What is at-issueness? An experimental comparison of diagnostics

Word count (including references, excluding abstract and supplements): 10,338

Abstract   At-issueness is a central concept in theoretical semantics and pragmatics, but there is no consensus about how it should be defined or diagnosed (e.g., Tonhauser 2012; AnderBois et al. 2015; Snider 2017a; b; 2018; Tonhauser et al. 2018; Koev 2018; Faller 2019; Korotkova 2020). We report the results of six experiments that investigate whether five diagnostics for at-issueness yield consistent results when applied to the same set of contents in English. Diagnostics are compared based on whether they reproduce contrasts reported in prior work, in particular (i) positional effects for appositive NRRCs found in Syrett & Koev (2015) and (ii) fine-grained differences among the complements of clause-embedding predicates reported in Degen & Tonhauser (2025). The results reveal substantial differences between diagnostics, indicating that they are not interchangeable. None of our experiments observed the positional effect for NRRCs (i), and only the diagnostics that embed the tested contents in polar questions come close to observing the predicate-level differences (ii). Our generalization, that diagnostics measuring the at-issueness of contents embedded in questions yield greater by-content differentiation than diagnostics embedding them in assertions, indicates that the speech act in which a content is presented plays a key role in affecting diagnostic outcomes. We argue that this effect of speech act on at-issueness ratings is due to how questions and assertions relate at-issue and not-at-issue content to speaker commitments, and cannot be explained by different notions of at-issueness assumed in the literature.

Keywords: at-issueness, diagnostics, experimental pragmatics

1   Introduction

At-issueness is a key concept in theoretical semantics and pragmatics,[1] used in the analysis of various phenomena, including presupposition, conventional implicature, evidentials, expressives, and co-speech gestures (e.g., Karttunen & Peters 1979; Horton & Hirst 1988; Abbott 2000; Faller 2002; 2019; Potts 2005; Simons et al. 2010; Lee 2011; Tonhauser et al. 2018; Esipova 2019; Korotkova 2020; Barnes & Ebert 2023; Chen 2024; Scontras & Tonhauser 2025). It is generally understood as distinguishing the main point of an utterance (at-issue content) from propositions conveying background information (not-at-issue content), but there is no consensus how at-issueness should be diagnosed, and competing definitions reflect different assumptions about its underlying nature (e.g., Tonhauser 2012; Snider 2017a; b; 2018; Tonhauser et al. 2018; Koev 2018; Faller 2019; Esipova 2019; Korotkova 2020). Consequently, empirical claims about whether a given expression contributes at-issue or not-at-issue content may be relative to specific diagnostics, and different diagnostics may not target the same underlying phenomenon.

Four commonly used at-issueness diagnostics are illustrated in (1–4) for sentence-medial appositive non-restrictive relative clauses (NRRCs), which are typically taken to contribute not-at-issue content (Potts 2005). Accordingly, participants are expected to: give low question-answer-match ratings under the QUD diagnostic (1) (Amaral et al. 2007; Lee 2011; Tonhauser 2012; Chen 2024); judge that the speaker is not asking about the NRRC content under the 'asking-whether'

---

[1] Potts (2015) credits the term to Bill Ladusaw, "who began using it informally in his UCSC undergraduate classes in 1985" (p. 2).

diagnostic (2) (Tonhauser et al. 2018; Solstad & Bott 2024; Degen & Tonhauser 2025); give low naturalness ratings under the direct-dissent diagnostic (3), (Faller 2002; 2006; Papafragou 2006; Amaral et al. 2007; Murray 2010; 2014; AnderBois et al. 2010; 2015; Tonhauser 2012; Syrett & Koev 2015); and prefer a *yes*-response when denying the NRRC content under the 'yes, but' diagnostic (4), (Xue & Onea 2011; Cummins et al. 2013; Destruel et al. 2015).

(1)   QUD diagnostic

   Nora: *What did Greg buy?*
   Leo: *Greg, who bought a new car, is envied by his neighbor.*

   Question to participants: How well does Leo's response fit Nora's question?

(2)   'asking whether' diagnostic

   Nora: *Is Greg, who bought a new car, envied by his neighbor?*

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

(3)   Direct-dissent diagnostic

   Nora: *Greg, who bought a new car, is envied by his neighbor.*
   Leo: *No, that's not true, he didn't buy a new car.*

   Question to participants: How natural is Leo's rejection of Nora's utterance?

(4)   'yes, but' diagnostic

   Nora: *Greg, who bought a new car, is envied by his neighbor.*
   Leo: *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 assumptions underlying these diagnostics can be organized around three properties of at-issue content identified in Tonhauser (2012) as motivating diagnostic strategies:
   (i) at-issue content addresses the question under discussion (see also Amaral et al. 2007; Simons et al. 2010),
   (ii) the at-issue content of questions determines the relevant set of alternatives, and
   (iii) at-issue content can be directly assented or dissented with (see also Shanon 1976; Groenendijk & Stokhof 1984; Faller 2002; 2006; Murray 2010).

The QUD diagnostic (1) targets property (i), assuming that discourse is structured around addressing a *Question Under Discussion* (QUD) (Roberts 2012; Ginzburg 1996) and that at-issue content is the part of an utterance intended to address that question (Amaral et al. 2007; Simons et al. 2010). Assuming that only at-issue content can felicitously address an established QUD, the diagnostic presents the target content in a response to a question that makes it relevant. Question-answer match ratings are expected to be high when the content can be construed as at-issue and low otherwise.

   The 'asking whether' diagnostic (2) targets property (ii), testing whether the target content is interpreted as determining the question alternatives. Participants judge whether a speaker, who asks a polar question containing the target content, is asking whether that content is true. Such 'asking whether' judgments are expected to be high when the target content is at-issue and low otherwise.

   The direct-dissent diagnostic (3) and the 'yes, but' diagnostic (4) target property (iii), assuming that only at-issue content can be directly affirmed or denied, whereas denying not-at-issue content requires more indirect discourse moves (Shanon 1976; Faller 2002; 2006; Potts 2005; Papafragou 2006; Amaral et al. 2007). Accordingly, the direct-dissent diagnostic assumes that directly dissenting with the target content is judged natural if the content is at-issue and unnatural otherwise. The 'yes, but' diagnostic tests whether a target content can be denied while responding *yes* to

the assertion as a whole, under the assumption that *yes, but* responses provide an indirect denial suitable for dissenting with not-at-issue content (Xue & Onea 2011).

Although these diagnostics are widely used, there is no consensus on whether the properties they target correspond to a single underlying notion of at-issueness. Tonhauser (2012) argues that diagnostics targeting properties (i)–(iii) can be understood as probing a common underlying notion of at-issueness, adopting the QUD-based definition in (5).

(5)   QUD at-issueness: (based on Simons et al. 2010, p. 323)
       A content *m* is at-issue in a context *c* iff

       a.   *m* is relevant[2] to the QUD in *c*, and
       b.   the speaker has a recognizable intention to address the QUD via *m*.

While this definition aligns naturally with property (i), it is less clear whether all diagnostics, in particular those targeting property (iii), track this QUD-based notion. Evidence for empirical differences between diagnostics comes from appositive NRRCs: although they are widely taken to contribute not-at-issue content, a forced-choice continuation study by Syrett & Koev (2015) found that sentence-final appositive NRRCs pattern as more at-issue than medial ones, under a version of the direct-dissent diagnostic, which targets property (iii). Snider (2017b) argues that diagnostic targeting property (iii) are sensitive to this medial-final distinction, but diagnostics targeting properties (i) or (ii) are not. Similar diagnostic differences have been reported for sentence-final slifting parentheticals (Koev 2018) and evidential inferences (Korotkova 2020).

These contrasts have motivated arguments that at-issueness is not a uniform notion (Koev 2018), but that property (iii) diagnostics are linked to an assertion-based notion of at-issueness, grounded in analyzing assertions as proposals to update the common ground (Stalnaker 1978; Clark & Schaefer 1989; Ginzburg 1995; Farkas & Bruce 2010). On this view, the at-issue content of an assertion is the proposition proposed to be added to the common ground, whereas not-at-issue content is presupposed (already entailed in the common ground; Stalnaker 1973; 2002), or added without being proposed (Murray 2010; 2014; AnderBois et al. 2010; 2015). This notion is defined as follows:[3]

(6)   Proposal at-issueness: (Koev 2013, pp. 50–51)
       A proposition *p* is 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*.

Although empirical differences between at-issueness diagnostics have been reported across a range of constructions, they have not, to date, been evaluated in a systematic experimental comparison that applies multiple diagnostics to the same contents. At the same time, the diagnostic differences that have been reported can be alternatively interpreted in ways that do not require positing multiple underlying notions of at-issueness. One line of work (Snider 2017b; a; 2018; Korotkova 2020) argues that diagnostics targeting properties (i) and (ii) genuinely diagnose QUD at-issueness,[4]

---

[2] *m* may be a proposition or a question denotation, and relevance to the QUD can be defined in terms of contextual entailment of a partial or complete answer (Roberts 2012; Simons et al. 2010), or based on inferences about the probability of QUD-answers (Büring 2003; Beaver & Clark 2008).

[3] Further definitions of at-issueness proposed in the literature include analyses in terms of restrictive vs. non-restrictive modification (Esipova 2019; 2021), and coherence-based discourse structure (Hunter & Asher 2016; Jasinskaja 2016). Clarifying the theoretical relations among the various definitions of at-issueness proposed in the literature remains an important question for future research.

[4] Property (ii) can be seen as a forward-looking counterpart of property (i) (Snider 2017b): while property (i) concerns addressing an established QUD, property (ii) concerns determining the alternatives a question introduces. Under standard assumptions that explicit questions introduce or update the QUD (Roberts 2012; Ginzburg 1996), property (ii) aligns with the QUD-based definition, but it is incompatible with the proposal-based definition, which applies only to assertions.

whereas diagnostics targeting property (iii) do not diagnose at-issueness at all, but instead track the availability of the target content for propositional anaphora, such as response particles (*yes/no*) or propositional pro-forms (e.g., *that* in *that's not true*).

