for at-issueness yield consistent results when testing the same stimuli. Our findings reveal significant differences across diagnostics, indicating they are not interchangeable. Since the diagnostics target distinct theoretical conceptions of at-issueness, these differences offer insight into their comparability.

The four diagnostics we tested are illustrated in (1–4) for sentence-medial non-restrictive relative clauses (NRRCs), which are usually taken to contribute non-at-issue content. As appositive content is generally taken to be not-at-issue, participants are expected to: Give low naturalness ratings under the QUD diagnostic (1) and the direct dissent diagnostic (3), not interpret the speaker to be asking about the content under the ‘asking-whether’ diagnostic in (2), will choose one of the *yes*-responses under the ‘yes, but’ diagnostic in (4).

- (1) QUD diagnostic (e.g., [Tonhauser 2012](#); [Chen 2024](#))
A: *What did Greg buy?*
B: *Greg, who bought a new car, is envied by his neighbor.*
Question to participants: How well does B’s response fit A’s question?
- (2) ‘asking whether’ diagnostic (e.g., [Tonhauser et al. 2018](#); [Solstad & Bott 2024](#))
Is Greg, who bought a new car, envied by his neighbor?
Question to participants: Is the speaker asking whether Greg bought a new car?
- (3) Direct dissent diagnostic (e.g., [Tonhauser 2012](#); [Syrett & Koev 2015](#))
A: *Greg, who bought a new car, is envied by his neighbor.*
B: *No, that’s not true, he didn’t buy a new car.*
Question to participants: How natural is B’s rejection of A’s utterance?
- (4) ‘yes, but’ diagnostic (e.g., [Xue & Onea 2011](#); [Destruel et al. 2015](#))
A: *Greg, who bought a new car, is envied by his neighbor.*
B: *Yes, but he didn’t buy a new car. /*
Yes, and he didn’t buy a new car. /
No, he didn’t buy a new car.

Task for participants: Choose the response that sounds best.

The diagnostics reflect different theoretical conceptions of at-issueness (Koev 2018), and they have led to different empirical results, discussed below.

1.1 QUD-based diagnostics

The diagnostics in (1) and (2) are based on the assumption that discourse is organized around addressing a question under discussion (QUD) (Roberts 1996; Ginzburg 1996), and that the at-issue content of an utterance addresses a QUD that is established by the preceding discourse (Amaral et al. 2007)¹. This notion, defined explicitly in Simons et al. 2010, is labeled Q(uestion)-at-issueness in Koev’s 2018 overview:

- (5) Q-at-issueness: (based on Simons et al. 2010: 26, Koev 2018: 2)
 A content m is Q-at-issue in a context c iff
- a. m is relevant to the QUD in c , and
 - b. p is appropriately conventionally marked relative to the QUD.

Here, m may be either a propositional content or a question meaning. Relevance to the QUD is defined as follows:

- (6) Relevance to the QUD in context c (based on Simons et al. 2010: 13)
- a. A proposition p is relevant the QUD iff it contextually entails in c a partial or complete answer to the QUD.
 - b. A question q is relevant to the QUD, iff it has an answer that is relevant to the QUD.

1.1.1 QUD-diagnostic

The QUD-diagnostic from Tonhauser 2012 operationalizes Q-at-issueness through naturalness judgments. It builds on two assumptions:

- i. An overt question explicitly introduces a QUD.²
- ii. An utterance is felicitous only if its at-issue content is relevant to the QUD (Amaral et al. 2007; Tonhauser 2012).

To test whether a given content m can be construed as Q-at-issue, participants are presented with a context that establishes a QUD via an overt question, followed by a response that includes m . For instance, (1) is used to diagnose the status of the content m of the appositive RC (Greg bought a car) conveyed by B’s utterance U , by presenting it as a response to a question Q that m is relevant to (What did Greg buy?), and asking a naturalness rating for U as a response to Q .

- (1) A: *What did Greg buy?*
 B: *Greg, who bought a new car, is envied by his neighbor.*
 Question to participants: How well does B’s response fit A’s question?

If m (Greg bought a car) is interpreted as addressing the QUD, the response should receive high naturalness ratings. However, responses like (1B) typically receive low ratings, suggesting that m is not at-issue, that is, even though m is relevant to Q and thereby satisfies the first part of the definition in (5a). The low naturalness should, therefore, reflect that m is not-at-issue due to the second part of the definition in (5b): The low ratings for (1B) support the claim that appositive RCs are not appropriately conventionally marked to contribute at-issue content.

¹ is this the right reference?

² add reference

1.1.2 Asking whether

Because the definition in (5) references the preceding context, Koev (2018) suggests that QUD-at-issueness is a backward-looking notion of at-issueness. However, overt questions may explicitly raise a QUD³, and thereby make a content Q-at-issue in the subsequent discourse. This is what is targeted by the ‘asking whether’ diagnostic in (2) (Tonhauser et al. 2018), based on the assumption that it is the at-issue content of interrogatives that partitions the context set, as opposed to their non-at-issue content (p.502).

- (2) *Is Greg, who bought a new car, envied by his neighbor?*

Question to participants: Is the speaker asking whether Greg bought a new car?

explain explain If participants respond "no," this suggests that the appositive content (Greg bought a new car) is not part of the at-issue content of the interrogative, providing evidence that it is not Q-at-issue. This diagnostic thus complements the QUD-diagnostic by probing the at-issueness of content from the perspective of explicitly raised questions rather than previously established ones.

1.2 Proposal at-issueness

The direct dissent diagnostic (3) and the ‘yes, but’ diagnostic (4) reflect the notion of P(roposal)-at-issueness, based on the assumption that at-issue content contributes to the main assertion of an utterance, which is taken to constitute a proposal to update the common ground.

- (7) P-at-issueness: (Koev 2013; 2018)

A proposition *p* is P-at-issue in a context *c* iff

- a. *p* is a proposal in *c* and
- b. *p* has not been accepted or rejected in *c*.

1.2.1 Direct dissent/assent

- (3) A: *Greg, who bought a new car, is envied by his neighbor.*

B: *No, that's not true, he didn't buy a new car.*

Question to participants: How natural is B's rejection of A's utterance?

1.2.2 yes, but

- (4) A: *Greg, who bought a new car, is envied by his neighbor.*

B: *Yes, but he didn't buy a new car. /*

Yes, and he didn't buy a new car. /

No, he didn't buy a new car.

Task for participants: Choose the response that sounds best.

1.3 Previous findings

Prior research has identified disagreements, potentially arising from diagnostic differences:

1.3.1 Medial appositives.

Based on impressionistic judgment data, Koev 2018 argues that medial appositives can be Q-at-issue, but not P-at-issue. An experimental study in Syrett & Koev 2015 found that sentence-medial

³ add reference

appositives are less at-issue than sentence-final ones using the direct dissent test, Drozdov 2024 found no difference with the ‘asking whether’ diagnostic.

