

Challenges for the operationalization of focus-sensitivity in a cross-linguistic context

Deniz Özyıldız, Ciyang Qing, Floris Roelofsen, Wataru Uegaki, Maribel Romero

August 30, 2024

Abstract

A number of analyses of clause-embedding predicates make crucial reference to the notion of *focus-sensitivity*. Whether a predicate is focus-sensitive or not is argued to affect mood selection (Villalta, 2008), homogeneity inferences (Wehbe and Flor, 2022), and compatibility with (different kinds of) interrogative complements (Romero, 2015; Uegaki and Sudo, 2019). This makes it important to be able to decide reliably whether a given clause-embedding predicate, in a given language, is focus-sensitive or not. In this paper, we critically assess the applicability, within and across languages, of three empirical tests for focus-sensitivity: One based on truth value judgments; another one based on judgments of coherence or contradiction; and a third one, our original proposal, based on judgments about the presence or absence of a particular inference. We identify the main limitations of each test, and suggest how they can be used in unison to provide a diagnosis of focus-sensitivity that is reliable, and that minimizes effort on the part of researchers and language consultants.

1 Introduction

For sentences containing certain lexical items and constructions, such as *only* and *even*, adverbs of quantification like *always*, *because*-clauses and counterfactuals, differences in the position of focus, marked in capitals below, can affect their truth-conditions (Jackendoff, 1972; Rooth, 1985, a.o.). What we mean by focus-sensitivity, pending a more precise definition in Section 2, is that the position of focus affects the truth conditions of a sentence.¹ To see this, one can construct a truth value judgment test to detect focus-driven truth-conditional differences: In scenario (1), sentence (1a) is judged true whereas sentence (1b) is not; similarly, scenario (2) makes sentence (2a) true but not sentence (2b) (Rooth, 1996).

- | | | |
|-----|---|----------|
| (1) | Scenario: Mary introduced Bill and Tom to Sue, and did nothing else. | |
| | a. Mary only introduced Bill to SUE _F . | TRUE |
| | b. Mary only introduced BILL _F to Sue. | NOT TRUE |
| (2) | Scenario: In Saint Petersburg, if an officer escorted somebody, it was a ballerina. Ballerinas could be escorted by officers or by opera employees. | |
| | a. In Saint Petersburg, officers always escorted BALLERINAS _F . | TRUE |
| | b. In Saint Petersburg, OFFICERS _F always escorted ballerinas. | NOT TRUE |

This paper is concerned with a class of expressions whose potential focus-sensitivity has been much less explored, namely the class of clause-embedding or attitude predicates. Dretske (1972), for example, notes that the truth of a sentence with an attitude verb like *advise* depends on the location of contrastive focus (Dretske’s words) within the complement clause, as in (3). And one can indeed conceive of contexts in which one of these sentences is true, while the other isn’t.

¹Our discussion of focus is independent of the description of how focus is realized within a given language and other questions, like whether association with focus is conventional or non-conventional (Beaver and Clark, 2007). Some terminological choices we make are as follows: We sometimes call two strings of the form “...A B_F...” and “...A_F B...”, which only differ in the position of focus, the same sentence and sometimes, two different sentences. We use the terms “clause-embedding” and “attitude” verb or predicate interchangeably, and assume that verbs are a subset of predicates, the former including, e.g., adjectival forms like “be glad” or “be surprised.”

- (3) a. Alex advised Clyde [TO SELL HIS CAR]_F to Schulz [FOR \$30000]_F.
- b. Alex advised Clyde to sell his car [TO SCHULZ]_F for \$30000.

For some attitude predicates, it is relatively easy to see which ingredients of the context need to be manipulated to obtain focus-driven truth-conditional meaning contrasts. However, as we will see, it proves to be a difficult task to test for focus-sensitivity in a general way.

Recent work in the study of attitude verbs makes it important to have a test (or tests) for focus-sensitivity that can be constructed easily and that are broadly applicable. (Of course, before one can test for focus sensitivity in any given language, one must independently know how focus is marked in that language to begin with. In this paper, we will set this descriptive issue aside and illustrate with examples from English, which uses prosodic marking of focus. The tests that we survey here should be applicable to any language regardless of how focus is marked, so long as potential confounds—such as additional inferences triggered by clefts—are controlled for.) Such work concerns, on the one hand, what attitude verbs mean, and, on the other, possible correlations between the semantic properties of attitude verbs and their combinatorial properties—that is, whether they combine with declaratives, interrogatives, exclamatives, etc., and which subkinds thereof, e.g., indicative or subjunctive declaratives; constituent, polar, or alternative questions, and so on. What we refer to as *combinatorial properties* is also often referred to as *selectional properties* or *restrictions*—see, for example, Grimshaw (1979).

In terms of the semantics of attitude verbs per se, semanticists strive to uncover their core meaning components in order to offer more refined and descriptively adequate semantic analyses. See, e.g., Djärv (2023) on *know* and *believe*, Roberts (2021) on *believe*, Major (2021) on *say*, Özyıldız (2021) on *think*, White and Rawlins (2018a) on *decide*, among others. Focus-sensitivity is one such meaning component: Some predicates, like *want*, *be glad* or *advise* are described as focus sensitive, whereas ones like *know* or *believe* are usually not. Whether or not a predicate is focus-sensitive or not has to be determined, and reflected in its lexical entry (Dretske, 1972; Villalta, 2008).

It is also important to know whether different properties of attitude predicates interact with one another. This makes it possible to subsume one set of properties under another, or to explain one in terms of another. These interactions may concern different sets of semantic properties or, as mentioned, one set of semantic properties and one set of combinatorial ones. For example, Saebø (2007) and Egré (2008) argue that whether or not attitude predicates select declarative vs. interrogative clauses depends on whether they are veridical or not; and Zuber (1982), Theiler et al. (2019) and Mayr (2019), on whether they are neg-raising or not. Some of the empirical generalizations that this body of work puts forward and the analyses that follow from them also involve focus-sensitivity as a crucial explanatory factor. Here we mention the following three.²

A first generalization, due to Villalta (2008), concerns subjunctive mood in declarative complement clauses:

- (4) In Spanish, all attitude predicates that select subjunctive mood in their complement clause are focus-sensitive. (Villalta, 2008, 495 et seq.)

This generalization is argued to follow from the proposal that focus-sensitive predicates operate on a set of alternatives C to the proposition that they embed (the comparison class), and that the subjunctive makes an appropriate set C available, while the indicative makes any such C unavailable.

A second generalization concerns factive emotive predicates like *surprise* and *disappoint*, in (5).

- (5) Factive emotive predicates allow wh-questions but disallow alternative questions and polar questions as their complements. (Grimshaw, 1979; Lahiri, 1991; Guerzoni, 2007)

Romero (2015) derives this ban with the observation that emotive factives are focus-sensitive, hence operating on a comparison class C , and the proposal that emotive factives' assertion and presupposition impose contradicting requirements on C when they combine with alternative and polar questions.

A third hypothesized generalization concerns non-veridical preferential predicates like *hope*, in (6).