Further, theoretical proposals argue that QUD-based and proposal-based definitions are not genuinely distinct, but offer different analytical perspectives on the same underlying phenomenon: AnderBois et al. (2015), building on Farkas & Bruce (2010), argue that proposing content for the common ground is intrinsically linked to managing the QUD, and Faller (2019) analyzes assent and dissent as discourse moves that address the QUD. From this perspective, diagnostic divergence cannot be taken as direct evidence for multiple notions of at-issueness, but instead motivates careful analysis of how specific properties of diagnostics (e.g., whether they include propositional anaphora) may lead to diverging behavior independently of any differences in at-issueness.

In sum, there is currently no agreed-upon definition of at-issueness, and existing diagnostics may or may not be probing the same underlying phenomenon. This raises two central questions: whether reported empirical differences between diagnostics can be confirmed experimentally, and whether competing definitions of at-issueness target the same underlying phenomenon. This paper takes a first step toward addressing these issues through a systematic experimental comparison of diagnostics designed to address the following research question:

- Do the various diagnostics yield consistent results when applied to the same contents?

Initial experimental research on these questions has yielded mixed answers. First, as noted above, Syrett & Koev (2015) report a positional effect for appositive NRRCs under a direct-dissent diagnostic. Because this medial-final distinction has been argued to be specific to assent and dissent-based (property iii) diagnostics, a key question is whether the same contrast emerges under other diagnostics when applied to identical materials.

Beyond appositives, Tonhauser et al. 2018 compared two at-issueness diagnostics across a range of English expressions, including sentence-medial appositives and the complements of the clause-embedding predicates *discover, know,* and *annoyed* (p. 526ff). They found that the question-based 'asking whether' diagnostic yielded lower ratings and smaller differentiation between contents than the second at-issueness diagnostic they used, the 'are you sure?' diagnostic.[5] Despite these differences, ratings from the two diagnostics were correlated, suggesting that they may nonetheless be sensitive to a shared underlying phenomenon. The possibility that diagnostics may differ in how strongly they differentiate among contents raises another key question: whether different diagnostics systematically reproduce fine-grained by-content differences.

Building on these observations, we compare at-issueness diagnostics by asking whether they reproduce by-content differences reported in the literature. In addition to appositive NRRCs, we examine the complements of clause-embedding predicates, for which Degen & Tonhauser (2025) report fine-grained differences under the 'asking whether' diagnostic. Figure 1 shows the mean 'asking whether' ratings for the 20 English predicates used in their study.

This paper addresses the research question, whether the diagnostics yield consistent results when applied to the same target contents, through a systematic comparison of at-issueness diagnostics across six experiments. Our study compares the diagnostics by examining whether they reproduce contrasts between contents reported in prior work. Specifically, we ask whether different diagnostics converge on (i) the medial versus sentence-final distinction for appositive NRRCs and (ii) fine-grained differences among the contents of the complements of clause-embedding predicates.

---

[5] The 'are you sure?' diagnostic presents participants with dialogues like (i) (Tonhauser et al. 2018, p. 519) and collects ratings about whether Fred answered Carla's question (see also the 'really' test of Shanon 1976). This diagnostic cannot clearly be aligned with the three properties of at-issue content.

(i)   Fred: Martha's new car, a BMW, was expensive.
        Carla: Are you sure?
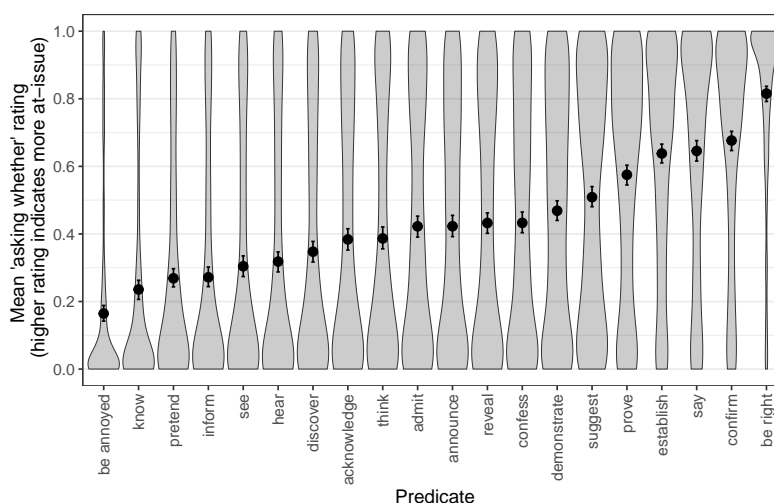   Fred: Yes, I am sure that Martha's new car is a BMW.

Figure 1: Mean 'asking whether' ratings for the contents of the clausal complements of 20
clause-embedding predicates, from Degen & Tonhauser 2025. Error bars indicate 95%
bootstrapped confidence intervals. Violin plots indicate the distribution of individual ratings.

Experiments 1–4 apply the four diagnostics to a set of seven contents in English, including
medial and final appositive NRRCs and the complements of selected clause-embedding predicates.
While none of these four experiments observed the positional effect for NRRCs found by Syrett
& Koev (2015), only Exp. 2 ('asking whether') comes close to observing the differences between
the complements of clause-embedding predicates observed in Degen & Tonhauser (2025), sug-
gesting that diagnostics do vary in whether they detect at-issueness differences between particular
expressions. The 'asking whether' diagnostic, as implemented in Exp. 2, which showed the greatest
differentiation between contents, is the only diagnostic which asked participants are asked to judge
the intentions of the speaker, and it is the only diagnostics where the target content is embedded in
a polar question.

To test whether the greater differentiation under the 'asking whether' diagnostic was due to the
polar question embedding or from the response task itself, we compared the 'asking whether'
diagnostic (Exp. 5) with a direct-response diagnostic (Amaral et al. 2007; Tonhauser 2012) that
also embeds content in a polar question but elicits acceptability judgments for a direct response
(Exp. 6), using the complement clauses of 20 clause-embedding predicates from Degen & Tonhauser
(2025). The two experiments yielded highly correlated results, indicating that the observed high
differentiation in Exp. 2 was driven by the polar question embedding, not the response task.

Our results, therefore, provide evidence that the speech act used to present the target content
plays an important role: embedding the target content in polar questions yield results with greater
by-content differentiation than embedding it in assertions.

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).[6] Each experiment tested the same set of contents associated with the seven types of

---

[6] The experiments, data and R code for generating the figures and analyses of the experiments reported in this paper are
available at https://anonymous.4open.science/r/at-issueness-diagnostics-232C/.

expressions in (7): sentence-medial and sentence-final NRRCs (7a–b), as well as the contents of the clausal complements of *be right, discover, confess, confirm, discover,* and *know* (7c–g).

(7) a. Content of sentence-medial NRRC
     Lucy, who broke the plate, apologized. ⤳ Lucy broke the plate

   b. Content of sentence-final NRRC
     The police found Jack, who saw the murder. ⤳ Jack saw the murder

   c. Content of the clausal complement of *know*
     Ann knows that Raul cheated on his wife. ⤳ Raul cheated on his wife

   d. Content of the clausal complement of *discover*
     Mary discovered that Denny ate the last cupcake. ⤳ Denny ate the last cupcake

   e. Content of the clausal complement of *be right*
     Tom is right that Ann stole the money. ⤳ Ann stole the money

   f. Content of the clausal complement of *confirm*
     Harry confirmed that Greg bought a new car. ⤳ Greg bought a new car

   g. Content of the clausal complement of *confess*
     Lucy confessed that Dustin lost his key. ⤳ 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. As discussed in §1, Syrett & Koev 2015 observed differences between sentence-medial and -final NRRCs using a variant of the direct-dissent diagnostic, and Degen & Tonhauser 2025 observed differences between *be right, discover, confess, confirm, discover,* and *know* 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 80 unique 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 1 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

As shown in Fig. 2, the response options differed by diagnostic. In Exp. 1 (QUD diagnostic, panel (a)), participants rated how well the response fit the question 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 judged whether the question was about the target content, using a slider marked 'no' on one end (coded as 0) and 'yes' on the other (coded as 1). In Exp. 3 (direct-dissent diagnostic, panel (c)), participants rated the naturalness or the direct dissent 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 chose the response that sounded best; the two indirect dissents were coded as 0 and the direct one as 1. Across the four experiments, the responses were coded so that 1 meant that the content to be diagnosed was rated as at-issue and 0 as not-at-issue.

(a) Exp. 1: QUD diagnostic

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

(c) Exp. 3: direct-dissent diagnostic

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

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

Each of the seven contents in (7) was combined with one of the seven items in (8) in each of the four experiments.

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

   Each experiment included two control stimuli serving as attention checks: one 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 slider (Exps. 1-3) or a 'yes' response (Exp. 4). See Supplement A for the control stimuli used in Exps. 1-4.

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

   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   Dependent measure and interpretation

Following prior work (e.g., Tonhauser et al. 2018), we interpret higher/lower mean ratings as indicating that content is more/less at-issue under a given diagnostic, with two caveats: First, we remain agnostic about whether at-issueness is an underlyingly binary or a gradient property (cf. Tonhauser et al. 2018; Barnes & Ebert 2023). If at-issueness is gradient, the extent to which a content is at-issue may be understood as the extent to which it is relevant to the QUD or the main assertion, in which case gradient mean ratings may be taken to reflect gradient relevance. If at-issueness is categorical, content is at-issue iff it addresses the QUD or assertive proposal, and not-at-issue otherwise; and gradient mean ratings could be attributed to uncertainty about what the QUD/proposal is. For example, our interpretation of a content in a given utterance being more/less at-issue may be interpreted as reflecting the frequency or ease with which a particular QUD is attributed to that utterance. **JT: is this really a caveat to the linking function? LH: It does seem to me that it is, but this is based on my best interpretation of your previous bullet points.** Second, we use the term "at-issueness diagnostic" descriptively throughout the paper, even though our findings may ultimately suggest that these diagnostics track distinct theoretical constructs. We return to this issue in the general discussion.

2.1.4   Data exclusion

We excluded the data of participants that were not self-reported native speakers of American English and of participants whose responses to one of the two control trials was more than 2 sd away from the group mean (Exps. 1–3) or whose responses to one of the two control trials was wrong (Exp. 4). Table 2 shows how many participants were excluded in each experiment, demographic information for the remaining participants, and the number of data points that entered into the analyses.