To investigate how consistent the diagnostics are, we conducted four experiments measuring the at-issueness of the same contents across diagnostics.

section 1 also needs to say something about how we will compare the diagnostics. my current thinking is: do they distinguish contents that have been distinguished in prior research (in the same direction)? then we can discuss why they might not: because they tap into different properties? because of the way in which the diagnostic was implemented?

JT: Degen & Tonhauser 2025 will hopefully be accepted soon, we should include something like this figure because prior results but also to motivate the inclusion of be right and confirm in the stimuli of the four experiments

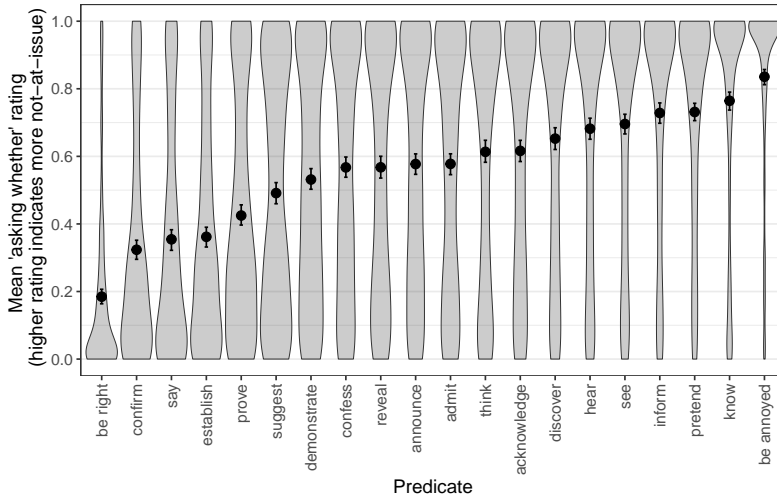


Figure 1: Mean ‘asking whether’ ratings for the contents of the clausal complements of 20 clause-embedding predicates, from Degen & Tonhauser 2025.

2 Experiments 1-4

To compare the results of at-issueness diagnostics, we conducted four experiments that each measured at-issueness with a different diagnostic, namely the QUD diagnostic (Exp. 1), the ‘asking whether’ diagnostic (Exp. 2), the direct dissent diagnostic (Exp. 3) and the ‘yes, but’ diagnostic (Exp. 4).⁴ To be able to compare the results of the diagnostics, the same seven contents, shown in (8), were investigated under the four diagnostics: the contents of sentence-medial and sentence-final NRRCs (8a)-(8b), as well as the contents of the clausal complements of *know*, *discover*, *confess*, *confirm* and *be right* (8c)-(8g). These seven contents were instantiated by the same items across the four experiments.

- (8) a. Content of sentence-medial NRRC
Lucy, who broke the plate, apologised. \rightsquigarrow Lucy broke the plate
 b. Content of sentence-final NRRC
The police found Jack, who saw the murder. \rightsquigarrow Jack saw the murder

⁴ The experiments, data and R code for generating the figures and analyses of the experiments reported in this paper are available at INSERT URL TO ANONYMOUS GITHUB REPO BEFORE SUBMISSION. All experiments were conducted with approval from the ethics review committee of [university name redacted for review].

- c. Content of the clausal complement of *know*
Ann knows that Raul cheated on his wife. \rightsquigarrow Raul cheated on his wife
- d. Content of the clausal complement of *discover*
Mary discovered that Denny ate the last cupcake. \rightsquigarrow Denny ate the last cupcake
- e. Content of the clausal complement of *be right*
Tom is right that Ann stole the money. \rightsquigarrow Ann stole the money
- f. Content of the clausal complement of *confirm*
Harry confirmed that Greg bought a new car. \rightsquigarrow Greg bought a new car
- g. Content of the clausal complement of *confess*
Lucy confessed that Dustin lost his key. \rightsquigarrow Dustin lost his keys

These seven contents were chosen because prior literature observed differences in at-issueness between two or more of these contents using a particular diagnostic for at-issueness. Specifically, as discussed in section §1, Syrett & Koev 2015 observed differences between sentence-medial and -final NRRCs using a variant of the direct dissent diagnostic, Tonhauser et al. 2018 observed differences between sentence-final NRRCs and the contents of the complements of *know*, *discover* and *confess* using the ‘asking whether’ diagnostic, and Degen & Tonhauser 2025 observed differences between *know*, *discover*, *confess*, *confirm* and *be right*, also using the ‘asking whether’ diagnostic. Thus, comparing these seven contents across the four diagnostics in Exps. 1-4 will allow us to assess whether the differences that emerge from one diagnostic also emerge from others.

In each experiment, participants read the stimuli and gave ratings corresponding to the diagnostics.

2.1 Methods

2.1.1 Participants

For each of the four experiments, we recruited unique 80 participants on Prolific. These participants had registered on the platform as living in the USA and as having English as their primary language. They had at least 50 previous submissions and an approval rate of at least 97%. Table 4 shows the age and gender distributions of the recruited participants.

	recruited	ages (mean age)	f/m/nb/dnd
Exp. 1 (QUD)	80	18-81 (43.8)	42/37/0/1
Exp. 2 (asking whether)	80	20-74 (38.5)	48/30/1/1
Exp. 3 (direct dissent)	80	18-77 (39.1)	50/28/1/1
Exp. 4 (yes, but)	80	19-67 (38.0)	48/30/2/0

Table 1: Information about the participants recruited in Exps. 1-4 (f = female, m = male, nb = nonbinary, dnd = did not disclose).

2.1.2 Materials and procedure

The four experiments measured the at-issueness of the seven contents in (8) with a different at-issueness diagnostic, namely the QUD diagnostic (Exp. 1), the ‘asking whether’ diagnostic (Exp. 2), the direct dissent diagnostic (Exp. 3) and the ‘yes, but’ diagnostic (Exp. 4). The examples in (9) illustrate how each diagnostic was implemented using the content of sentence-medial NRRCs (with the item ‘Lucy broke the plate’). In Exp. 1 (QUD diagnostic, (9a)), participants read a dialogue between two named speakers, where the first utters an interrogative sentence (the presumed QUD) that is about the content to be diagnosed and the second responds with a declarative sentence that contributes the content to be diagnosed. In Exp. 2 (‘asking whether’ diagnostic, (9b)), participants

read an interrogative sentence uttered by a named speaker, where the interrogative sentence contributes the content to be diagnosed. In Exp. 3 ('direct dissent' diagnostic, (9c)), participants read a dialogue between two named speakers, where the first utters a declarative sentence with the content to be diagnosed and the second directly dissents with the content to be diagnosed. Finally, in Exp. 4 ('yes, but' diagnostic, (9d)), participants read a dialogue between two named speakers where the first utters a declarative sentence that contributes the content to be diagnosed and the second responds with one of two indirect dissent variants (*yes, but...*, *yes, and...*) or with a direct dissent.