- (6) All non-veridical preferential predicates are anti-rogative, i.e., they do not take interrogative complements. (Uegaki and Sudo, 2019)

²Wehbe and Flor (2022) also find an interaction between focus-sensitivity, homogeneity and declarative vs. interrogative selection, which we do no more than mention, as a proper discussion of homogeneity would lead us far astray.

Here too, Uegaki and Sudo (2019) make use of the idea that non-veridical preferential predicates are focus-sensitive, and, when these predicates take interrogative complements, their asserted content turns out to be trivial given their presupposition.³

To argue for the fact that these classes of predicates are focus-sensitive, the authors cited make use of a truth value judgment test for focus-sensitivity found in Villalta (2008: sec. 7.1). This test involves constructing pairs of sentences like (3) that differ in the placement of focus, which leads to a difference in truth value in the appropriate contexts. We argue in Section 2, however, that this test faces certain limitations. First, constructing it requires knowing in advance certain non-obvious analytical components of the meaning of attitude predicates, e.g., that *want* and *be glad* have a bouletic and a doxastic meaning component (Heim, 1992; von Stechow, 1999), and second, there are predicates, e.g., *believe* and *know*, for which it cannot be constructed at all (for reasons that we outline below). This makes this particular test difficult to apply in a principled and uniform way to any given predicate in any given language.

Yet, the fact that focus-sensitivity is a key component in the semantics of certain attitude predicates, and the central role attributed to it in explaining mood and clause-type selection make it important to have a test for focus-sensitivity that *can* be applied to any given predicate in any given language in a principled and uniform way. Such a test would enable us to corroborate, invalidate or refine generalizations like the three above with data spanning the lexicon of a given language, as well as data from different languages (i.e., going beyond a selection of predicates in English and Spanish). It would also contribute to the possibility of uncovering new generalizations.

Our main goal in this paper is to work towards such a test for focus-sensitivity that can be uniformly applied to any attitude predicate in any language. To this end, we consider three different ways of testing for focus-sensitivity, which Sections 2, 3, and 4 are respectively dedicated to. Two of these come from Villalta (2008) and have been used elsewhere in the literature: One relies on truth value judgments, and the other on judgments of coherence and contradiction. The third one, inspired by an observation in Dretske (1972: p. 415 et sq.), is based on judgments of inference. We highlight potential confounds that these three tests for focus sensitivity might face, and the difficulties that arise while running them. We conclude that none of these tests are fully satisfactory on their own, and recommend that the inference-based test should be used in tandem with Villalta’s truth value judgment test.

2 A truth-based test

Original formulation of the truth value judgment test Our starting point is a truth value judgment task used as a focus sensitivity test by Villalta (2008: sec. 7.1), as well as Romero (2015) and Uegaki and Sudo (2019). To test whether a given clause-embedding predicate is focus sensitive, two string identical sentences are constructed, which differ only in the position of a narrowly focused constituent in the embedded clause. Example (7) illustrates two target sentences for ‘want.’⁴

³White (2021) uses large-scale experimental data and corpus examples to challenge the validity of several existing generalizations concerning the relation between the semantic and selectional properties of a clause-embedding predicate. In the case of Uegaki and Sudo’s generalization, White only provides several putative counterexamples from corpora as there is no large-scale dataset measuring focus sensitivity (presumably due to the lack of a general test for focus sensitivity).

⁴In (7), the constituent “on Tuesdays and Thursdays” is also where focus projects from, and where main sentential stress is expected to fall in broad(er) focus contexts (Selkirk 1996, a.o.). That is, the particular realization of (7a) could be used felicitously to answer questions like (ia) and (ib), in addition to (ic), where its correspondence with the *wh*- phrase is a way of identifying the size of a focused constituent.

- | | | |
|-----|--|---------------------------------|
| (i) | a. What does Lisa want John to do? | elicits embedded VP focus |
| | b. What does Lisa want? | elicits embedded sentence focus |
| | c. When does Lisa want John to teach syntax? | elicits embedded PP focus |

For uncontextualized realizations of (7a), this makes it difficult to tell when “on Tuesdays and Thursdays” is under narrow focus vs. when it is part of a larger focused constituent. Villalta does not address this confound, and we will not fully do so either. But, to the best of our ability to tell, placing main sentential stress on a constituent where focus cannot project from, does not affect the results that we report on. Two such configurations are provided in (ii), neither of which can be used to answer the broader focus eliciting questions in (ia) and (ib) without generating additional inferences.

- | | |
|------|--|
| (ii) | a. Lisa wants John to teach SYNTAX on Tuesdays and Thursdays. |
| | b. Lisa wants John to teach syntax ON TUESDAYS AND THURSDAYS in Murray Hall. |

- (7) a. Lisa wants John to teach syntax ON TUESDAYS AND THURSDAYS.
 b. Lisa wants JOHN to teach syntax on Tuesdays and Thursdays.

The truth of these sentences is judged in a context that constrains their relevant focus alternatives such as (8a), and that specifies relevant aspects of the attitude that the matrix subject bears to the embedded proposition, e.g., their preferences and beliefs, in (8b). The context in (8a) and (8b) is the one given by Villalta for the non-veridical preferential ‘want,’ which we separate into ‘broad’ and ‘immediate’ for expository purposes.⁵

- (8) a. **Broad context**
 In the linguistics department, at the faculty meeting, the teaching schedules of the different faculty members for the upcoming semester are discussed. There is only one syntactician in the department (John), one phonologist (Lisa), and two semanticists (Lara and Frank). John can only teach syntax. Lara can teach syntax and semantics. There is some controversy on which days John should teach his syntax classes. There are two options: he may teach syntax on Tuesdays and Thursdays, or he may teach syntax on Mondays, Wednesdays and Fridays.
- b. **Immediate context**
 Lisa’s preferences are the following: she would prefer it if Lara would teach syntax rather than John. But given that John has to teach syntax, she prefers it if he teaches on Tuesdays and Thursdays rather than on Mondays, Wednesdays and Fridays (because she wants the teaching slot on Mondays, Wednesdays and Fridays for her own phonology class, which cannot conflict with the syntax class).
- c. **Target sentences**
 (i) Lisa wants John to teach syntax ON TUESDAYS AND THURSDAYS. TRUE
 (ii) Lisa wants JOHN to teach syntax on Tuesdays and Thursdays. NOT TRUE

In the context given, (8ci) is judged true, while (8cii) is judged not to be true.⁶ This suggests that the position of focus under the scope of ‘want’ gives rise to truth conditional differences, which motivates the conclusion that ‘want’ is a focus sensitive predicate. (We will shortly see predicates under which the position of focus does *not* give rise to truth conditional differences.)

Predicates with different properties will require that the context be modified in different ways. For example, for the factive preferential ‘be glad,’ we have to assume, citing Villalta, that “at the end of the faculty meeting, *it is decided that* John is indeed going to teach syntax on Tuesdays and Thursdays” (compare with “There is some controversy on which days...” above). This is to satisfy the requirement that the complement of ‘be glad’ is presupposed.