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) mean 'asking whether' ratings in Exp. 2 ('asking whether' diagnostic), panel (c) shows mean naturalness ratings in Exp. 3 (direct-dissent diagnostic) and panel (d) the proportion of 'no' choices in Exp. 4 ('yes, but' diagnostic).

	exclusion criterion		remaining participants		data
	language	fillers	ages (mean age)	f/m/nb/dnd	points
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.
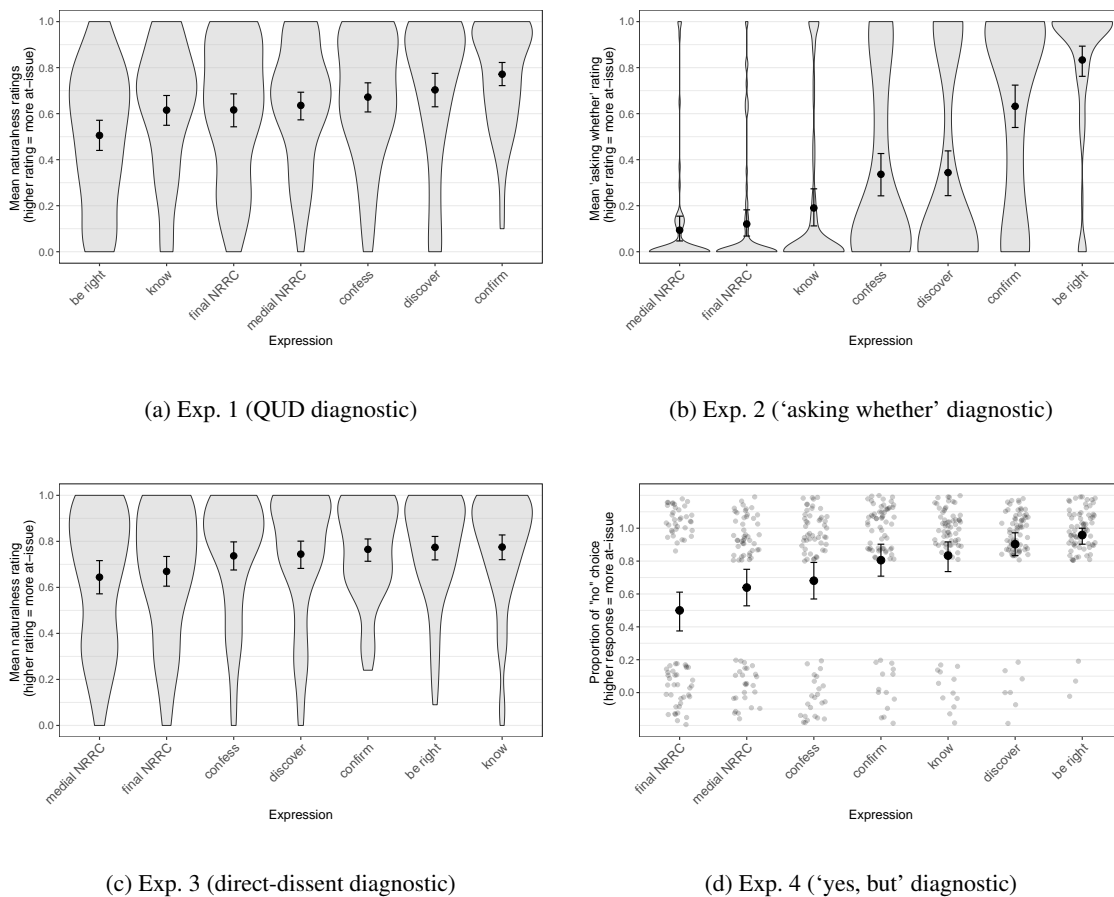

(a) Exp. 1 (QUD diagnostic)

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

(c) Exp. 3 (direct-dissent diagnostic)

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

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).

As is visually apparent in Fig. 3, the diagnostics differ in the overall range of mean responses across the seven contents (given as the difference between the largest and smallest by-content means). The range is largest in Exp. 2 ('asking whether' diagnostic), at .74 (from .1 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.

The experiments differ in whether they reproduce distinctions between contents reported in prior work. 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. In contrast, as shown in Fig. 3, no such difference is observed in Exps. 1–3; and in Exp. 4 ('yes, but' diagnostic), we find the opposite: sentence-final NRRCs are rated less at-issue than sentence-medial ones. Thus, none of the diagnostics as implemented in Exps. 1–4 replicate Syrett & Koev's finding.

Degen & Tonhauser (2025), using the 'asking whether' diagnostic, observed fine-grained differences in at-issueness among the complements of clause-embedding predicates, ordered as follows: *know < discover < confess < confirm < be right*. As shown in Fig. 3, Exp. 2 ('asking whether') largely reproduces this pattern, except that *confess* and *discover* do not differ.

The other three experiments show fewer differences between the mean at-issueness ratings among the clause-embedding predicates. In Exp. 1 (QUD diagnostic), the embedded content of most predicates shows little differentiation, with only two predicates showing a clearly different pattern: the embedded content of *confirm* is rated most at issue, and the embedded content of *be right* is rated less at-issue than the other predicates – the direction of this difference is the opposite to that observed in prior literature and the other experiments. In Exp. 3, no clear differences among clause-embedding predicates are found. Finally, in Exp. 4, the proportions of *yes*-responses are similar for many of the clause-embedding predicates, but the highest proportion is found for content of the complement of *be right*, and it is clearly different from the clause-embedding predicates on the lower end, such as *confirm* and *confess*.

The differences described above are supported by post-hoc pairwise comparisons of the estimated means or proportions for each content, using the `emmeans` package (Lenth 2023) in R (R Core Team 2016), illustrated in Fig. 4. These comparisons were based on mixed-effects beta regression models (Exps. 1–3) and a mixed-effects logistic regression model (Exp. 4), all fit using the `brms` package (Bürkner 2017) using weakly informative priors. The models predicted the ratings[7] from a fixed effect of expression (treatment-coded, with *be right* as reference level), with random intercepts for participants and items. 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.[8]

Accordingly, Figure 4 shows that the only clear positional difference for NRRCs is observed in Exp. 4 ('yes, but' diagnostic), where the sentence-medial ones are rated as more at-issue than sentence-final ones, contrary to previous findings. The other three experiments show no differences between sentence-medial and sentence-final NRRCs. The differences between clause-embedding predicates observed in prior work using the 'asking whether' diagnostic are largely replicated in Exp. 2, except for *confess* and *discover*, which do not differ from each other. The other three experiments show fewer differences between clause-embedding predicates, with Exp. 3 showing no differences at all. In particular, the embedded content of *be right* is rated less at-issue than that of most other clause-embedding predicates in Exp. 1, but more at-issue than content of embedded clauses under most other predicates in Exp. 4.

We also find that the four experiments differ in the relative ratings assigned to the seven contents, with only limited overlap. Across all experiments, *confirm* and *discover* consistently received higher ratings (at least numerically) than *confess*, which in turn received higher ratings than medial and final NRRCs. For all other pairs of expressions, however, the ordering was not consistent. In particular, the ranking of *be right* relative to most other contents is inconsistent across experiments

---

[7] 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.

[8] The full model outputs are available in the folder `results+analysis/exps1-4/` in the repository available at https://anonymous.4open.science/r/at-issueness-diagnostics-232C/.

(a) Exp. 1 (QUD diagnostic)

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

(c) Exp. 3 (direct-dissent diagnostic)

(d) Exp. 4 ('yes, but' diagnostic)
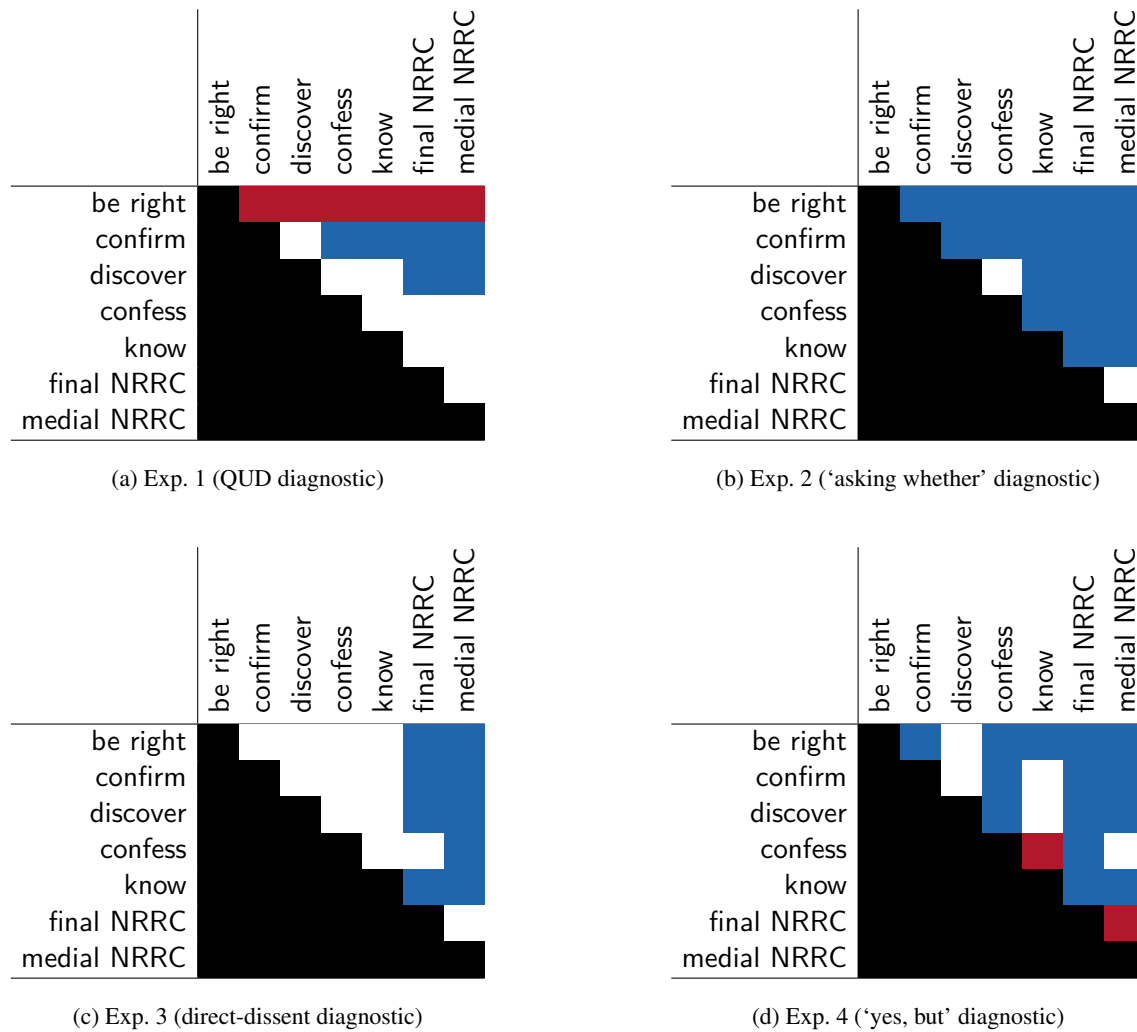
Figure 4: Pairwise differences between expressions, ordered from top to bottom and left to right by increasing mean in Exp. 2 ('asking whether' diagnostic). A white cell means that the 95% HDI of the pair of the row expression and the column expression includes 0, a red cell means that the 95% HDI does 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).