- (9) Implementations of the diagnostics in Exps. 1-4
- a. Exp. 1 (QUD diagnostic)
Nora: *What did Lucy break?*
Leo: *Lucy, who broke the plate, apologized.*
 - b. Exp. 2 ('asking whether' diagnostic)
Nora: *Did Lucy, who broke the plate, apologize?*
 - c. Exp. 3 ('direct dissent' diagnostic)
Nora: *Lucy, who broke the plate, apologized.*
Leo: *No, she didn't break the plate.*
 - d. Exp. 4 ('yes, but' diagnostic)
Nora: *Lucy, who broke the plate, apologized.*
Nina: *Yes, but she didn't break the plate.*
Yes, and she didn't break the plate.
No, she didn't break the plate.

As shown in Fig. 2, the response options in each of the four experiments differed depending on the diagnostic. In Exp. 1 (QUD diagnostic, panel (a)), participants were asked how well the response fits the question and they gave their response on a slider marked 'totally doesn't fit' on one end (coded 0) and 'totally fits' on the other end (coded as 1). In Exp. 2 ('asking whether' diagnostic, panel (b)), participants were asked whether the question is about the content to be diagnosed and they gave their response on a slider marked 'no' on one end (coded as 1) and 'yes' on the other (coded as 0). In Exp. 3 (direct dissent diagnostic, panel (c)), participants were asked how natural the direct dissent and participants gave their response on a slider marked 'totally unnatural' (coded as 0) on one end and 'totally natural' on the other (coded as 1). Finally, in Exp. 4 ('yes, but' diagnostic, panel (d)), participants were asked to choose the response that sounded best; the two indirect dissents were coded as 1 and the direct one as 0. Across the four experiments, the responses were coded as 0 or 1 in such a way that 0 meant that the content to be diagnosed was rated as at-issue and 1 meant that the content was rated as not-at-issue.

Each of the seven contents in (8) was instantiated by one of the seven items shown in (10) in each of the four experiments.

- (10)
- a. Jack saw the murder.
 - b. Raul cheated on his wife.
 - c. Ann stole the money.
 - d. Danny ate the last cupcake.
 - e. Lucy broke the plate.
 - f. Dustin lost his key.
 - g. Greg bought a new car.

Each experiment also included two control stimuli, which functioned as attention checks: one stimulus was expected to receive a response at one end of the slider (Exps. 1-3) or a 'no' response (Exp. 4); the other control stimulus was expected to receive a response at the other end of the

Nora: *What did Ann steal?*
Leo: *The manager reported Ann, who stole the money.*

How well does Leo's response fit Nora's question?

(a) Exp. 1: QUD diagnostic

Charlotte: *Did the manager report Ann, who stole the money?*

Is Charlotte asking whether Ann stole the money?

(b) Exp. 2: 'asking whether' diagnostic

Dawn: *The neighbor envies Greg, who bought a new car.*
Charlotte: *No, he didn't buy a new car.*

How natural is Charlotte's rejection of Dawn's utterance?

(c) Exp. 3: 'direct dissent' diagnostic

Vincent: *The boss scolded Dustin, who lost his key.*

Nina:

- ☐ *Yes, but he didn't lose his key.*
- ☐ *Yes, and he didn't lose his key.*
- ☐ *No, he didn't lose his key.*

Please choose the response by Nina that sounds best to you.

(d) Exp. 4: 'yes, but' diagnostic

Figure 2: Sample trials in (a) Exp. 1, (b) Exp. 2, (c) Exp. 3, and (d) Exp. 4.

slider (Exps. 1-3) or a 'yes' response (Exp. 4). See Supplement A for the control stimuli used in Exps. 1-4.

In each of the four experiments, each participant's set of items was generated by randomly combining each of the seven contents in (8) with a unique content in (10). Participants completed a total of 9 trials, namely 7 target trials and the same 2 control trials. Trial order was randomized.

After completing the experiment, participants filled out a short optional demographic survey. To encourage truthful responses, participants were told that they would be paid no matter what answers they gave in the survey.

2.1.3 Data exclusion

We excluded the data of participants who did not self-identify as native speakers of American English and of participants whose responses to either one of the two control trials was more than 2 sd away from the group mean (Exps. 1-3) or whose responses to either one of the two control

trials was wrong (Exp. 4). Table 2 shows how many participants were excluded in each experiment, the properties of the remaining participants, and the number of data points that entered into the analyses.

	exclusion criterion		remaining participants		data points
	language	fillers	ages (mean age)	f/m/nb/dnd	
Exp. 1 (QUD)	1	10	18-81 (41.1)	36/32/0/1	621
Exp. 2 (asking whether)	2	4	22-74 (38.7)	45/27/1/1	666
Exp. 3 (direct dissent)	2	7	18-77 (39.5)	44/25/1/1	639
Exp. 4 (yes, but)	4	4	19-67 (38.5)	43/27/2/0	648

Table 2: Information from Exps. 1-4 about the number of participants whose data was excluded based on their self-declared language (variety) and the fillers, about the remaining participants, and about the number of data points that entered into the analysis.

2.2 Results

Fig. 3 plots the results of the four experiments by the expression that is associated with the seven target contents: panel (a) shows the mean naturalness ratings in Exp. 1 (QUD diagnostic), panel (b) the mean ‘asking whether’ ratings in Exp. 2 (‘asking whether’ diagnostic), panel (c) the mean naturalness ratings in Exp. 3 (‘direct dissent’ diagnostic) and panel (d) the proportion of ‘no’ choices in Exp. 4 (‘yes, but’ diagnostic). We observe that the results of the four experiments differ in the range of the (mean or proportion of) ratings, that is, the difference between the largest and smallest ratings. **JT: redo, now that plots different** The range is largest in Exp. 2 (‘asking whether’ diagnostic), at .74 (.01 to .83) and smallest in Exp. 3 (‘direct dissent’ diagnostic), at .13 (.64 to .78). The results of Exp. 1 (QUD diagnostic, with a range of .27 (.51 to .77) and Exp. 4 (‘yes, but’ diagnostic), with a range of .46 (.5 to .96), fall in-between. This result suggests that the four diagnostics that were implemented in Exps. 1-4 differ in how much they differentiate between the seven contents investigated, with the ‘asking whether’ diagnostic showing the most differentiation and the ‘direct dissent’ and the QUD diagnostic showing the least differentiation.