- (9) a. Lisa is glad that John teaches syntax ON TUESDAYS AND THURSDAYS. TRUE
 b. Lisa is glad that JOHN teaches syntax on Tuesdays and Thursdays. NOT TRUE

And, in practice, the contexts, the target sentences and possibly other task-related information involved in the tests presented in this paper would all seem to bias consultants into narrow focus construals. The context in (8), for example, explicitly contrasts two sets of days, which evokes narrow focus alternatives to sentences with sentential stress on “on Tuesdays and Thursdays.”

⁵A veridical predicate entails, and a factive predicate presupposes its propositional complement. We subscribe to the view that emotive predicates like ‘be glad’ presuppose their complement in ordinary contexts, as opposed to merely presupposing that their subject *believes* the complement (Klein 1975, a.o.). This is relevant because it affects whether the embedded proposition should be made true by the context given or not. This original context does not explicitly state that it is decided that John will teach syntax and that Lisa knows this, but this is implicit in “But given that John has to teach syntax, she prefers it if...” In other words, the context is such that Lisa truthfully believes that John will teach syntax. It is not required that the belief be truthful for ‘want,’ but for factive predicates like ‘be glad,’ it is.

⁶For these examples, Villalta does not say “false” but rather “not true,” presumably in order to leave room for the possibility that the non-true sentences could come out as undefined. Romero (2015) proceeds similarly for “surprise” and “be happy.” In a refinement, Harner (2016) draws a distinction between semantic and pragmatic focus sensitivity, where the position of focus makes a difference between true and false in the former, and between felicitous and infelicitous in the latter. Relying on this distinction, Uegaki and Sudo (2019) explain the anti-roгатivity of non-veridical preferential predicates through their semantic focus sensitivity, and exclude predicates like “decide” based on the commitment that these are pragmatically focus sensitive. For simplicity, this paper only concerns the basic distinction between predicates that are focus sensitive and those that are not, and therefore we do not break down “not true” responses into “false” and “infelicitous” responses. However, the test can be straightforwardly extended to incorporate a finer-grained distinction between falsity and infelicity in the responses, as long as the consultant can reliably distinguish between the two (though see Matthewson, 2004; Deal, 2015).

Given Lisa’s preferences as specified in (8b), here too, (9a) is judged true and (9b) not true, motivating the conclusion that ‘be glad’ is a focus-sensitive predicate.

Villalta contrasts the situation with ‘want’ and ‘be glad’ with the factive doxastic predicate ‘know.’ About the pair of sentences in (10), she writes that “[they] are both true under the same circumstances: all contexts that make one of them true make the other true as well.” (That is, each sentence entails the other, which amounts to saying that the sentences are truth conditionally equivalent.)

- (10) a. Lisa knows that John teaches syntax ON TUESDAYS AND THURSDAYS.
 b. Lisa knows that JOHN teaches syntax on Tuesdays and Thursdays.

Because the position of focus under the scope of ‘know’ is not perceived to give rise to a truth conditional difference, the predicate is classified as non-focus-sensitive. A larger list of predicates, classified in terms of whether they are focus sensitive or not according to Villalta’s judgment, is found in the appendix of her paper. Importantly though, no sample context is provided by Villalta (2008) for any of the verbs that were judged non-focus-sensitive.

In the remainder of this subsection, we will explore how such contexts may be constructed. To this end, we will first dissect Villalta’s original context for *want* into its crucial ingredients. Then, we will seek parallel ingredients for *believe/know* and build the corresponding, extended Villalta-style contexts. Finally, we will return to *want* and show that, if we employ the extension recipe we use for *believe/know*, *want* will incorrectly be judged non-focus-sensitive. The lack of general applicability of this truth-value judgment test will lead us to the proposal that it should be supplemented by other tests for focus sensitivity.

Key characteristic properties of the truth value judgment test Our main goal in this paper is to provide a general recipe for constructing a focus sensitivity test that can be applied to any predicate in any given language. To probe whether Villalta’s truth value judgment task is a candidate for such a test, we would like to know whether it is possible, and what it would take to construct contexts for predicates like ‘know,’ ‘believe,’ and the like, which are parallel to what we are given for ‘want’ and ‘be glad.’ The reason that we are interested in constructing contexts is not merely to extend Villalta’s test by making as few changes as possible, but also because the task requires a linguistic consultant to give a pair of truth value (and felicity) judgments, which cannot be accessed for sentences presented out of context (Matthewson, 2004). Thus, we will identify the key properties that characterize the contexts given by Villalta for ‘want’ and ‘be glad,’ generalize over them, and attempt to extend the task to new predicates.

Before discussing these properties, we provide the following working definitions of focus sensitivity at the sentential and sub-sentential levels (11). In particular, the definition of focus sensitivity for a sub-sentential expression E (11b) ensures that E counts as focus sensitive only when it properly contributes to the focus sensitivity of a sentence that contains it. These definitions help understand why the truth-based test (and other tests we will discuss in later sections) has the formal properties that it has.

- (11) a. **Definition of a focus-sensitive sentence**
 A sentence S is focus sensitive iff there exist a context C and a sentence S' that is only different from S in terms of the placement of focus, such that S and S' have different truth values in C .
 b. **Definition of a focus-sensitive sub-sentential constituent**
 A sub-sentential expression E is focus sensitive iff there exists a focus-sensitive sentence S containing E whose focus sensitivity is caused by E rather than some other independent factor(s).

Based on these definitions, we can derive a more explicit definition of a focus-sensitive clause-embedding predicate (12).

- (12) **Definition of a focus-sensitive clause-embedding predicate**
 A clause-embedding predicate P is focus sensitive iff there exist a context C and two clauses S and S' that are only different in terms of the placement of focus such that (i) $\lceil x Ps S \rceil$ and $\lceil x Ps S' \rceil$ have different truth values in C and (ii) the difference in the truth values cannot be attributed to factor(s) independent from the use of P .

With this definition at hand, we now discuss the properties of the contexts required by the truth-based test. One general assumption that is (tacitly) present in the context provided by Villalta is that there is only

one syntax class offered in the semester, that it is taught by only one instructor and on one set of days (as is usually the case in linguistics departments). This means that specifying one instructor name and one set of days is enough to fully resolve the question of who will teach syntax when. To see why this is important, let us briefly assume that there are *two* syntax classes, one taught by Lara for senior graduate students and one by John for incoming graduate students, and that both classes occur on Tuesdays and Thursdays. In this context, (13a) is true, but (13b) is infelicitous.

- (13) **Context:** It is common knowledge that one syntax class is taught by Lara and a second syntax class is taught by John, and that both classes occur on Tuesdays and Thursdays.
- a. Lisa knows that John teaches syntax ON TUESDAYS AND THURSDAYS.
 - b. #Lisa knows that JOHN teaches syntax on Tuesdays and Thursdays.
 - c. Lisa knows that JOHN AND LARA teach syntax on Tuesdays and Thursdays.

The reason behind the infelicity of (13b) is that focus on *John* strongly invites a reading under which the sentence is taken to be an exhaustive answer to the question “Which individual(s) x are such that Lisa knows that x teaches syntax on Tuesdays and Thursdays?” and it is not. In contrast, the sentence in (13c) with both instructors named and narrowly focused is felicitous (and true).