(e.g., compared to appositive NRRCs): The embedded content of *be right* received the lowest ratings in Exp. 1, but was among the most at-issue in the other three experiments. The embedded content of *know* was rated (numerically) lower than that of *confirm* in Exps. 1 and 2, but higher in

	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.

Exps. 3 and 4. This difference between the results of the experiments is quantified in the Spearman rank correlations shown in Table 3.[9]

The rank correlations are particularly low for Exp. 1 compared to the other three experiments, at least partly because of the low relative ranking of *be right*. These results suggest that the four diagnostics as implemented in Exps. 1–4 interact differently with the seven contents investigated.

2.3 Discussion

Exps. 1–4 were designed to compare four at-issueness diagnostics that have been widely used in the literature. To assess whether the diagnostics yield consistent findings when applied to the same target contents, we asked whether they reproduce contrasts reported in prior work, particularly (i) the medial versus sentence-final distinction for appositive NRRCs, and (ii) fine-grained differences among the contents of the complements of clause-embedding predicates. The results of Exps. 1–4 suggest that the four diagnostics do differ in whether they reproduce these contrasts, and in the relative order of the seven contents investigated. Three main differences between the four experiments emerge:

i. The four experiments differ in the extent to which they differentiate between the seven contents investigated, as evidenced by where they find by-content differences: Only the 'asking whether' diagnostic (Exp. 2) comes close to comes close to observing fine-grained differences between the contents of the complements of clause-embedding predicates as in Degen & Tonhauser (2025).

ii. They differ in the relative order of the seven contents – most notably in the case of the complement of *be right*, which ranks low under the QUD diagnostic but high under the other diagnostics.

iii. None of the diagnostics as implemented here reproduce the positional effect for appositive NRRCs reported by Syrett & Koev (2015). Exps. 1-3 show no difference between medial and sentence-final NRRCs, while Exp. 4 (the 'yes, but' diagnostic) yields the opposite pattern: here, medial NRRCs were judged as more at-issue than final ones.

This section discusses the implications of these findings for the interpretation of the diagnostics and for theories of at-issueness.

2.3.1 Positional effects for appositive NRRCs?

One factor that might explain the difference between our findings and those of Syrett & Koev (2015), apart from the stimuli, is the response task. Although both studies tested appositive NRRCs using dissent-based diagnostics presenting the target content in an assertion, their forced-choice

---

[9] The Spearman rank correlation coefficient, a value between -1 and 1, is a nonparametric measure of rank correlation: the higher the absolute value of 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.

implementation differs from our Exps. 3 and 4. In Exp. 3 (direct dissent), we elicited naturalness ratings rather than forced-choice judgments. Even in Exp. 4 ('yes, but'), which also used a forced-choice continuation task, the response options were different. In our implementation (based on Xue & Onea 2011), all options rejected the target content, and participants chose between expressing direct dissent (*no*), indirect dissent (*yes, but*), or assent without contrast (*yes, and*). In Syrett & Koev's study, all options expressed direct dissent (*no*), and participants chose whether to reject the appositive content or the main clause content, as illustrated in (9).

(9)   Syrett & Koev 2015, p. 17
       A: My friend Sophie, a classical violinist, performed a piece by Mozart.
       B1: No, she's not. (target: NRRC)
       B2: No, she didn't. (target: main clause)

While the findings suggest that positional effects for appositive NRRCs may not be robust across diagnostics, the corresponding task differences suggest that even small changes in the response task used to operationalize a diagnostic, such forced-choice alternatives, might reverse the observed pattern. More generally, different diagnostics vary in whether they detect any positional difference for appositive NRRCs at all. This suggests that previously reported effects may reflect task- or diagnostic-specific effects rather than stable properties of NRRC interpretation. A natural follow-up would be to test whether applying their particular forced-choice schema to our stimuli would reproduce their effect.

2.3.2   Interactions with discourse requirements: the case of *be right*

The diagnostics as implemented here yielded different rank orders of contents. At first glance, one might take such differences as evidence that the diagnostics track distinct underlying notions of at-issueness. However, the most pronounced ranking difference, between the low ranking of the complement of *be right* in Exp. 1 (QUD diagnostic) and its high ranking in the other experiments, does not appear to reflect a difference in at-issueness. Instead, it arises from how the QUD diagnostic interacts with the requirements that *be right* places on the prior discourse.

As shown in panel (a) of Fig. 3, participants gave relatively low naturalness ratings to responses like that in (10a), indicating that they did not judge the response as fitting the question.

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

We hypothesize that this is because *be right* in (10) presupposes that Danny has previously committed to the proposition that "Lucy broke the plate" (Abusch 2010; Anand & Hacquard 2014). In the three diagnostics in (10b–d), no previous discourse context is given, so this presupposition can be accommodated. In contrast, the preceding question in the QUD-diagnostic (10a) conflicts

with the presupposition, making it difficult to accommodate. Specifically, we hypothesize that *be right* signals that the question "whether Lucy broke the plate" is salient in the preceding discourse. This allows us to understand the ill-formedness based on QUD-based discourse structure, because this presupposed question is a subquestion (in the sense of Roberts 2012) of the question in (10a) "What did Lucy break", since every complete answer to the latter entails an answer to the former. The progression from the presupposed subquestion question to the explicitly given superquestion violates Roberts' constraint on QUD stacks (p. 6:15), which allows only the reverse order (from superquestions to subquestions).

As a result, low QUD-match ratings for (10a) are predicted, not because the embedded clause is not at-issue, but because the utterance presupposes an incoherent discourse context.[10] This highlights that diagnostics may interact with contextual requirements independently of at-issueness, and such interactions must be taken into account when interpreting the results.

2.3.3 Why does the 'asking whether' diagnostic show greater differentiation?

A striking difference between the results of the experiments is that Exp. 2 ('asking whether') showed the greatest differentiation between contents, whereas the other three experiments exhibited a smaller range of means, and fewer statistically supported differences between investigated contents. Among the first 4 experiments, Exp. 2 was the only one that came close to differentiating fine-grained by-content differences as reported in Degen & Tonhauser (2025).

Considering the question whether the existing definitions of at-issueness target the same underlying phenomenon, we might consider whether this empirical difference between diagnostics aligns with the divide emphasized in prior work between diagnostics that have been associated with a QUD-based definition (QUD, 'asking whether') and those that have been argued to target a proposal-based definition (direct-dissent, 'yes, but'). However, this distinction could not explain why the 'asking whether' diagnostic behaves differently from the QUD diagnostic, since both have been linked to the QUD-based definition of at-issueness.

To explain the greater by-content differentiation of the 'asking whether' diagnostic, we therefore focus on properties that set it apart from the other three diagnostics. Two differences are particularly salient: First, it is the only diagnostic where the target content itself occurs in a polar question, whereas in the other diagnostics, the target content appears in a declarative assertion. Second, it is the only diagnostic where participants are asked directly about the intentions of the speaker, while the other ones collect measures that are more indirectly tied to semantic interpretation, such as acceptability ratings and continuation choices. Either of these properties might be important for explaining the greater differentiation.

These considerations motivate two hypotheses, which we tested in Exps. 5–6. The following section reports on these experiments, which compare the 'asking whether' diagnostic to a diagnostic that also embeds the target content in a polar question but elicits naturalness ratings for a direct response to assess at-issueness. Both experiments yielded similarly fine-grained distinctions among contents, supporting the conclusion that the increased differentiation observed in Exp. 2 is driven primarily by interrogative embedding rather than by the response task.

---

[10] 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

3   Experiments 5 and 6

Exps. 5 and 6 investigate whether the greater differentiation observed in Exp. 2 ('asking whether') was due to (i) embedding target contents in polar questions or (ii) the response task, testing two hypotheses:

i. **Question-embedding hypothesis:** Differentiation between contents is greater when contents are embedded in a polar question than in a declarative assertion.
ii. **Response-task hypothesis:** Differentiation is greater when participants are asked directly what the utterance is about, compared to tasks such as acceptability judgments or forced-choice continuations.

To test these hypotheses, we compared two diagnostics that both embed the target content in a polar question but differ in the response task. Exp. 5 used the 'asking whether' diagnostic as in Exp. 2 (11a) and Exp. 6 uses a direct-response diagnostic (Amaral et al. 2007; Tonhauser 2012), shown in (11b).

(11)   Implementation of the diagnostics in Exps. 5–6

    a.   Exp. 5 ('asking whether' diagnostic)
         Nora: *Is Tom right that Lucy broke the plate?*
         Question to participants: Is Nora asking whether Lucy broke the plate?
    b.   Exp. 6 (direct-response diagnostic)
         Nora: *Is Tom right that Lucy broke the plate?*
         Leo: *Yes, she didn't break the plate.*
         Question to participants: How natural is Leo's response to Nora's question?

This direct-response diagnostic targets the assumption that at-issue content is interpreted as determining the question alternatives (property (ii)). Participants read a dialogue between two named speakers, where the first utters a polar question, which contributes the target content. They then rate the naturalness of a *yes* response that asserts the target content. A response to a question is felicitous only if the target content provides a possible answer, that is, a question alternative (von Stechow 1991). Accordingly, naturalness ratings are expected to be high when the target content is at-issue and low otherwise.