We also observe that the results of the four experiments differ in the relative ratings that the seven contents received. The only two contents that received consistent relative ratings across all four experiments are the contents of the complement of *discover* and *confess*: The content of the complement of *discover* received higher ratings (at least numerically) across all four experiments than that of *confess*. There is no other pair of expressions for which that is the case. For instance, whereas the content of the complement of *confirm* received (numerically) higher ratings than that of *know* in Exps. 1 and 2, the opposite pattern is observed in Exps. 3 and 4. This difference between the results of the experiments is quantified in the Spearman rank correlations **JT: redo** in Table 3.⁵ The rank correlations are particularly low for for Exp. 1 compared to the other three experiments, which is due at least in part to the fact that the content of the complement of *be right* is the most not-at-issue content of the seven contents in Exp. 1, but among the least not-at-issue in the other three experiments. These results suggests that the four diagnostics as implemented in Exps. 1-4 interact differently with the seven contents investigated.

⁵ The Spearman rank correlation coefficient, a value between -1 and 1, is a nonparametric measure of rank correlation: the higher the coefficient, the more the relation between the two variables can be described using a monotonic function. If the coefficient is positive, the value of one variable tends to increase with an increase in the other. In the case of our experiments, a coefficient of 1 for two experiments would mean that there is a perfectly monotone increasing relation between the mean ratings of the seven contents in the two experiments: for any two contents *c1* and *c2*, if *c1* ranks below *c2* in one experiment (that is, the mean rating of *c1* is lower than that of *c2*), then that ranking is preserved in the other experiment.

	Exp. 1	Exp. 2	Exp. 3	Exp. 4
Exp. 1 (QUD diagnostic)		.11	-.29	-.18
Exp. 2 ('asking whether' diagnostic)			.64	.79
Exp. 3 ('direct dissent' diagnostic)				.79

Table 3: Spearman rank correlations between the results of Exps. 1-4.

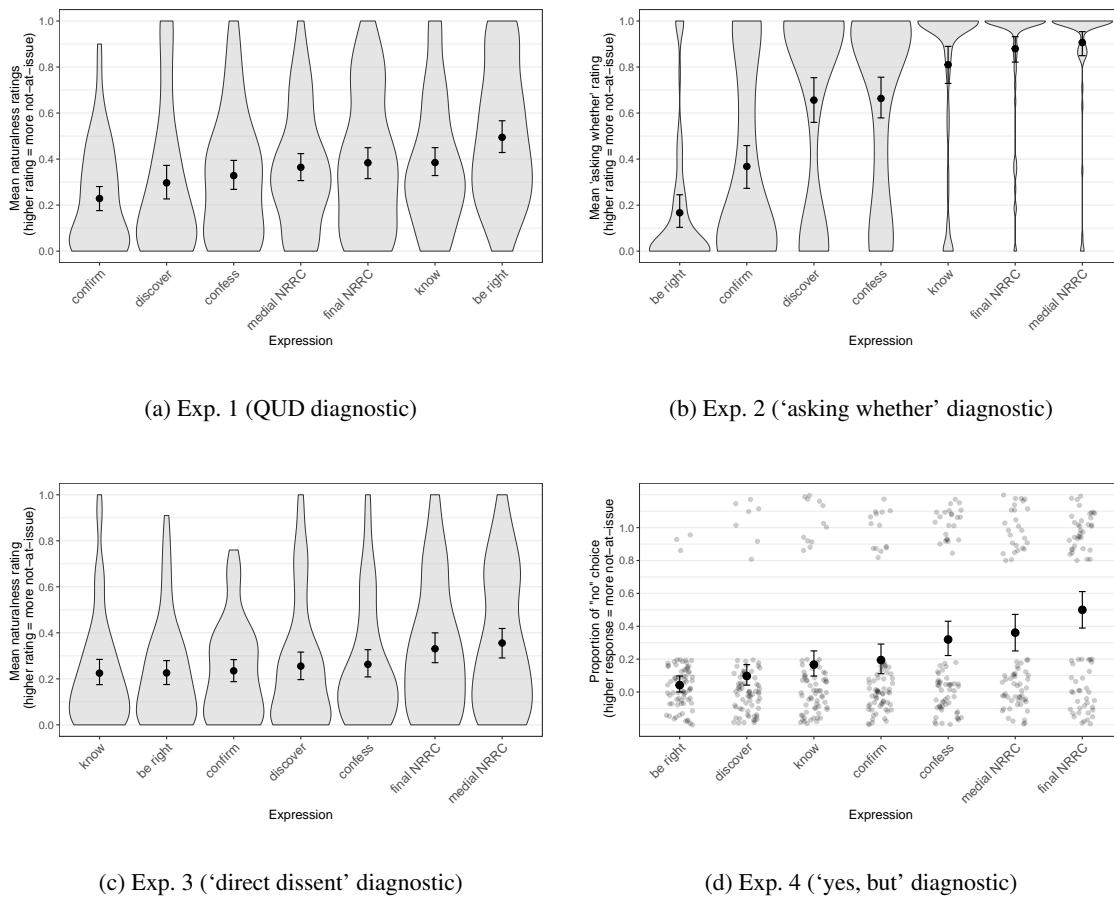


Figure 3: Results of Exps. 1–4. Panels (a)–(c) show the mean responses by expression for (a) Exp. 1 (QUD diagnostic), (b) Exp. 2 ('asking whether' diagnostic), and (c) Exp. 3 ('direct dissent' diagnostic); panel (d) shows the proportion of 'no' choices by expression in Exp. 4 ('yes, but' diagnostic). Error bars indicate 95% bootstrapped confidence intervals. Violin plots in panels (a)–(c) show the kernel probability density of individual participants' ratings. Gray dots in panel (d) represent individual participant responses ('no' vs. 'yes', jittered vertically and horizontally for legibility).

Finally, the results of the experiments differ with respect to whether distinctions between contents from prior literature are observed. Fig. 4 presents the results of posthoc pairwise comparisons of the estimated means/proportions for each content using the 'emmeans' package (Lenth 2023) in R (R Core Team 2016). The input to the pairwise comparisons were mixed-effects beta regression models (Exps. 1-3) or a mixed-effects logistic regression model (Exp. 4) with weakly informative

priors that were fit using the ‘brms’ package (Bürkner 2017). The models predicted the ratings⁶ from a fixed effect of expression (with treatment coding and ‘be right’ as reference level) and included random by-participant and by-item intercepts. The output of the pairwise comparison were 95% highest density intervals (HDIs) of estimated marginal mean differences between each of the expressions. We assume that two contents differ if their HDI does not include 0.⁷

Recall that Syrett & Koev 2015, using a variant of the direct dissent diagnostic, found that sentence-medial NRRCs are more not-at-issue than sentence-final ones. As shown in Figs. 3 and 4, no difference is observed in Exps. 1-3; in Exp. 4 (‘yes, but’ diagnostic), sentence-final NRRCs are more not-at-issue than sentence-medial ones. This result suggests that none of the diagnostics as implemented in Exps. 1-4 predict that sentence-medial NRRCs are more not-at-issue than sentence-final ones, contrary to Syrett & Koev’s result.