Not controlling for exhaustivity effects, then, is a potential confound for the truth value judgment test for focus sensitivity. To avoid it, the test needs to be formulated against an *exhaustivity neutral context*. Failing to do so could give rise to contrasts like the one between (13a) and (13b) that are indeed driven by the position of focus, but lead to the wrong conclusion that a predicate like ‘know’ is focus sensitive. In general, this is a risk of false positives, i.e., of over-classifying predicates that are not focus sensitive as focus sensitive.

Unless stated otherwise, all of the contexts provided in this paper are exhaustivity neutral. Concretely, they always involve the background assumption that any given class has only one instructor, and that it occurs on only one day of the week. We will also simplify ‘Tuesdays and Thursdays’ to simply ‘Tuesdays,’ as this is shorter and less dependent on any knowledge of how US linguistics departments work (except when we are directly citing Villalta’s examples in the next section).

With this in mind, we are ready to move on to a second defining characteristic of Villalta-style contexts, namely, that they are built on the two necessary requirements in (14).

- (14) I. **The conflicting attitude requirement:**
- Lisa has a negative preference for John being the teacher.
(But having no preference should also suffice.)
 - Lisa has a positive preference for the class to be taught on Tuesdays.
- II. **The substrate attitude requirement:**
Lisa believes that John will be the teacher (truthfully, at least in the case of ‘be glad’).

In the case of ‘want’ and ‘be glad,’ and preferentials in general, the *conflicting attitude* requirement references the main semantic component of the attitude verb: The agent’s preferences. The *substrate attitude* requirement references belief. We believe that this is due to the fact that preferential predicates induce a preference-based ordering on a modal base consisting of an agent’s doxastic alternatives (Heim, 1992; von Stechow, 1999). Intuitively, the proposition desired is the ordinary semantic value of the embedded clause (e.g., “that John will teach syntax on Tuesdays”) and the proposition believed corresponds roughly to the great union of a contextually restricted subset C of the focus semantic value of the embedded clause (e.g., “that John will teach syntax on some days of the week”).

Let us briefly see that the substrate attitude requirement is necessary. Assume the background of the linguistics department faculty meeting from (8), but that it has not been decided who will teach syntax. The latter assumption implies that Lisa has no particular beliefs about who will teach syntax. As before, she has a dispreference for John, and a preference for the Tuesday slot.

- (15) a. **Broad context**
There are two people qualified to teach syntax: John and Lara. There is some debate as to who will in fact teach the course, but at the end of the meeting, no decision has been reached.
- b. **Immediate context**

Lisa would prefer it if Lara teaches syntax, and she would prefer the class to be held on Tuesdays. It is not the case that Lisa believes that John will teach syntax.

c. **Target sentences**

- | | | |
|------|--|----------|
| (i) | Lisa wants John to teach syntax ON TUESDAYS. | NOT TRUE |
| (ii) | Lisa wants JOHN to teach syntax on Tuesdays. | NOT TRUE |

In this context, neither of the sentences in (15c) are true. In particular, for (15ci), the inclusion of ‘John’ in the embedded clause conflicts with the assumption that it hasn’t been decided yet who will teach syntax, as there are other people teaching syntax in Lisa’s doxastic alternatives. One would have to use a definite description (‘the person who will teach syntax’) or an impersonal construction (‘Lisa wants syntax to be taught on Tuesdays’) instead.

A final general remark about these two requirements is that they need to be filled in by statements that are consistent with each other. (In practical terms, the substrate attitude statement needs to be weaker than or independent from the non-positive conflicting attitude statement.) In the case of preferentials, consistency is possible because the conflicting attitude and the substrate requirements involve different modalities, respectively bouletic and doxastic, and having a negative preference for John to be the teacher does not contradict the belief that he will be. We will explore different ways of satisfying consistency without appealing to a difference in modality in our discussion of “believe” and “know” below.

This leads us to the following methodological guideline for generalizing Villalta’s focus sensitivity test and applying it to new predicates. First, for a given predicate P one must construct a context that satisfies the conflicting attitude requirement and the substrate attitude requirement. This will involve choosing what these attitudes will be. For some predicates like ‘want,’ the choice might be straightforward, but for others, like ‘believe,’ initial choices might result in a contradiction. (We will shortly see why.) However, given that the conflicting and substrate attitudes may be chosen arbitrarily, it should always be possible to avoid contradiction, should one arise, and construct a context without one. With this context, the next step is to run a truth value judgment task on a pair of sentences, schematized in (16), which minimally differ in the position of focus in the embedded clause. Important is that one of the focused constituents must be the target of the positive attitude (e.g., “on Tuesdays” above).

- (16) a. x Ps that [...X...Y_F...]
 b. x Ps that [...X_F...Y...]

If the first sentence is judged true, but the second is not true, then the predicate P is focus sensitive. If the position of focus does not make a difference in truth value, then we fail to conclude that the predicate is focus sensitive—to be able to draw the stronger conclusion that a predicate is not focus sensitive, one must exhaust all possible contexts. Exhausting all possible contexts is technically necessary because the failure of one context to bring out a truth conditional difference between the two sentences does not imply that we will find no context in which a truth conditional difference will arise. Thus, the best we can hope for given this test is that we will *fail to find evidence* that some predicates are focus-sensitive (rather than being able to find evidence that they are not focus-sensitive).

Extending the test to ‘believe’ and ‘know’ Let us now return to the doxastic predicates ‘believe’ and ‘know.’ We are now in a position to try to extend Villalta’s focus sensitivity test to these predicates in a conservative manner. We first attempt to identify the attitude that is relevant for the conflicting attitude requirement: In the case of preferential predicates, the attitude was ‘having a preference.’ Let us then assume that in the case of doxastic predicates, it is doxastic—namely, the attitude expressed by the verb. This yields (17) and (18). (In the case of (18a), we assume that the factive inference usually associated with ‘know’ is not projected, and read the assumption as “Lisa has no knowledge one way or the other.”)

(17) **The conflicting attitude requirement for ‘believe’**

- | | |
|----|---|
| a. | Lisa neither believes that John will teach syntax, nor that he will not. ⁷ |
| b. | Lisa believes that syntax will be taught on Tuesdays. |

⁷Unopinionatedness in the doxastic case (cf. lack of preference in the preferential case) is the weakest assumption that will allow us to run the test. One could alternatively assume the stronger belief that John will not teach syntax (cf. dispreference).

(18) **The conflicting attitude requirement for ‘know’**

- a. Lisa neither knows that John will teach syntax, nor that he will not.
- b. Lisa knows that syntax will be taught on Tuesdays.

Now, the question is what the substrate attitude might be. In the case of preferentials, we had predicates that made reference to both an agent’s desires and their beliefs, and the substrate attitude was doxastic. In the case of ‘believe’ and ‘know,’ this option is at first sight unavailable: These predicates respectively only make reference to an agent’s doxastic and epistemic states. For concreteness, if, for ‘believe,’ we took the substrate in (19), a contradiction would arise with (17a), making it look like a Villalta-style context cannot be constructed.