Both experiments measured at-issueness of the contents of the complements of the 20 clause-embedding predicates from Degen & Tonhauser 2025:

(12)   20 clause-embedding predicates from Degen & Tonhauser 2025:

         acknowledge, admit, announce, be annoyed, be right, confess, confirm, demonstrate, discover, establish, hear, inform, know, pretend, prove, reveal, say, see, suggest, think

Both experiments also assessed the projection of the complement contents. This data was reported in Hofmann et al. 2024. Here we focus on the at-issueness data, which have not yet been reported.

If the greater differentiation between contents found in Exp. 2 (compared to Exps. 1, 3, and 4) was driven primarily by presenting the target content in a polar question, we expect Exps. 5 and 6 to show similar fine-grained by-content differentiation of at-issueness. In contrast, if Exp. 5 again shows greater differentiation than Exp. 6, this would support the response-task hypothesis, indicating that the task, rather the question embedding, drives the effect.

3.1 Methods

3.1.1 Participants

We recruited 300 participants on Amazon's Mechanical Turk platform for Exp. 5 and 250 partici-
pants on Prolific for Exp. 6.[11] Participants recruited on Mechanical Turk had U.S. IP addresses
and at least 99% of previous HITs approved. Participants recruited on Prolific had registered as
U.S.-born native speakers of English residing in the USA and had an approval rate of at least 99%.
Table 4 summarizes the age and gender distributions of the recruited participants.

	recruited	ages (mean age)	f/m/nb/dnd
Exp. 5 (asking whether)	300	19-74 (38.2)	–/–/–/–
Exp. 6 (direct assent)	250	18-58 (25.5)	201/43/6/0

Table 4: Information about the participants recruited in Exps. 5-6 (f = female, m = male, nb
= nonbinary, dnd = did not disclose; gender information was not collected in Exp. 5).

3.1.2 Materials and procedure

As shown in Fig. 5, the response task differed between experiments. In Exp. 5 ('asking whether'
diagnostic, panel (a)), participants judged whether the question was about the content of the
complement clause, using a slider marked 'no' on one end (coded as 0) and 'yes' on the other
(coded as 1). In Exp. 6 (direct-response diagnostic, panel (b)), participants rated whether the
response to the first speaker sounds good, on a slider marked 'no' (coded as 0) on one end and 'yes'
on the other (coded as 1). In the initial instructions, participants were asked to imagine that they
are at a party and that, when walking into the kitchen, they overhear somebody say something to
somebody else. Across both experiments, the responses were coded so that 1 meant that the content
of the embedded clause was rated as at-issue and 0 as not-at-issue.

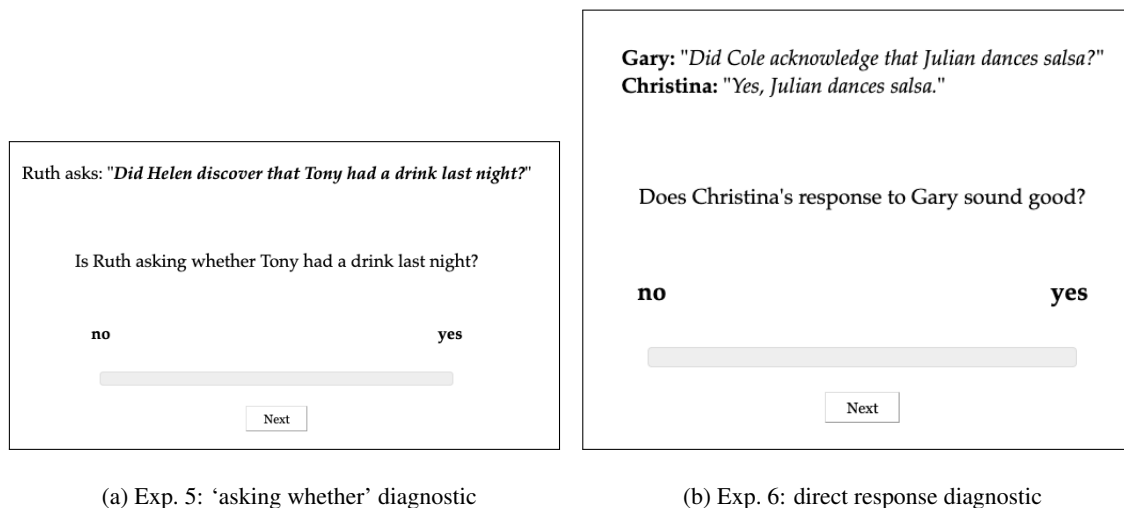(a) Exp. 5: 'asking whether' diagnostic   (b) Exp. 6: direct response diagnostic

Figure 5: Sample trials in (a) Exp. 5 and (b) Exp. 6.

In both experiments, target stimuli were created by combining each of the 20 clause-embedding
predicates in (12) with one of 20 items (listed in Supplement B). The same six control stimuli were

---

[11] Exp. 5 was run in August 2019 and Exp. 6 in August 2021.

included in both experiments. The contents targeted in these controls (see Supplement C) were expected to be at-issue and served as attention checks. For each participant, the 20 predicates were randomly paired with 20 unique items, yielding one target stimulus per predicate. Participants thus saw 26 stimuli in total: 20 target items and 6 controls.[12] Trial order was randomized for each participant. 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.

3.1.3 Data exclusion

We excluded the data of participants who did not self-identify as native speakers of American English, of participants whose responses to the projection or at-issueness controls were more than 2 sd away from the group mean, and of participants who always selected roughly the same point on the response scale for the target stimuli. **JT: did we do this latter one in exps 1-4 as well? LH: I don't think so, and if it should be the same in both, I would suggest omitting this latter criterion here, since it does not seem to make much of a difference.** To identify such participants, we first identified participants whose mean variance on the target stimuli was more than 2 sd below the group mean variance and then manually inspecting their response patterns. Due to a programming error, 5 participants took Exp. 5 more than once. Since we were not able to identify which submission was their first submission, the data of these participants was also excluded. Table 5 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
	language	controls	variance	ages (mean age)	f/m/nb/dnd	total	points
Exp. 5 (asking whether)	7	35	0	21-74 (39.2)	–/–/–/–	242	6292
Exp. 6 (direct assent)	5	24	1	18-58 (24.9)	187/28/5/0	220	5720

Table 5: Information from Exps. 5-6 about the number of participants whose data was excluded based on their self-declared language and language variety ('language'), the controls, and the variance of their responses, about the remaining participants, and about the number of at-issueness data points that entered into the analysis.

3.2 Results and discussion

Fig. 6 shows the results of the two experiments by embedding predicate: panel (a) plots mean 'asking whether' ratings in Exp. 5, and panel (b) plots mean naturalness ratings in Exp. 6 (direct-response).

Overall, the results of two experiments yield highly similar patterns. The ranges of by-content means are comparable (Exp. 5 range: .75, .07 to .82; Exp. 6 range: .73, .09 to .82), and the relative orderings of contents by their means align closely, as reflected in a strong Spearman rank correlation ($r_s = .93$).

Crucially, both experiments show fine-grained by-content differentiation. For instance, for the subset of predicates also tested in Exps. 1–4, Exps. 5 and 6 reproduce the ordering expected based on Degen & Tonhauser (2025) and show pairwise differences in the predicted directions for adjacent contents (*know < discover < confess < confirm < be right*). Overall, the results in Exps. 5 and 6 are similar in revealing fine-grained differences among many rank-adjacent predicates.

---

[12] Each participant saw their set of 26 stimuli twice, once in the projection block and once in the at-issueness block. Block order was randomized. As mentioned above, we focus here on the at-issueness data.
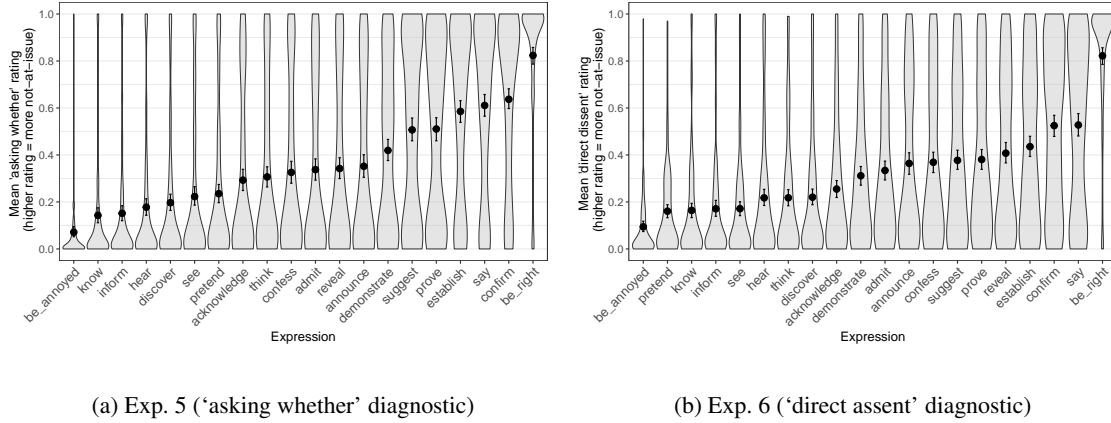
(a) Exp. 5 ('asking whether' diagnostic)   (b) Exp. 6 ('direct assent' diagnostic)

Figure 6: Results of Exps. 5–6. The panels show the mean ratings by expression for (a) Exp. 5 (asking whether diagnostic) and (b) Exp. 2 (direct assent diagnostic). Error bars indicate 95% bootstrapped confidence intervals. Violin plots show the kernel probability density of individual participants' ratings.

There are also some differences between the experiments, concerning a small number of local reversals in the middle of the scale. For example, the content of the complement of *demonstrate* comes out as more at-issue than that of *announce*, *confess*, and *reveal* under the 'asking whether' diagnostic as implemented in Exp. 5, but as less at-issue than the content embedded under these predicates under the implementation of the direct-response diagnostic in Exp. 6.

Fig. 7 presents the results of post-hoc pairwise comparisons of the estimated means for each content. As in Section 2, we used `emmeans` (Lenth 2023) in R (R Core Team 2016) applied to mixed-effects beta regression models fit with `brms` (Bürkner 2017) using weakly informative priors. The models predicted ratings[13] from a fixed effect of content (treatment-coded, with *be right* as the reference level). The pairwise comparisons support the above generalizations: while the two diagnostics as implemented here differ in which specific contrasts they detect, their overall patterns are highly consistent. Both diagnostics reveal fine-grained at-issueness differences between contents. At the same time, there are some differences, namely some contrasts that reached significance in experiment did not in the other.