Recall also that Tonhauser et al. 2018 and Degen & Tonhauser 2025, using the ‘asking whether’ diagnostic, observed that the content of the clausal complement of *know* was more not-at-issue than that of *discover*, which in turn was more not-at-issue than that of *confess*, which in turn was more not-at-issue than that of *confirm*, which in turn was more not-at-issue than that of *be right*. As shown in Figs. 3 and 4, these differences are replicated in Exp. 2 (‘asking whether’ diagnostic), except for the contents of the complement of *confess* and *discover*. In Exp. 1 (QUD diagnostic), the content of the complement of *confirm* is less not-at-issue than that of *confess* and *know*, but the only other difference that is observed is that the content of the complement of *be right* is less at-issue than those of the other predicates – the direction of this difference is the opposite from that observed in prior literature and the other experiments. In Exp. 3..., Finally, in Exp. 4, the content of the complement of *be right* is more at-issue than those of *confirm*, *confess* and *know*, that of *confirm* is more at-issue than that of *confess*, and that of *discover* is more at-issue than that of *confess*. These results suggest that the diagnostics, as implemented in Exps. 1-4, differ in whether they predict differences in at-issueness between the contents of the complements of the five clause-embedding predicates included in the experiments.

⁶ To model the ratings in Exps. 1-3 using a beta regression, the ratings were first transformed from the interval [0,1] to the interval (0,1) using the method proposed in Smithson & Verkuilen 2006.

⁷ The full model outputs are available in the folder **where?** in the repository linked in footnote ??.

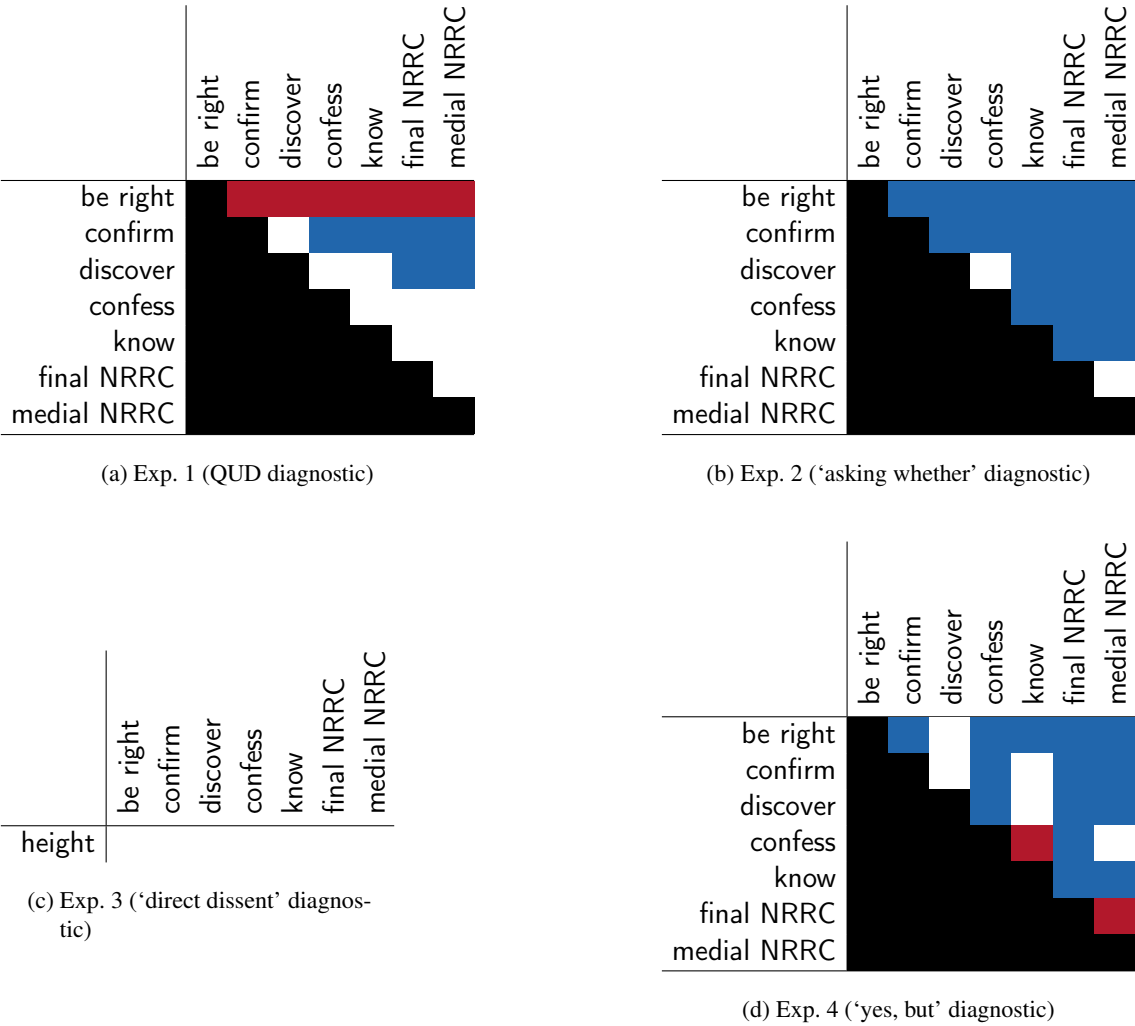


Figure 4: Pairwise differences between expressions, ordered from top to bottom and left to right by increasing mean in Exp. 2 ('asking whether' diagnostic). A white cell means that the 95% HDI of the pair of the row expression and the column expression includes 0, a red cell means that the 95% HDI does not include 0 and that the coefficient is positive (the row expression received a higher rating than the column expression), and a blue cell means that the 95% HDI does not include 0 and the coefficient is negative (the row expression received a lower rating than the column expression).

bla

2.3 Discussion

Exps. 1-4 were designed to compare the results of four different diagnostics of at-issueness that have been used in prior literature. The results of the experiments suggest that the diagnostics, in the particular way in which they were implemented in the experiments, differ on several dimensions. First, they differ in the extent to which they differentiate between the seven contents investigated, with the ‘asking whether’ diagnostic of Exp. 2 showing the most differentiation and the ‘direct dissent’ diagnostic of Exp. 3 showing the least. Second, the diagnostics as implemented differ in the way in which they interact with the seven contents. In particular, while the results of none of the experiments differentiated between sentence-medial and -final NRRCs, the results of Exp. 2 (‘asking whether’ diagnostic) distinguished between most of the contents of the complements of the five clause-embedding predicates, whereas the results of Exp. xx (‘’ diagnostic) distinguished between the least of these contents.