(19) **Unreasonable candidate substrate attitude requirement for ‘believe’**

Lisa believes that John will be the teacher. Contradicts (17a)

One could stop here and declare that ‘believe’ and (changing what needs to be changed) ‘know’ are not focus sensitive (or rather, that we fail to find evidence for their focus sensitivity), but that conclusion, while correct, is too hasty.

Omitting the substrate requirement would resolve the issue, but example (15) has shown that it is necessary for Villalta’s test to work, and omitting it would make the comparison between doxastic and preferential predicates non-minimal. Instead, we explore the possibility of finding suitable substrate attitudes for doxastic predicates that do not give rise to a contradiction with either one of the attitudes from the conflicting attitude requirement. Our guiding intuition stems from the observation that ‘know’ and ‘believe’ quantify over different information states: What an agent knows, and what they believe. We now ask whether it makes sense to make the following assumption to satisfy the substrate attitude requirement with ‘know.’

(20) **Reasonable candidate substrate attitude requirement for ‘know’**

Lisa *believes* that John will be the teacher.

Notice here that this assumption is consistent with Lisa having no *knowledge* that John will be the teacher, in (18a). If we allow for this possibility for ‘know,’ there is no reason not to allow it for ‘believe.’ We simply take a notion of commitment to a proposition, in (21), that is weaker than belief, namely, considering that something is more likely than something else.

(21) **Reasonable candidate substrate attitude requirement for ‘believe’**

Lisa *considers it slightly more likely* that John will be the teacher than anyone else.

Note, again, that this assumption is compatible with the assumption that the agent is unopinionated (has no belief one way or another) with respect to whether John will be the teacher, in (17a). (A similar exercise can be done for other doxastic predicates like ‘be certain.’)

On the basis of these assumptions, we can proceed to construct Villalta-style contexts for ‘know’ and ‘believe,’ and perform a truth value judgment task on the sentences in (22b) and (23b).

(22) **Villalta-style context for ‘know’**

a. **Immediate context**

At the faculty meeting, it is decided that syntax will be taught on Tuesday. The meeting ends before a decision can be made about *who* will teach the course. However, John has usually been volunteering to teach syntax in recent years, and Lisa believes that he will do it again.

b. **Target sentences**

- (i) Lisa knows that John will teach syntax ON TUESDAYS. NOT TRUE
- (ii) Lisa knows that JOHN will teach syntax on Tuesdays. NOT TRUE

With ‘know,’ neither one of these sentences is judged to be true. The intuition is that they are odd, because it is not certain that John will teach syntax—this has not been publicly decided yet—and that the verb’s factive presupposition is therefore not satisfied. In addition to this, to the extent that one can interpret these sentences charitably, they ascribe to Lisa a level of certainty that goes beyond what is licensed in context: Perhaps John has told her that he fully intends to teach syntax, and this licenses us to say that she knows it. But, not only is this a way of extending the context as a repair strategy, but also, even with such an

extension, the position of focus does not have a truth conditional effect: To the extent that the sentences become acceptable, they are both true.

Let us now turn to ‘believe.’

(23) **Villalta-style context for ‘believe’**

a. **Immediate context**

At the faculty meeting, it is decided that syntax will be taught on Tuesday. The meeting ends before a decision can be made about *who* will teach syntax. Lara and John can both teach the course and neither has signed up to teach anything yet. Lisa knows that John enjoys teaching more than Lara, and suspects that it’s slightly more likely that he will sign up rather than her.

b. **Target sentences**

- (i) Lisa believes that John will teach syntax ON TUESDAYS. NOT TRUE
- (ii) Lisa believes that JOHN will teach syntax on Tuesdays. NOT TRUE

Here too, neither one of these sentences are true.⁸ The intuition is that both ascribe to Lisa a degree of commitment to the proposition that John will teach syntax that is greater than the one warranted by the context. (The reader concerned with the possibility that ‘believe that p’ is less natural in contexts where p is true is invited to modify the context in such a way that it is still unclear, but likely, that syntax will be taught on Tuesdays.)⁹

The interim conclusion here is that even under our best attempts to bring out the focus sensitivity of ‘believe’ and ‘know’ by using the test adapted from Villalta, this result does not follow: We fail to find evidence that these two predicates are focus sensitive. This is a point in favor of this test.

A reassessment of test contexts for ‘want’ We have seen that our ability to construct the relevant contexts for Villalta’s focus sensitivity test depends on our ability to determine conflicting and substrate attitudes. In the case of preferentials, our choice was tacitly guided by the knowledge that these verbs reference both an agent’s desires and their beliefs. In the case of doxastics, no natural way of satisfying the conflicting and the substrate attitude requirements was available, and we allowed ourselves some flexibility and arbitrariness in determining the substrate attitude. Returning to preferentials, this opens up the possibility of instantiating their substrate attitude requirement with an attitude different from belief. We illustrate an alternative here, and show that the consequence of such a manipulation is that a predicate like ‘want’ no longer comes out as focus sensitive.

Let us instantiate the conflicting attitude requirement for ‘want’ in the way that we have already seen in (14). Instead of picking the belief that John will teach syntax as the substrate attitude, however, let us say that this is instantiated by Lisa not having a preference one way or another regarding whether John teaches syntax.

(24) a. **Conflicting attitude requirement for ‘want’**

- (i) Lisa has a negative preference for John teaching syntax.
- (ii) Lisa has a positive preference for syntax to be taught on Tuesday.

b. **Alternative, non-doxastic substrate attitude for ‘want’**

Lisa doesn’t care whether or not John teaches syntax.

⁸We thank an anonymous reviewer for noting that, in their judgment, both of these sentences are true (independently of the confounds discussed in the next footnote). This might be because some speakers find that considering *p* more likely than $\neg p$ is sufficient for believing that *p*. For such speakers, this would then contradict the first conflicting attitude, namely that Lisa is not opinionated with respect to whether John will be the teacher or not. While the reviewer’s judgments can still lead to the conclusion that we fail to find evidence for focus sensitivity in the case of *believe*, the immediate context for this reviewer is no longer a Villalta-style context in the way that we have characterized them, since the conflicting attitude requirement does not hold. Therefore they do not affect our general point that there is no general way here to construct a relevant Villalta-style context.

⁹There are two ways of reading (22bi) and (23bi) as *true* in contexts, like the ones given, in which Lisa has no knowledge or belief with the content that John will teach syntax. First, the sentence might be reinterpreted with an embedded conditional: “Lisa knows/believes that *if John teaches syntax*, he will teach it on Tuesdays.” Second, it is in principle possible to read “John” *de re* and access truth conditions of the form “Lisa knows/believes that the person who will teach syntax will teach it on Tuesdays.” This is not an issue in the contexts given here as the *speaker* would not assent to the proposition that John will teach syntax either, as this has not been decided yet. These effects are not driven by focus placement, but they are issues that one might encounter while testing for focus sensitivity and would have to control for.