Taken together, Exps. 5 and 6 suggest that the two diagnostics that embed target contents implemented here ('asking whether' and direct-response) yielded highly similar results for the 20 contents investigated, including fine-grained by-predicate differentiation. This supports the idea that presenting the target content in a polar question is what drives the increased by-content differentiation observed in Exps. 2, 5, and 6 compared to Exps. 1, 3, and 4, where the target content was presented in declarative assertions.

Nonetheless, small differences remain, for example in which by-content contrasts reached significance and in the precise rank ordering of contents. Since the two experiments used identical embedding contexts for the target content and both tested property (ii), which has been linked to a QUD-based definition of at-issueness, these differences may stem from the response tasks used: judging what a question is about (Exp. 5) versus evaluating the naturalness of a response (Exp. 6). This suggests that response task differences can affect at-issueness judgments in subtle ways.

---

[13] To model the ratings 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.
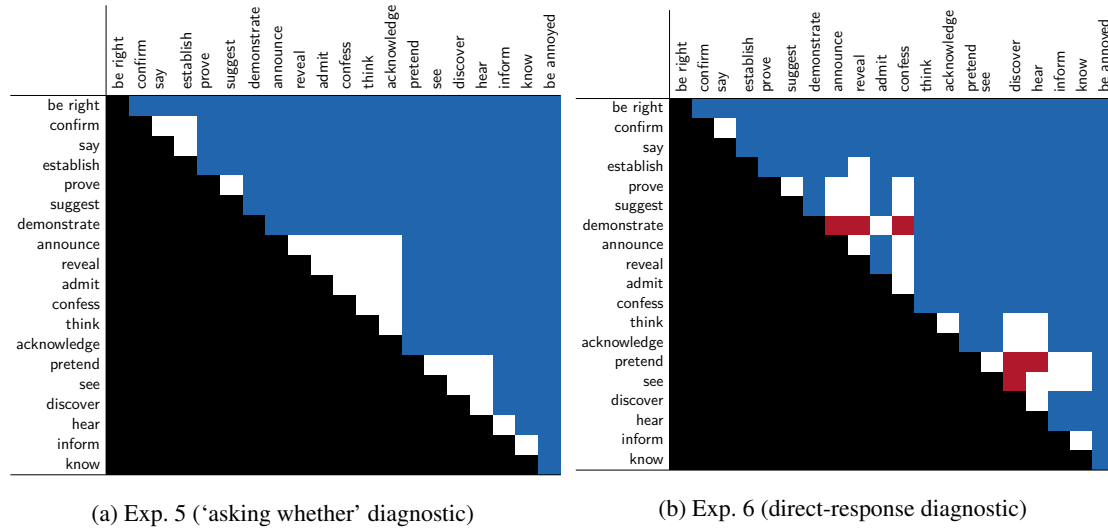
(a) Exp. 5 ('asking whether' diagnostic)   (b) Exp. 6 (direct-response diagnostic)

Figure 7: Pairwise differences between expressions, ordered from by increasing mean in Exp. 5 ('asking whether'). White cells indicate that the 95% HDI of the difference includes 0. Red cells indicate a positive difference (the row expression received a higher rating than the column expression), and blue cells indicate a negative difference.

4 General discussion

This paper reported the results of six experiments designed to investigate whether at-issueness diagnostics yield consistent results when applied to the same target contents. Exps. 1–4 show that, as implemented, the diagnostics do not converge, suggesting that they are not interchangeable.

Taken together, the experiments provide converging evidence that presenting the target content in a polar question yields greater by-content differentiation than presenting it in an assertion: When target contents were presented in a polar question, as in the 'asking whether' diagnostic (Exps. 2, 5) and the direct-response diagnostic (Exp. 6), we observed high by-content differentiation, reproducing fine-grained differences reported in prior work. In contrast, when the same contents were presented in an assertion, as in the QUD diagnostic (Exp. 1) and the dissent-based diagnostics (Exps. 3-4), the resulting ratings showed less differentiation. An implication is that clause type and the illocutionary force of the utterance in which the target content is presented needs to be taken into consideration, when interpreting outcomes of at-issueness diagnostics.

Beyond this main result, we observed additional differences between diagnostics that are specific to particular tested contents. First, diagnostics varied in whether they detected positional effects for appositive NRRCs, and even in the direction of such effects (Section 2.3.1). This was the case for our dissent-based diagnostics (Exps. 3–4) and Syrett & Koev's (2015) study, despite their shared reliance on the underlying assumptions that at-issue content cannot be directly dissented with (property (iii)). The studies differ in the choice or implementation of the response tasks, suggesting that relatively small design choices, such as the alternatives in a forced-choice continuation task, can affect results.

Second, the complement of *be right* ranked low under the QUD diagnostic (Exp. 1) despite coming out as the most at-issue content in all other experiments (Section 2.3.2). We argued that this reflects an interaction between the QUD diagnostic and the discourse presuppositions of *be right*, rather than a systematic difference in at-issueness.

A methodological implication of these diagnostic differences is that they cannot be treated as interchangeable. Theoretical claims based on one diagnostic should therefore be relativized to

the diagnostic used (see also Snider 2017b; a; 2018; Koev 2018; Korotkova 2020). Therefore, empirical investigations should carefully consider that results may depend on whether the target content was presented in an assertion or polar question, contextual requirements of the expressions used, or the particular task design.

The remainder of this section discusses further implications of our findings, offering discussion on the role of question embedding for diagnosing at-issueness (Section 4.1), and on what the observed differences might imply for the question of whether there are distinct underlying phenomena associated with the notions of QUD at-issueness and proposal at-issueness (Section 4.2).

4.1   Speech act, at-issueness, and commitment

We suggest that the lower by-content differentiation observed in diagnostics that embed target content in assertions, compared to polar questions, may reflect a ceiling effect in assertion-based diagnostics. We hypothesize that this could result from how the illocutionary force of assertions relates propositional content to speaker commitment in different ways: in assertions, but not in questions, the speaker commits to the truth of the at-issue main clause content.

To consider the argument in detail, we begin by noting that at-issue content and not-at-issue content contributed by polar questions differ in their relationship to speaker commitments. The at-issue contribution of a polar question such as (13) determines the set of question alternatives, without committing the speaker to any of the alternatives (13a).

(13)   *Is Greg, who bought a car, envied by his neighbor?*

    a.   At-issue contribution: determining the question alternatives:
        {Greg, who bought a car, is envied by his neighbor,
                            Greg, who bought a car, is not envied by his neighbor}
        $\nRightarrow$ *Greg is envied by his neighbor.*

    b. $\Rightarrow$ *Greg bought a car.*   (not-at-issue content)

In contrast, the not-at-issue NRRC content (13b) is assumed to be true across all question alternatives, and thus throughout the entire context set, giving rise to projection (Abusch 2010; Simons et al. 2010; Abrusán 2011; Tonhauser et al. 2018; Degen & Tonhauser 2025). As a result, the projective not-at-issue content and the at-issue main clause content differ in their *epistemic status* relative to the speaker commitments: Because the not-at-issue proposition projects, it is treated as part of the speaker's commitments, while the at-issue main clause content is not.

In assertions, the contrast between the epistemic status of not-at-issue content and the at-issue main-clause content is weakened. An assertion the proposes to add the at-issue main clause content the common ground, and also commits the speaker to its truth (Gunlogson 2001). For instance, the at-issue contribution in (14), commits the speaker to the main clause content (14a). In addition, the assertion in (14) comes with speaker commitment to the not-at-issue NRRC content (14b) as well.

(14)   *Greg, who bought a car, is envied by his neighbor.*

    a. $\Rightarrow$ *Greg is envied by his neighbor.*   (at-issue main clause content)
    b. $\Rightarrow$ *Greg bought a car.*   (not-at-issue content)

We suggest that because the at-issue main clause content and not-at-issue content have the same epistemic status in assertions relative to speaker commitments, the two types of content are harder to distinguish for participants. This reasoning helps explain why the two question-based diagnostics yielded highly similar results. Both the asking whether diagnostic used in Exps. 2 and 5 and the direct-response diagnostic used in Exp. 6 embed the target content *p* in a polar question, as in (15),

where the target content is (15a), and the utterance signals that speaker is committed to the main clause at-issue content (15b).

(15)   *Helen confirmed that Tony had a drink last night.*
    a.   p : *Tony had a drink last night*   (target content)
    b.   q : *Helen confirmed* p   (main-clause content)

Now, if (15a) (p) is interpreted as more at-issue, then participants will give responses that are consistent with this more at-issue interpretation. If p is interpreted as less at-issue, construe q as the primary at-issue contribution. In this case, asserting q : *Helen confirmed* p also leads the speaker to commit to the target content p. As a result, the supposed not-at-issue content p would be similar in its epistemic status to the at-issue main clause content p, because the speaker is committed to both, which could make them harder to distinguish. The epistemic status of the at-issue main clause content and not-at-issue content is therefore highly similar in assertions.

We suggest that this similarity makes the two types of content harder to distinguish in assertion-based diagnostics. This would lead to not-at-issue content being rated more similarly to at-issue content, giving rise to overall higher ratings for not-at-issue content and reduced differentiation across contents, essentially creating a ceiling effect.

For diagnostics that embed target content in polar questions, the distinct epistemic status between not-at-issue content and the at-issue main clause content avoids this difficulty in differentiation. **LH: We could also look at response times across experiments to look for support for this hypothesis.** For example, for the polar question embedding in (16), the target content is (16a).

(16)   *Did Helen confirm that <u>Tony had a drink last night</u>?*
    a.   p : *Tony had a drink last night*   (target content)
    b.   q : *Helen confirmed p*   (main-clause content)

Here, if (16a) (p) is interpreted as more at-issue, then participants will interpret the question as being (more) about whether p. If p is interpreted as less at-issue, participants will interpret the question as being (more) about whether q. In that case, p is likely to project, leading to speaker commitment. However, since the main clause content of the question (about whether p) does not come with speaker commitment, the at-issue main clause content and the not-at-issue content differ in their epistemic status.