One of the most striking differences between the results of the experiments concerns the content of the complement of *be right* in Exp. 1 (QUD diagnostic) vs. the other three experiments. As shown in panel (a) of Fig. 3, participants gave relatively high naturalness ratings to responses like that in (11a), suggesting that they did not take the response to fit the question. As discussed above, the content of the complement of *be right* emerged as the most not-at-issue in Exp. 1.

- (11) a. Exp. 1 (QUD diagnostic) with *be right*
Nora: *What did Lucy break?*
Leo: *XX is right that Lucy broke the plate.*
- b. Exp. 2 (‘asking whether’ diagnostic)
Nora: *Is xx right that Lucy broke the plate?*
- c. Exp. 3 (‘direct dissent’ diagnostic)
Nora: *xx is right that Lucy broke the plate.*
Leo: *No, she didn’t break the plate.*
- d. Exp. 4 (‘yes, but’ diagnostic)
Nora: *xx is right that Lucy broke the plate.*
Nina: *Yes, but she didn’t break the plate.*
Yes, and she didn’t break the plate.
No, she didn’t break the plate.

Why might this be? We hypothesize that this is because *be right* signals that the question of whether Lucy broke the place must be salient in discourse. But no such discourse context is given and it is difficult to accommodate, so low naturalness ratings. This doesn’t happen in the other diagnostics, shown in (11b) to (11d).⁸ So, when using a particular diagnostic, one also has to worry about whether the expression might presuppose a particular context, if one wants to ask about how well an utterance fits the context.

As mentioned above, none of our experiments replicated the effect reported in Syrett & Koev 2015, that sentence-final NRRCs are more at-issue than sentence-medial ones. This includes Exp. 3, which implemented a variant of the ‘direct dissent’ diagnostic that was used in Syrett & Koev 2015. A difference between our Exp. 3 and Exp. 2 in Syrett & Koev 2015, aside from the stimuli, is that we asked for naturalness ratings and they implemented it as a forced choice: do you choose to dissent with the main clause content or the content of the NRRC. An interesting follow-up would be to see whether our stimuli with the forced choice can reproduce their result.

⁸ When *be right* is excluded, the Spearman rank correlations are:

	Exp. 1	Exp. 2	Exp. 3	Exp. 4
Exp. 1 (QUD diagnostic)		.77	.09	.31
Exp. 2 (‘asking whether’ diagnostic)			.66	.66
Exp. 3 (‘direct dissent’ diagnostic)				.77

One of the most noticable differences between the results of the experiments is that Exp. 2 has much larger range and differentiation among the contents. Why might this be? It is the only diagnostic where the expression/content occurs in a polar question. We hypothesize that the polar question embedding highlights at-issueness differences. Content is differently suitable to partition the context set, with some content really well able to do this and other content not, and other content inbetween. All the other diagnostics rely on anaphoric accessibility of the proposition or how well it can address a question in prior discourse.

To investigate whether it is really embedding polar question that creates this differentiation, it would be good to investigate a diagnostic that also embeds the expression/content in a polar question but uses a different assumption about at-issueness. This is what we do in the next section.

3 Experiments 5-6

Exp. 5 is just like Exp. 2 except that we investigate the contents of the complements of the 20 clause-embedding predicates investigated in [Tonhauser et al. 2018](#) and [Degen & Tonhauser 2025](#). Exp. 6 also investigates these contents, but uses a variant of the ‘direct dissent’ diagnostic in Exp. 3. As shown in (12b), the expression/content occurs in a polar question and the response directly dissents with the content.

- (12) a. Exp. 5 (‘asking whether’ diagnostic)
Nora: *Is xx right that Lucy broke the plate?*
 Question to participants: Is Nora asking whether Lucy broke the plate?
- b. Exp. 6 (‘direct dissent’ diagnostic)
Nora: *Is XX right that Lucy broke the plate?*
Leo: *No, she didn’t break the plate.*
 Question to participants: How natural is Leo’s response to Nora’s question?

Both experiments were run a while back already. They also included a projection block (data reported in [Hofmann et al. 2024](#)). The results of the at-issueness block have not yet been reported.

3.1 Methods

3.1.1 Participants

For both experiments, we recruited unique 80 participants on Prolific. These participants had registered on the platform as living in the USA and as having English as their primary language. They had at least 50 previous submissions and an approval rate of at least 97%. Table 4 shows the age and gender distributions of the recruited participants.

	recruited	ages (mean age)	f/m/nb/dnd
Exp. 1 (QUD)	80	18-81 (43.8)	42/37/0/1
Exp. 2 (asking whether)	80	20-74 (38.5)	48/30/1/1
Exp. 3 (direct dissent)	80	18-77 (39.1)	50/28/1/1
Exp. 4 (yes, but)	80	19-67 (38.0)	48/30/2/0

Table 4: Information about the participants recruited in Exps. 1-4 (f = female, m = male, nb = nonbinary, dnd = did not disclose).

3.1.2 Materials and procedure

3.1.3 Data exclusion

3.2 Results

3.3 Discussion

4 General discussion

5 Conclusion

The conclusion is the last numbered section, and any ensuing sections are unnumbered.

Abbreviations (if applicable)

ACC = accusative, DAT = dative, DEM = demonstrative, NOM = nominative, PL = plural, SG = singular

For the standard abbreviations to be used here, refer to the [Leipzig glossing rules](#).

Data availability/Supplementary files (if applicable)

The journal encourages authors to make all data associated with their submission openly available, according to the FAIR principles (Findable, Accessible, Interoperable, Reusable). More information can be found [here](#).

If data/supplementary files are to be associated with the accepted paper, one of the options below should be followed:

- i. upload the files to your chosen open repository and make note of the DOI that they will provide (most suitable for datasets or information that act as foundations to the research being published. This option makes the files more findable and more citable). We recommend an open repository such as [osf.io](#), which allows you to create a "project" under which you can upload relevant files (datasets, analysis scripts, experimental materials, etc.). The project will be associated with a unique DOI. You can then include in your manuscript a citation of the OSF entry and/or a link to the project page on OSF, to direct interested readers to the supplementary materials. During review, please be sure that the link to the repository is anonymized to maintain a fully double masked review process. Instructions for doing this on the OSF may be found [here](#). If you'd like to learn more about best practices for ensuring reproducibility, see [Laurinavichyute and Vasisht \(2021\)](#). Please contact us if you would like more information or advice about hosting your data on an open repository.
- ii. upload the files to the journal system during the submission process, as 'data files'. The journal will then host them as part of the publication and provide them with a DOI (most suitable for non-data files or very short pieces of information, although option 1 is also suitable for these if the author prefers).