The observation, of course, is that this choice of a substrate attitude contradicts Lisa’s negative preference in (24a). Consequently, if we had picked ‘not care’ as the substrate attitude for ‘want’ and had not considered any alternatives, we would have had to conclude that it is not possible to construct a Villalta-style context for the predicate. This, in turn, would have led to the conclusion that ‘want’ is not focus sensitive—or, more accurately, to the conclusion that we have failed to find evidence that it is. This is an undesirable result.

The contradiction arises here because we chose negative preference as one of our conflicting attitudes. If we assume instead that the first conflicting attitude is that Lisa has no particular preference for John, like we did for ‘know’ and ‘believe,’ no contradiction would arise. But then, the substrate attitude, being equivalent to that conflicting attitude, becomes redundant. And we have seen, in (15), that a distinct substrate attitude was required for the test under discussion to work. Let us now take the same substrate attitude as the one for ‘believe’ instead, stated in (25).

- (25) **Alternative doxastic substrate attitude for ‘want’**
 Lisa considers it slightly more likely that John will teach syntax.

We are now in a position to construct a context for ‘want’ that doesn’t involve contradiction.

- (26) a. **Immediate context**
 Lisa has a negative preference for John teaching syntax, and a positive preference for the class to be held on Tuesdays. It is not yet decided who will teach the class, but Lisa knows that John enjoys teaching more than Lara, and suspects that it’s slightly more likely that he will sign up rather than her.
- b. **Target sentences**
- | | | |
|------|--|----------|
| (i) | Lisa wants John to teach syntax on TUESDAYS. | NOT TRUE |
| (ii) | Lisa wants JOHN to teach syntax on Tuesdays. | NOT TRUE |

In this context, however, the position of focus does not affect the truth value of the sentence: both (26bi) and (26bii) are false. In particular, (26bi) feels false because it ascribes to Lisa a preference for John, which contradicts our assumption that she has a negative preference for him being the teacher.¹⁰ What we see here is, again, that with these particular instantiations of the conflicting and the substrate attitude requirements, we would fail to find evidence that ‘want’ is focus sensitive—an undesirable result.

Conclusions on the truth value judgment based test In this section, we have seen how Villalta’s original focus sensitivity test works for preferential predicates and how it may in principle be extended to other predicates. Such extensions require the identification of a pair of conflicting attitudes and a substrate attitude that are of a distinct from but consistent with one another, e.g., a pair of desires and a belief. These are the building blocks of a context in which a truth value judgment test can be run on a pair of sentences that differ only in focus position. A difference in truth value indicates that a predicate is focus sensitive; no difference indicates a lack of evidence for focus sensitivity.

The conflicting attitude requirement and the substrate attitude requirement were instantiated naturally in the case of preferential predicates, but testing the focus sensitivity of doxastic predicates required making arbitrary decisions about what the substrate attitude should be. Even though the test correctly identifies ‘want’ and ‘be glad’ as focus sensitive and its extension classifies ‘believe’ and ‘know’ as non-focus sensitive, there are several reasons for looking for an alternative diagnostic for focus sensitivity.

Whether the result of the test can be trusted relies on the identification of the right substrate attitude: If we pick belief, preferentials come out as focus sensitive, if we pick desire (‘not care’) or a form of doxastic commitment that is weaker than belief (‘consider it likely’), the same predicates do not come out as focus sensitive. We do not want the result and the reliability of a test to depend on such a choice. Although it may be impossible to construct a linguistic test that is completely free from arbitrary decision points, it is desirable to have a test that invokes a minimal amount of arbitrary choices per predicate. Moreover, there is no reason to expect that we should know *in advance* and *in a general way* which candidates should be

¹⁰This intuition is somewhat surprising because we have seen cases where “John” did not need to be the target of a positive preference when it was unfocused. We speculate that in sentences like ‘S wants p’ one can only escape from x’s preferences those ‘parts’ of S that x believes.

considered for the substrate attitude associated with a given predicate in a given language—especially when the researcher is not a native or competent speaker of the language that they are investigating.¹¹

3 A coherence-based test

The second test for focus sensitivity involves judging whether a piece of dialogue is coherent or contradictory. We will refer to this test as the *coherence-based test*.

Villalta (2008: 497–498) introduces what we will call a *minimalistic* variant of the coherence-based test. Her examples illustrate a contrast between *know* and *want*. For *know*, she points out that B’s reply to A in (27) is contradictory.

- (27) A Lisa knows that John teaches syntax ON TUESDAYS AND THURSDAYS.
 B No, that’s not true. Lisa knows that JOHN teaches syntax on Tuesdays and Thursdays.

This is in contrast with *want*, for which B’s reply in (28) is judged to be not contradictory.

- (28) A Lisa wants John to teach syntax ON TUESDAYS AND THURSDAYS.
 B No, that’s not true. Lisa wants JOHN to teach syntax on Tuesdays and Thursdays.

In general, if a predicate P is focus sensitive, given the definition (12), there exist a context and two embedded clauses S and S' that only differ in the placement of focus such that $\lceil x Ps S \rceil$ and $\lceil x Ps S' \rceil$ have different truth values in C . Therefore, it should be coherent for one to deny the former and assert the latter instead.

However, not all speakers find (28) coherent. In fact, Villalta (2008:498, fn. 11) acknowledges herself that “[f]or some speakers, a more explicit context is necessary to make this a natural dialogue.” This motivates her to further introduce another variant of the coherence-based test, which she attributes to Jenny Doetjes. We will call it the *naturalistic* variant. In this variant, B’s reply is elaborated with a more explicit context, which we highlight in bold (29).

- (29) B: Well, that’s not really true, **as she doesn’t mind which day these classes take place, as long as John is the one who does the teaching**, so one should rather say that Lisa wants JOHN to teach syntax on Tuesdays and Thursdays.

While this additional contextual information makes the dialogue more natural and more clearly consistent, we note that it is predicate specific. Consider *decide* for instance. What would be an appropriate context to include in the dialogue? A simple-minded substitution based on (29) would be non-sensical (30).

- (30) A: Lisa decided that John would teach syntax ON TUESDAYS AND THURSDAYS.
 B: Well, that’s not really true, **#as she didn’t decide which day these classes would take**

¹¹A reviewer suggests that in general one could take (the conjunction of) the presuppositions of $\lceil x Ps A B_F \rceil$ as the substrate attitude. In the case of *want*, $\lceil x wants A B_F \rceil$ presupposes that x believes A , which is indeed the substrate attitude used in Villalta’s context. The reviewer also suggests a general formulation of the conflicting attitude requirement, according to which the conflicting attitude requirement for $\lceil x Ps A B_F \rceil$ is that (i) $\lceil x Ps A \rceil$ is not true and (ii) $\lceil x Ps B \rceil$ is true. However, as the reviewer noted themselves, in the case of *believe*, given that $\lceil x believes A B_F \rceil$ presupposes that x believes A , the substrate attitude according to this recipe would be that x believes A , but this is not in fact applicable since it contradicts the first half of the conflicting attitude requirement (i) as formulated by the reviewer. Also, as discussed earlier, this attitude is also not an applicable substrate attitude under our formulation of the conflicting attitude requirement (17). In light of this, the reviewer speculates that the test may be amended so that it classifies a predicate as not focus sensitive when the substrate attitude constructed according to this recipe contradicts the conflicting attitude requirement.