In sum, the differences in by-content variation between diagnostics that embed contents in polar questions versus assertions can be understood as a result of how the illocutionary force of assertions and polar questions interacts with a further property of at-issue content, namely that the speaker is overall more likely to be committed to not-at-issue content. This will be the case either by entailment (like when the embedded content in (14+15) is not-at-issue), or by projection from entailment-cancelling environments like polar questions (like when the embedded content in (13+16) is not-at-issue).

To our knowledge, no previous literature has proposed that the speech act in which target content is presented can itself be a major driver (or inhibitor) of differentiation. Here, we saw that polar questions more transparently reveal how different contents can be interpreted as at-issue or not-at-issue, and suggested that this may be because the two contents differ in their epistemic status. Assertions, in contrast, commit the speaker both to the at-issue assertion and the not-at-issue content, thereby obscuring the distinction. Consequently, diagnostics that embed the target content in questions provide a particularly sensitive test of at-issueness. Embedding target content in questions therefore allows for detection of more fine-grained differences in at-issueness associated with the conventional meaning of particular expressions, whereas assertions introduce

a confounding factor: the speaker commits to both types of content, reducing differentiation in at-issueness judgments.

- **JT: the paragraph around 12 is about a sentence-medial NRRC, which is maximally different in terms of at-issueness and speaker commitment from the at-issue content, which is at-issue and the speaker is not committed. you suggested above that speaker commitment is relevant. so when we have content that is not as clearly not-at-issue as sentence-medial NRRCs, like the CC of "discover" or of "reveal", why does embedding in questions help bring out differences between them? i don't see how the example in 12 helps explain this, or something is missing**
  LH: It's less that question embeddings help to bring out the differences, but more that assertions obscure the differences. I hope that's clearer now.
- **JT: does this imply that embedding under negation should also not work so well (commitment to negated at-issue content) but embedding under "perhaps" (or another epistemic modal) should? don't we have experiments like Exp 6 but with embedding under "perhaps" and with embedding under negation?**
  LH: It shouldn't, but I hadn't really considered AI content in various non-veridical environments when I wrote the first draft. In the current version, it should become clearer that I am talking about AI main clause content. In assertions with negation or "perhaps", there is still at least the main clause content that the speaker is committed to.

4.2 Diagnostic differences and notions of at-issueness

Because the diagnostics are grounded in distinct assumptions about what it means for content to be at-issue, the observed differences between diagnostics also bear on the broader question of whether existing diagnostics track a common underlying notion of at-issueness or operationalize distinct phenomena. While many of the diagnostic differences in our data find plausible explanations in methodological and contextual factors, we found no obvious empirical split between QUD-at-issueness and proposal at-issueness. This section considers what our results suggest for this issue.

As reviewed in Section 1, diagnostics differ both in the properties they target and in the theoretical notions they are taken to diagnose. Recall the definitions of QUD at-issueness and proposal at-issueness, repeated here:

(5)   QUD at-issueness: (based on Simons et al. 2010, p. 323)
        A content m is at-issue in a context c iff

        a. m is relevant[14] to the QUD in c, and
        b.   the speaker has a recognizable intention to address the QUD via m.

(6)   Proposal at-issueness: (Koev 2013, pp. 50–51)
        A proposition p is 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.

Table 6 summarizes how the four diagnostics from Exps. 1–4 have been mapped onto these notions in the literature. Three broad positions can be distinguished. On one view, all diagnostics ultimately target a single underlying phenomenon (Tonhauser 2012; AnderBois et al. 2015; Faller 2019), so any empirical differences must be attributed to independent factors. A second view holds that some diagnostics track QUD at-issueness while others track anaphoric availability (Snider 2017a; b; 2018; Korotkova 2020), predicting systematic differences between diagnostics that involve propositional anaphora and those that do not. A third view posits genuinely distinct notions, such as QUD

at-issueness and proposal at-issueness (Koev 2018), which predicts that at least some diagnostic differences reflect this theoretical divide.

Diagnostic	Property targeted by diagnostic	Underlying notion assumed to correspond to property		
		Uniform at-issueness Tonhauser 2012, AnderBois et al. 2015, Faller 2019	Anaphoric availability Snider 2017a, Korotkova 2020	Dual-notion Koev 2018
QUD diagnostic	(i) addressing QUD	QUD at-issueness	QUD at-issueness	QUD at-issueness
Asking whether	(ii) determining alternatives			–
Direct dissent	(iii) direct dissent/ assent		Anaphoric availability	Proposal at-issueness
'Yes, but'				

Table 6: Mapping between at-issueness diagnostics, the properties they target (Tonhauser 2012), and the underlying theoretical notions assumed in different theoretical approaches.

Under the view that assumes a uniform notion of at-issueness (Tonhauser 2012; AnderBois et al. 2015; Faller 2019), all diagnostics target the same underlying notion of at-issueness and all diagnostic differences would have to be explained by factors independent of at-issueness. The anaphoric availability view (Snider 2017a; a; 2018; Korotkova 2020) assumes that some diagnostics targeting properties (i) and (ii) genuinely target QUD at-issueness and diagnostics targeting property (iii) diagnose anaphoric availability. This predicts systematic differences between diagnostics that do and do not involve propositional anaphora. Under the dual-notion position (Koev 2018), which assumes separate underlying phenomena associated with the two notions defined above, at least some diagnostic differences should be explained by the distinction. Here, we consider our main findings in light of these positions.

First, the contrast that aligns with the clearest difference between diagnostics we found – whether target contents are presented in polar questions versus assertions – may appear related to the distinction between QUD-based and proposal-based definitions of at-issueness, since both involve a question versus assertion contrast. However, as discussed in Section 2.3.3, this pattern is not predicted by the differences in the underlying notions targeted by the diagnostics that have been proposed in the literature. The dual-notion view could only provide an explanation for differences between diagnostics targeting property (i) (the QUD diagnostic) and those targeting property (iii) (direct dissent and 'yes, but'). The anaphoric-availability view would be supported by differences between diagnostics targeting properties (i) or (ii)[15] such as the QUD diagnostic and 'asking whether' and those targeting property (iii). Our main contrast, however, cuts across these groupings. Although the main diagnostic difference is not straightforwardly predicted by either the dual-notion view or the anaphoric-availability view, it remains in principle compatible with all three positions.

Second, the diagnostics differ in how they interact with contextual requirements imposed by particular expressions. A clear example is the behavior of *be right* under the QUD diagnostic, where low QUD-match ratings do not plausibly reflect not-at-issueness, but instead arise from a pragmatic incompatibility: *be right* presupposes a discourse structure that conflicts with the assumptions made by the diagnostic (see Section 2.3.2). A related point is raised by Snider

---

[15] The status of the direct-response diagnostic further complicates the picture. It targets property (ii), which Snider aligns with QUD-based at-issueness, but it also involves the response particle yes, which has been argued to diagnose anaphoric availability in property (iii) diagnostics. It is therefore unclear how this diagnostic should be classified under the anaphoric-availability view.

(2017a; 2018) who emphasized that the direct-dissent and 'yes but' diagnostic involve propositional anaphora in the interpretation in the response particles *yes/no*. Different diagnostics impose different requirements on discourse structure, common ground, and anaphoric accessibility, all of which can independently affect acceptability judgments. This supports the conclusion that some differences between diagnostics arise from interactions with expression-specific contextual requirements rather than from differences in at-issueness per se.

Third, some small differences between diagnostics can be attributed to response-task effects. This is evident in the subtle differences between Exps. 5 and 6 (see Section 3.2), which differed only in response format, and in our failure to replicate Syrett & Koev's 2015 reported positional effect for appositive NRRCs (see section 2.3.1). While they found sentence-final appositives to be more at-issue than sentence-medial ones under a version of the direct-dissent diagnostic, our assent/dissent-based *yes, but* diagnostic in Exp. 4 showed a difference in the opposite direction, while the other experiments found no difference between them (Exp.1–3). Taken together, these results provide no evidence for a stable notion of at-issueness under which sentence-final appositives are systematically more at-issue than medial ones (or the other way round), and they highlight how even small task differences can affect outcomes.

At the same time, among most pairs from Exps. 1–4, the Spearman rank correlation was at least moderate (when excluding *be right*, see footnote 10), suggesting some convergence among diagnostics, and that the diagnostics, as implemented there, are at least partially sensitive to a shared underlying property.

Our data does not provide conclusive evidence about whether or not the diagnostics track fundamentally different underlying notions of at-issueness. However, the fact that many of the differences find explanations outside this question, seems more compatible with assuming that the diagnostics offer different, partially overlapping windows onto a single discourse phenomenon, that interacts with multiple discourse factors (speaker commitments, speech acts, contextual requirements, and the given alternatives) in complex and sometimes subtle ways, so that even small differences in task design can make a difference.

5   Conclusion

In a series of six experiments, this study investigated how different diagnostics for at-issueness behave when applied to the same linguistic stimuli, using five different diagnostics: the QUD diagnostic, the 'asking whether' diagnostic, the direct-dissent diagnostic, the 'yes, but' diagnostic, and the direct-response diagnostic. Our findings provide experimental confirmation for claims that there are empirical differences between at-issueness diagnostics (Snider 2017b; a; 2018; Koev 2018; Faller 2019; Korotkova 2020). The diagnostics yielded different results, with some diagnostics showing greater differentiation among contents than others. The main finding was embedding the target content in questions leads to greater differentiation among contents than embedding it in assertions. We suggested that this finding highlights the need to consider the speech act context when interpreting results from at-issueness diagnostics. An important question for future work remains open: what the results of different diagnostics might reveal about whether or not they reflect a shared underlying notion of at-issueness.

Data accessibility statement

The experiments, data and R code for generating the figures and analyses of the experiments reported in this paper are available at https://anonymous.4open.science/r/at-issueness-diagnostics-232C/.

Ethics and consent

All experiments were conducted with approval from the ethics review committee of [university name redacted for review].

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. Publisher: Elsevier

Abrusán, Márta. 2011. Predicting the presuppositions of soft triggers. *Linguistics and Philosophy* 34(6). 491–535. https://doi.org/10.1007/s10988-012-9108-y

Abusch, Dorit. 2010. Presupposition triggering from alternatives. *Journal of Semantics* 27(1). 37–80.

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.

Anand, Pranav & Hacquard, Valentine. 2014. Factivity, belief and discourse. In Crni, Luka & Sauerland, Uli (eds.), *The Art and Craft of Semantics: A Festschrift for Irene Heim*, 69–90. MIT Working Papers in Linguistics. https://semanticsarchive.net/Archive/jZiNmM4N/.