In both cases, a 'Data availability' or 'Supplementary files' section must be added prior to the reference list that provides a title and very short summary of the files for each file. If option 1 was selected, you should also provide the DOI in this section. For example:

Supplementary file 1: Appendix. Scientific data related to the experiments. DOI: <https://doi.org/10.5334/gjgl.310.s1>

Ideally, supplementary files are also cited in the main text.

Please note that neither of the above two options will result in the files being typeset, so please ensure that they are in publishable format when you upload the accepted paper.

Ethics and consent (if applicable)

Research involving human subjects, human material, or human data, must have been performed in accordance with the Declaration of Helsinki. Studies must have been approved by an appropriate ethics committee and the authors should include a statement in the article text detailing this approval, including the name of the ethics committee and reference number of the approval, or mention any exemptions to ethical approval that apply to their research. The identity of research subjects should be anonymised whenever possible. For research involving human subjects, informed consent to participate in the study must be obtained from participants (or their legal guardian).

Funding information (if applicable)

Should the research have received a funding grant then the grant provider and grant number should be detailed.

Acknowledgements (optional)

The authors wish to thank Martin Haspelmath for providing the generic style sheet for linguistics, and Kai von Fintel for giving permission to use and modify the *Semantics & Pragmatics* Latex template, bibliography style, and document class.

Competing interests (required)

If any of the authors have any competing interests then these must be declared. Guidelines for competing interests can be found [here](#). If there are no competing interests to declare then the following statement should be present: ‘The author(s) has/have no competing interests to declare’.

Authors’ contributions (optional)

A sentence or a short paragraph detailing the roles that each author held to contribute to the authorship of the submission. Individuals listed must fit within the definition of an author, as per our [Author Guidelines](#).

References

- Abbott, Barbara. 2000. Presuppositions as nonassertions. *Journal of Pragmatics* 32(10). 1419–1437. [https://doi.org/https://doi.org/10.1016/S0378-2166\(99\)00108-3](https://doi.org/https://doi.org/10.1016/S0378-2166(99)00108-3). Publisher: Elsevier
- Amaral, Patricia & Roberts, Craige & Smith, E Allyn. 2007. Review of the logic of conventional implicatures by Chris Potts. *Linguistics and Philosophy* 30. 707–749. Publisher: Springer.
- AnderBois, Scott & Brasoveanu, Adrian & Henderson, Robert. 2010. Crossing the appositive/at-issue meaning boundary. In *Semantics and linguistic theory*, vol. 20. 328–346. <http://journals.linguisticsociety.org/proceedings/index.php/SALT/article/view/2551>.
- Ariel, Mira. 1988. Referring and accessibility. *Journal of linguistics* 24(1). 65–87. Publisher: Cambridge University Press.
- Beaver, David. 2010. Have you noticed that your belly button lint colour is related to the colour of your clothing. *Presuppositions and Discourse: Essays Offered to Hans Kamp*.
- Beaver, David & Roberts, Craige & Simons, Mandy & Tonhauser, Judith. 2009. Investigating properties of projective meaning. Ms.
- Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1). 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Chen, Yuqiu. 2024. Presuppositions at the Semantics-Pragmatics Interface <https://ediss.uni-goettingen.de/handle/11858/15164>.
- Degen, Judith & Tonhauser, Judith. 2021. Prior beliefs modulate projection. *Open Mind* 5(1). 59–70. https://doi.org/10.1162/opmi_a_00042
- Degen, Judith & Tonhauser, Judith. 2022. Are there factive predicates? An empirical investigation. *Language* 98(3). 552–591.
- Degen, Judith & Tonhauser, Judith. 2025. Projection inferences: On the relation between prior beliefs, at-issueness, and lexical meaning. Manuscript under review.

- Destruel, Emilie & Onea, Edgar & Velleman, Daniel & Bumford, Dylan & Beaver, David. 2015. A cross-linguistic study of the non-at-issueness of exhaustive inferences. In Schwarz, Florian (ed.), *Experimental approaches to presupposition*, 135–156. Springer. <https://doi.org/10.1007/978-3-319-07980-6>
- Drozdo, Katharina. 2024. Projection and At-Issueness of Nonrestrictive Relative Clauses.
- Faller, M. T. 2003. Semantics and pragmatics of evidentials in Cuzco Quechua (Peru) <https://elibrary.ru/item.asp?id=5280271>.
- Ginzburg, Jonathan. 1996. Interrogatives: Questions, facts and dialogue. *The handbook of contemporary semantic theory* 5(18). 359–423. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=595cd12adcf5e27f900c47778695412e89711481>. Publisher: Citeseer.
- Ginzburg, Jonathan & Sag, Ivan A. 2000. *Interrogative investigations: The Form, Meaning, and Use of English Interrogatives* (CSLI lecture notes 123). Stanford, CA: CSLI Publications.
- Hofmann, Lisa & de Marneffe, Marie-Catherine & Tonhauser, Judith. 2024. Projection variation: Is the family of sentences really a family? *Sinn und Bedeutung* 28. 422–440.
- Horton, Diane & Hirst, Graeme. 1988. Presuppositions as beliefs. In *Coling Budapest 1988 Volume 1: International Conference on Computational Linguistics*. <https://aclanthology.org/C88-1052.pdf>.
- Hothorn, Torsten & Bretz, Frank & Westfall, Peter. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50(3). 346–363.
- Hunter, Julie & Abrusán, Márta. 2017. Rhetorical Structure and QUDs. In Otake, Mihoko & Kurahashi, Setsuya & Ota, Yuiko & Satoh, Ken & Bekki, Daisuke (eds.), *New Frontiers in Artificial Intelligence*, vol. 10091, 41–57. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-50953-2_4. http://link.springer.com/10.1007/978-3-319-50953-2_4. Series Title: Lecture Notes in Computer Science
- Jasinskaja, Katja. 2016. Not at issue any more. *Ms. University of Cologne* https://dslc.phil-fak.uni-koeln.de/sites/dslc/katja_files/jasinskaja_any_more.pdf.
- Karttunen, Lauri & Peters, Stanley. 1979. Conventional implicature. In Oh, Choon-Kyu & Dinneen, David A. (eds.), *Presuppositions* (Syntax and Semantics Vol.11), 1–56. New York: Academic Press.
- Kiparsky, Paul & Kiparsky, Carol. 1968. *Fact*. Linguistics Club, Indiana University.
- Koev, Todor. 2018. Notions of at-issueness. *Language and Linguistics Compass* 12. e12306. <https://doi.org/https://doi.org/10.1111/lnc3.12306>
- Koev, Todor K. 2013. *Apposition and the structure of discourse*. Rutgers The State University of New Jersey, School of Graduate Studies. <https://search.proquest.com/openview/3686668834d9802d690c6e574100c8e0/1?pq-origsite=gscholar&cbl=18750>.
- Korotkova, Natasha. 2020. Evidential meaning and (not-)at-issueness. *Semantics and Pragmatics* 13. article 4.
- Lenth, Russell V. 2023. *emmeans: Estimated marginal means, aka least-squares means*. <https://CRAN.R-project.org/package=emmeans>. R package version 1.8.8.
- Murray, Sarah E. 2014. Varieties of update. *Semantics and Pragmatics* 7(2). 1–53. <https://doi.org/10.3765/sp.7.2>
- Onea, Edgar. 2016. *Potential questions at the semantics-pragmatics interface*, vol. 33. Brill. [https://books.google.com/books?hl=en&lr=&id=OyYiDAAQBAJ&oi=fnd&pg=PP5&dq=Onea,+E.+\(2016\).+Potential+questions+at+the+semantics-pragmatics+interface&ots=DeTvzGNNZW&sig=8ZCgDj7FO-M73aRoyy3TLi08pZA](https://books.google.com/books?hl=en&lr=&id=OyYiDAAQBAJ&oi=fnd&pg=PP5&dq=Onea,+E.+(2016).+Potential+questions+at+the+semantics-pragmatics+interface&ots=DeTvzGNNZW&sig=8ZCgDj7FO-M73aRoyy3TLi08pZA).
- Potts, Christopher. 2005. *The Logic of Conventional Implicatures*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199273829.001.0001>
- Potts, Christopher. 2007. Into the Conventional-Implicature Dimension. *Philosophy Compass* 4. 665–679.