While we think the reviewer’s general recipe to formulate the conflicting attitude requirement and the substrate attitude is certainly worth exploring further, at the moment we do not know whether it will work for all focus sensitive predicates. That is, it remains an empirical hypothesis whether all focus sensitive predicates will exhibit focus sensitivity in contexts constructed according to this recipe.

Moreover, since there is no general guarantee that the conflicting attitude requirement and the substrate attitude constructed according to the recipe are compatible, the researcher will need to ask the consultant whether this is the case for the specific predicate under investigation. That is, the researcher first asks the consultant whether it is possible to construct a context that satisfies certain requirements, and if so, the researcher then applies the truth-based test using such a context. In this respect, the procedure here is very similar to our overall recommendation in section 4.3, and is very different from the truth-based test (on its own) since the latter does not involve eliciting a context from the consultant.

place, as long as John would be the one who does the teaching, so one should rather say that Lisa decided that JOHN would teach syntax on Tuesdays and Thursdays.

Of course, we should not conclude from the infelicity of (30) that *decide* is not focus sensitive, because we can make the dialogue coherent by using a different context (31).

- (31) B: Well, that’s not really true, **as she didn’t decide which day these classes would take place — that part had been determined long ago by someone else**, so one should rather say that Lisa decided that JOHN would teach syntax on Tuesdays and Thursdays.

Note, however, that the second half of the context needs to be adjusted in a predicate-specific way. Therefore, the naturalistic variant of the coherence-based test faces the same limitation as the truth-based test discussed above, i.e., there is no general recipe to construct the relevant contexts.¹²

Moreover, in our experience, even the “naturalistic” variant is not so natural after all. Even for canonical focus-sensitive predicates such as *hope* and *want*, for which we can directly use relevant contexts in the literature, our consultants often found the dialogues complicated and artificial. Of course, it may well be that we were facing this difficulty because we were not using the most natural dialogues. But this in fact illustrates another limitation of the test, i.e., it is difficult to know what dialogues would be good to use. What counts as a good dialogue will probably differ from language to language. In addition, because the naturalistic variant requires an explicit context in the dialogue, such a context needs to be properly translated into the target language and integrated into the dialogue. This will likely involve language-specific adjustments that will reduce the uniformity of the test across languages.

Finally, the coherence-based test involves a denial in the dialogue. This raises a general concern about what the denial is doing and whether it is doing the same thing across languages. This is particularly so for the minimalistic variant in (27). The consultant might judge the dialogue consistent by imagining a context where Lisa has a preference for John to teach syntax as well as a preference for it to be on Tuesdays and Thursdays, but the Question Under Discussion (QUD, à la Roberts 2012) is about Lisa’s preference about who will teach syntax on Tuesdays and Thursdays rather than her preference about when John will teach syntax. In such a context, B’s response can be seen as a metalinguistic move to correct A’s placement of narrow focus just to ensure that it is congruent with the QUD, and therefore it does not provide evidence that the predicate *want* is focus sensitive. While this possibility might not be available in the English version of (27) because B’s response uses *that’s not true*, there is no guarantee that the target language has such a truth predicate, or that it is natural to use it in a denial. For instance, if in the target language it is only possible to use a counterpart of *that’s incorrect* in English, then B’s response can be more easily judged consistent under the metalinguistic correction interpretation above. As a result, all predicates may **appear** to be focus sensitive due to the general constraint on focus-question congruence, which is not the kind of focus sensitivity we are interested in (recall our definition in (12)). Of course, one can always ask consultants to rule out such metalinguistic corrections when judging the consistency of the resulting dialogue. However, the distinction can be quite subtle, and is likely to make the results less comparable crosslinguistically. For the naturalistic variant (31), the concern that B’s correction is about a general felicity condition on focus-QUD congruence is alleviated because the additional context makes it clearer that the correction rather has to do with the meaning of the specific predicate *decide*. But note that B’s response uses explicitly metalinguistic comments: *well, not really* and *one would rather say*. While such expressions can help make the dialogue more natural, they may well introduce their own idiosyncratic effects that reduce the crosslinguistic uniformity of the test. Therefore a test that avoids such confounding factors altogether would be better.

In sum, the coherence-based test probes for focus sensitivity by checking whether it is consistent to deny a sentence with narrow focus in one position and assert a string identical sentence with narrow focus in a different position.¹³ However, it is difficult to construct good dialogues (let alone creating a general recipe for

¹²In principle, one could only include the first half of the context, for which there is indeed a general recipe, i.e., *not P which day these classes take place* (modulo the need for paraphrases when P does not embed questions). However, this would make the dialogue less natural and does not avoid the limitations we will discuss below.

¹³In fact, a different test based on this idea would be to directly test for the consistency of sentences such as *Ann wants Lisa to teach syntax on TUESDAYS, but she doesn’t want LISA to teach syntax on Tuesdays*. Such a test also faces the concerns that the sentence may not sound very natural and that negation might be used meta-linguistically.

Additionally, for a neg-raising predicate *P*, *Ann Ps φ* and *Ann does not P φ'* (where φ and φ' only differ in the placement of narrow focus) is likely interpreted as $P\varphi \wedge P\neg\varphi'$, which is a stronger claim than the intended $P\varphi \wedge \neg P\varphi'$. Therefore, it is

constructing good dialogues for any given predicate in any given language) that sound natural and guarantee that the denial is not just a metalinguistic correction. This is particularly so for the minimalistic variant. In order to make the dialogues more natural, researchers need to provide more elaborate contexts, but again there is no general recipe to construct such contexts. In the next section, we discuss an alternative test that avoids these limitations.

4 An inference-based test

4.1 Basic structure

In the previous sections, we identified two major issues with existing tests for focus-sensitivity: (a) the need to construct predicate specific contexts, for which no general recipe is available, and (b) the confounding effects of denial. In the truth-based test and the naturalistic variant of the coherence-based test, the researcher needs to construct predicate-specific contexts, even though they might not know a priori which contexts are relevant for a given predicate. This issue makes it impossible to construct a diagnostic that is general enough to be applied to different predicates in different languages. Without knowing what the relevant contexts are, researchers are simply unable to construct the corresponding tests. The minimalistic variant of the coherence-based test is free from this problem, as it can be constructed without reference to a predicate-specific context. However, it faces problems due to the unnaturalness of the dialogues that its target sentences require and the complications introduced by the use of denial.

Here we introduce a third test (which to our knowledge has not been explicitly formulated as a general test) that aims to address both of these issues by means of probing the consultant’s judgments about (*contextual*) *entailment* rather than truth or coherence.¹⁴

We first demonstrate how to use it to test the focus sensitivity of the additive particle *too*. The test involves asking the consultant whether two premises, which differ in the placement of focus, lead to the same conclusion (32).

- (32) **Conclusion:**
Someone other than Alice will visit Sue.