Anand, Pranav & Korotkova, Natasha. 2024. Facts, intentions, questions: English come-to-knowpredicates in deliberative environments. In *Proceedings of the Amsterdam Colloquium*. 15–21. https://platform.openjournals.nl/PAC/article/view/21764.

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.

AnderBois, Scott & Brasoveanu, Adrian & Henderson, Robert. 2015. At-issue proposals and appositive impositions in discourse. *Journal of Semantics* 32(1). 93–138. https://doi.org/10.1093/jos/fft014

Bade, Nadine. 2024. New data on the 'triggering problem'for presuppositions. In *Semantics and Linguistic Theory*. 197–212. https://journals.linguisticsociety.org/proceedings/index.php/SALT/article/view/34.010.

Bade, Nadine & Schlenker, Philippe & Chemla, Emmanuel. 2024. Word learning tasks as a window into the triggering problem for presuppositions. *Natural Language Semantics* 32(4). 473–503. https://doi.org/10.1007/s11050-024-09224-5. https://link.springer.com/10.1007/s11050-024-09224-5

Barnes, Kathryn & Ebert, Cornelia. 2023. The information status of iconic enrichments: modelling gradient at-issueness. *Theoretical Linguistics* 49(3-4). 167–223. https://doi.org/10.1515/tl-2023-2009. https://www.degruyter.com/document/doi/10.1515/tl-2023-2009/html

Beaver, David & Clark, Brady. 2008. *Sense and Sensitivity: How Focus Determines Meaning*. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781444304176

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

Büring, Daniel. 2003. On D-Trees, Beans, And B-Accents. *Linguistics and Philosophy* 26(5). 511–545. https://doi.org/10.1023/A:1025887707652. https://link.springer.com/10.1023/A:1025887707652

Chen, Yuqiu. 2024. *Presuppositions at the Semantics-Pragmatics Interface*: Georg-August-Universität Göttingen Doctoral dissertation. https://ediss.uni-goettingen.de/handle/11858/15164.

Clark, Herbert H. & Schaefer, Edward F. 1989. Contributing to discourse. *Cognitive science* 13(2). 259–294. https://www.sciencedirect.com/science/article/pii/0364021389900086.

Cummins, Chris & Amaral, Patricia & Katsos, Napoleon. 2013. Backgrounding and accomodation of presuppositions: An experimental approach. In *Proceedings of Sinn und Bedeutung*, vol. 17. 201–218.

Degen, Judith & Tonhauser, Judith. 2025. Projection inferences: On the relation between prior beliefs, at-issueness, and lexical meaning. *Glossa Psycholinguistics* 4(1). https://escholarship.org/uc/item/7hb2h628.

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, Floran (ed.), *Experimental approaches to presupposition*, 135–156. Springer. https://doi.org/10.1007/978-3-319-07980-6

Esipova, Maria. 2019. *Composition and projection in speech and gesture*. New York, NY: New York University Doctoral dissertation. https://www.researchgate.net/publication/371702326_Composition_and_projection_in_speech_and_gesture.

Esipova, Maria. 2021. On not-at-issueness in pictures. *Glossa: a journal of general linguistics* 6(1). https://www.glossa-journal.org/articles/10.5334/gjgl.1314/?utm_source=TrendMD&utm_medium=cpc&utm_campaign=Glossa%253A_a_journal_of_general_linguistics_TrendMD_0.

Faller, Martina. 2006. Evidentiality below and above speech acts. Unpublished manuscript. https://semanticsarchive.net/Archive/GZiZjBhO.

Faller, Martina T. 2002. *Semantics and pragmatics of evidentials in Cuzco Quechua (Peru)*: Stanford University Doctoral dissertation. https://www.proquest.com/dissertations-theses/semantics-pragmatics-evidentials-cuzco-quechua/docview/305523331/se-2?accountid=10957.

Faller, Martina T. 2019. The discourse commitments of illocutionary reportatives. *Semantics and Pragmatics* 12. 8:1–53. https://doi.org/10.3765/sp.12.8. https://semprag.org/index.php/sp/article/view/sp.12.8

Farkas, Donka & Bruce, Kim. 2010. On reacting to assertions and polar questions. *Journal of Semantics* 27(1). 81–118. https://doi.org/10.1093/jos/ffp010. https://scholar.archive.org/work/hymiwpqf5fdazop52liak5zuoa/access/wayback/https://people.ucsc.edu/~farkas/papers/assertion_question.pdf

Ginzburg, Jonathan. 1995. Resolving Questions I, II. *Linguistics & Philosophy* 18. 459–527, 567–609.

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.

Groenendijk, Jeroen Antonius Gerardus & Stokhof, Martin Johan Bastiaan. 1984. *Studies on the Semantics of Questions and the Pragmatics of Answers*: Univ. Amsterdam PhD Thesis.

Gunlogson, Christine. 2001. *True to form: Rising and falling declaratives as questions in English*: University of California Santa Cruz Doctoral dissertation. http://hdl.handle.net/1802/231.

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.

Hunter, Julie & Asher, Nicholas. 2016. Shapes of Conversation and At-Issue Content. *Semantics and Linguistic Theory* 1022–1042. https://doi.org/10.3765/salt.v26i0.3946. https://journals.linguisticsociety.org/proceedings/index.php/SALT/article/view/26.1022

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.

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.

Lee, Jungmee. 2011. *Evidentiality and its interaction with tense: Evidence from Korean*: The Ohio State University PhD Thesis. https://rave.ohiolink.edu/etdc/view?acc_num=osu1306940284.

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. 2010. *Evidentiality and the structure of speech acts*: Rutgers The State University of New Jersey-New Brunswick Doctoral dissertation.

Murray, Sarah E. 2014. Varieties of update. *Semantics and Pragmatics* 7. 2:1–53. https://doi.org/10.3765/sp.7.2. https://semprag.org/index.php/sp/article/view/sp.7.2

Papafragou, Anna. 2006. Epistemic modality and truth conditions. *Lingua* 116(10). 1688–1702. https://www.sciencedirect.com/science/article/pii/S0024384106000805. Publisher: Elsevier.

Potts, Christopher. 2005. *The Logic of Conventional Implicatures*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199273829.001.0001

Potts, Christopher. 2015. Presupposition and Implicature. In Lappin, Shalom & Fox, Chris (eds.), *The Handbook of Contemporary Semantic Theory*, 168–202. Wiley 1st edn. https://doi.org/10.1002/9781118882139.ch6. Koev

R Core Team. 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

Roberts, Craige. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5. 6:1–69. https://doi.org/10.3765/sp.5.6. https://semprag.org/index.php/sp/article/view/sp.5.6

Schlenker, Philippe. 2021. Triggering presuppositions. *Glossa: a journal of general linguistics* 6(1). https://www.glossa-journal.org/articles/10.5334/gjgl.1352/. Publisher: Open Library of Humanities.

Scontras, Gregory & Tonhauser, Judith. 2025. Projection without presupposition: A model for know. In *Sinn und Bedeutung (SuB) 29*. To appear.

Shanon, Benny. 1976. On the two kinds of presuppositions in natural language. *Foundations of language* 14(2). 247–249. https://www.jstor.org/stable/25170057.

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. 2017a. At-issueness anaphoric availability. *Proceedings of the Linguistic Society of America* 2. 39–1. http://journals.linguisticsociety.org/proceedings/index.php/PLSA/article/view/4089.

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.

Snider, Todd N. 2017b. *Anaphoric reference to propositions*. Ithaca, NY: Cornell University Doctoral dissertation. https://ecommons.cornell.edu/server/api/core/bitstreams/3794923d-

d85f-4c7c-85ac-8106f83152b5/content.

Solstad, Torgrim & 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. 1973. Presuppositions. *Journal of Philosophical Logic* 4. 447–57.

Stalnaker, Robert C. 1978. Assertion. In Cole, Peter (ed.), *Pragmatics* (Syntax and Semantics 9), 315–332. Leiden, Netherlands: Brill. https://doi.org/10.1163/9789004368873_013.

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 & 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

von Stechow, Arnim. 1991. Focusing and backgrounding operators. In Abraham, Werner (ed.), *Discourse particles: Descriptive and theoretical investigations on the logical, syntactic and pragmatic properties of discourse particles in German*, 37–84. Amsterdam, Netherlands: John Benjamins.

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)

B 20 clauses

The contents of the following 20 clauses, which realized the complements of the 20 clause-embedding predicates, were investigated in Exps. 5–6:

i.	Mary is pregnant.	xi.	Danny ate the last cupcake.
ii.	Josie went on vacation to France.	xii.	Frank got a cat.
iii.	Emma studied on Saturday morning.	xiii.	Jackson ran 10 miles.
iv.	Olivia sleeps until noon.	xiv.	Jayden rented a car.
v.	Sophia got a tattoo.	xv.	Tony had a drink last night.
vi.	Mia drank 2 cocktails last night.	xvi.	Josh learned to ride a bike yesterday.
vii.	Isabella ate a steak on Sunday.	xvii.	Owen shoveled snow last winter.
viii.	Emily bought a car yesterday.	xviii.	Julian dances salsa.
ix.	Grace visited her sister.	xix.	Jon walks to work.
x.	Zoe calculated the tip.	xx.	Charley speaks Spanish.

C Control stimuli in Exps. 5–6

The control stimuli in Exps. 5–6 were the contents of the main clause polar questions in (5). The non-restrictive relative clauses (NRRCs), given in parentheses in (5), were included in Exp. 6, where at-issueness was measured with an assent diagnostic. The control stimuli here consisted of two clauses (like the target stimuli), to allow the relevant speaker to assent with one of two clauses.

(5) Sentences for control stimuli in in question embedding experiments (Exps. 5–6)
a. Do these muffins (, which are really delicious,) have blueberries in them?
b. Does this pizza (, which I just made from scratch,) have mushrooms on it?
c. Was Jack (, who is my long-time neighbor,) playing outside with the kids?
d. Does Ann (, who is a local performer,) dance ballet?
e. Were John's kids (, who are very well-behaved,) in the garage?
f. Does Samantha (, who is really into fashion,) have a new hat?

We expected participants to give low responses on the at-issueness diagnostics for the control stimuli in (5), indicating that the main clause content is at-issue.