- R Core Team. 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. <https://www.R-project.org/>.
- Riester, Arndt. 2019. Constructing QUD trees. In *Questions in discourse*, 164–193. Brill. <https://brill.com/downloadpdf/display/book/edcoll/9789004378322/BP000006.pdf>.
- Roberts, Craige. 1996. Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. In Yoon, Jae Hak & Kathol, Andreas (eds.), *Ohio State University Working Papers in Linguistics*, vol. 49. The Ohio State University, Department of Linguistics.
- Roberts, Craige. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics* 5. 1–69.
- Roberts, Craige & Simons, Mandy & Beaver, David & Tonhauser, Judith. 2009. Presupposition, conventional implicature and beyond: A unified account of projection. In *New Directions in the Theory of Presupposition, ESSLLI workshop, Toulouse*.
- Simons, Mandy & Tonhauser, Judith & Beaver, David & Roberts, Craige. 2010. What projects and why. In *Semantics and linguistic theory*, vol. 20. 309–327.
- Smithson, Michael & Verkuilen, Jay. 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods* 11. 54–71.
- Snider, Todd. 2018. Distinguishing at-issueness from anaphoric potential: A case study of appositives. In *West Coast Conference on Formal Linguistics (WCCFL)*, vol. 35. 374–381.
- Solstad, Torggrim & Bott, Oliver. 2024. Cataphoric resolution of projective content: The case of occasion verbs. *Semantics and Pragmatics* 17. 11:1–66. <https://doi.org/10.3765/sp.17.11>. <https://semprag.org/index.php/sp/article/view/sp.17.11>
- Stalnaker, Robert C. 1999. *Context and content: Essays on intentionality in speech and thought*. Clarendon Press.
- Stalnaker, Robert C. 2002. Common ground. *Linguistics and Philosophy* 25. 701–721. <https://www.jstor.org/stable/pdf/25001871.pdf>.
- Syrett, Kristen & Koev, Todor. 2015. Experimental evidence for the truth conditional contribution and shifting information status of appositives. *Journal of Semantics* 32(3). 525–577. <https://doi.org/10.1093/jos/ffu007>. Publisher: Oxford University Press
- Tonhauser, Judith. 2012. Diagnosing (not-) at-issue content. *Proceedings of Semantics of Under-represented Languages of the Americas (SULA)* 6. 239–254.
- Tonhauser, Judith. 2016. Prosodic cues to presupposition projection. In *Semantics and Linguistic Theory*, vol. 26. 934–960.
- Tonhauser, Judith & Beaver, David I. & Degen, Judith. 2018. How projective is projective content? Gradience in projectivity and at-issueness. *Journal of Semantics* 35(3). 495–542. <https://doi.org/10.1093/jos/ffy007>
- Xue, Jingyang & Onea, Edgar. 2011. Correlation between presupposition projection and at-issueness: An empirical study. In *Proceedings of the ESSLLI 2011 workshop on projective meaning*. 171–184.

Supplements

A Control stimuli in Exps. 1-4

The examples in (1)–(4) provide the two control stimuli used in each of Exps. 1-4. For the a.-examples, participants were expected to give a ‘totally fits’ response (Exp. 1), a ‘yes’ response (Exp. 2), a ‘totally natural’ response (Exp. 3), and a ‘no’ response (Exp. 4); for the b.-examples, the opposite response was expected. The numbers after each example identify the mean ratings (Exps. 1-3) or the proportion of ‘no’ responses (Exp. 4) after excluding participants who did not

self-identify as native speakers of American English (but before excluding participants on the basis of these controls), showing that the control stimuli worked as intended.

- (1) Control stimuli in Exp. 1 (QUD diagnostic)
 - a. Mary: Which course did Ava take?
John: She took the French course. (.97)
 - b. Jennifer: What does Betsy have?
Robert: She loves dancing salsa. (.07)
- (2) Control stimuli in Exp. 2 ('asking whether' diagnostic)
 - a. Mary: Did Arthur take a French course?
Question to participants: Is Mary asking whether Arthur took a French course? (.96)
 - b. Robert: Does Betsy have a cat?
Question to participants: Is Robert asking whether Betsy loves apples? (.02)
- (3) Control stimuli in Exp. 3 ('direct dissent' diagnostic)
 - a. Mary: Arthur took a French course.
Lily: No, he took a Spanish course. (.87)
 - b. Robert: Betsy has a cat.
Maximilian: No, she doesn't like apples. (.05)
- (4) Control stimuli in Exp. 4 ('yes, but' diagnostic)
 - a. Mary: Arthur took a French course.
Lily: Yes, but Lisa loves cats. / Yes, and he didn't take a French course. / No, he didn't take a French course. (.95)
 - b. Robert: Betsy has a cat.
Maximilian: Yes, but she is good at math. / Yes, and she loves it so much. / No, she doesn't like apples. (0)