Concretely, the consultant is presented with the first premise (33) and is asked whether the conclusion (32) necessarily follows from it. In this particular case, given the meaning of the additive particle *too*, the answer to the first test question (34) should be YES.

- (33) **Premise 1:**
ALICE will visit Sue on Thursday, too.

- (34) **Test question 1:**
Suppose the premise (33) is true. Does it necessarily follow that someone other than Alice will visit Sue? (If not, please describe a situation where the premise is true but the conclusion is not.)

The consultant is then presented with the second premise (35), which differs from the first one (33) only in terms of the placement of focus, and is asked whether the same conclusion (32) necessarily follows from the second premise (35).

- (35) **Premise 2:**
Alice will visit Sue on THURSDAY, too.

- (36) **Test question 2:**
Suppose the premise (35) is true. Does it necessarily follow that someone other than Alice will visit Sue? (If not, please describe a situation where the premise is true but the conclusion is not.)

possible that the intended interpretation $P\varphi \wedge \neg P\varphi'$ is consistent but the strengthened $P\varphi \wedge P\neg\varphi'$ is not, and the consultant’s judgment is for the strengthened interpretation because it is more salient.

¹⁴Tonhauser and Matthewson (2016) argue that in general, to probe whether a certain inference arises from a certain linguistic construction, one should use a test not formulated in terms of the theoretical notion of *entailment* but rather formulated in terms of a more commonplace notion like *implication*, since consultants may not have sufficient training in linguistics (semantics in particular) to understand what is meant by *entailment*. In light of this, our formulation of the test below avoids the use of the term *entailment* and we call it an *inference-based* test.

In this case, the answer to the second test question (36) should be NO, and the consultant is expected to provide the a concrete scenario along the lines of (37) to justify their answer.

(37) **Sample answer to test question 2:**

No. It can be that Alice will visit Sue on Tuesday and Thursday, and nobody other than Alice will visit Sue. In this case, the premise (35) is true but the conclusion (32) is not.

What can researchers conclude from the answers to the two test questions above? If *too* were not focus sensitive, then the two premises would be expected to have identical truth conditions despite the their difference in focus placement. Consequently either both premises would entail the conclusion or neither of them would, i.e., if *too* were not focus sensitive, the answers to the two test questions would be the same. Given that the answers are different, researchers can conclude that *too* is in fact focus sensitive.

Crucially, this test not only allows researchers to learn whether the operator of interest (which in this case is the additive particle *too*) is focus sensitive, but also allows them to elicit relevant scenarios such as (37), which can subsequently be used for other tests that require concrete scenarios such as the truth value judgment test discussed earlier.

Now, in order to test whether a clause-embedding predicate is focus sensitive, we similarly need two premises that only differ in terms of the placement of focus in the embedded clause, and test whether their entailment patterns are the same as one another, or different. We are inspired, here, by an argument from Dretske (1972: 415–416) for the focus sensitivity of the predicate *advise*.¹⁵ Consider the triplet in (38), modified from Dretske’s original to fit the format of our entailment-based test below.

- (38) a. Alex advised Clyde to sell his car to Schulz for THIRTY THOUSAND DOLLARS. (Premise 1)
b. Alex advised Clyde to sell his car to SCHULZ for thirty thousand dollars. (Premise 2)
c. Alex advised Clyde to sell his car to Schulz. (Conclusion)

To sketch out a situation in which (38a) is true while (38c) is not, Dretske provides the dialogue in (39).

- (39) Clyde: Alex, I need your advice. I have a 1927 Lincoln in my garage that is in mint condition. I haven’t driven it for 35 years and it runs perfectly. Schultz, down the street, has expressed an interest in buying it and has offered me \$30,000 for it.

Alex: So what is your problem?

Clyde: Well, I thought maybe if I held on to it longer it would become even more valuable.

Alex: That isn’t very likely. Your car isn’t going to appreciate in value much more no matter how long you keep it, and you may never again receive such a fine offer. I advise you to sell it to him.

According to his judgment, this situation makes (38a) true since Alex’s advice resolves Clyde’s question about whether to sell the car for \$30,000 now or to wait. However, in the same situation, (38c) is not true since Alex’s advice was not about *who* to sell the car to. In effect, even though it is not worded this way, Dretske points out that (38a) doesn’t entail (38c).

Now, Dretske’s purpose was to argue for the focus sensitivity of specific clause-embedding predicates such as *advise*, and therefore he did not propose a general formulation of a test to diagnose focus sensitivity of clause-embedding predicates. Below we formulate such a general test and discuss its advantages and limitations.

First, consultants are presented with the following broad context (40).

(40) **Broad context:**

Prof. Smith is a professor in a linguistics department. The department will offer several courses in the next semester. Each course will be taught by exactly one professor on exactly one day of the week.

Lisa is a student in the department. She knows all the background information above, but she may or may not know the exact course schedule. That is, it is possible that she has full, no, or partial

¹⁵Villalta (2008) already credits Dretske for the formulation of her truth-value judgment task. We extend our gratitude to Felix Frühauf for drawing our attention to the fact Dretske’s discussion of focus sensitivity contains the seed for our entailment-based test as well.

information about the course schedule.

Then, consultants answer the two test questions in (41) and (42). Here, “*Ps*” stands in for a particular clause-embedding predicate in the simple present, like *wants* or *believes*.

- (41) a. **Premise 1:**
Lisa *Ps* that Peter will teach syntax on THURSDAY.
b. **Conclusion:**
Lisa *Ps* that Peter will teach syntax.

Test question 1:

Suppose Premise 1 is true. Does it necessarily follow that Lisa *Ps* that Peter will teach syntax? (If not, please describe a situation where the premise is true but the conclusion is not.)

- (42) a. **Premise 2:**
Lisa *Ps* that PETER will teach syntax on Thursday.
b. **Conclusion:**
Lisa *Ps* that Peter will teach syntax.

Test question 2:

Suppose Premise 2 is true. Does it necessarily follow that Lisa *Ps* that Peter will teach syntax? (If not, please describe a situation where the premise is true but the conclusion is not.)

If the answers to the two test questions are different, researchers conclude that *P* is focus sensitive. If the answers are the same, then researchers conclude that the predicate is not focus sensitive.¹⁶

This inference-based test avoids the two major limitations of the tests discussed in the previous sections. First, unlike the truth-based test and the naturalistic variant of the coherence-based test, the inference-based test is fully general in that researchers do not need to construct predicate-specific contexts. Second, unlike the minimalistic variant of the coherence-based test, the inference-based test does not involve denial and therefore can avoid any complications it may bring.

These features of the inference-based test make it an attractive and promising alternative to the previous tests. However, despite this initial appeal, the inference-based test also has its limitations, which we will discuss in the next section.

4.2 Limitations of the inference-based test

4.2.1 Difficulty in detecting (certain) non-entailments and articulating invalidating contexts

The first limitation of the inference-based test concerns its practical feasibility. The test crucially relies on the consultants’ ability to determine whether a conclusion always or necessarily follows from a premise (under certain contextual assumptions), i.e., whether the premise (contextually) entails the conclusion. Furthermore, when an inference is not an entailment, consultants need to be able to construct and articulate relevant invalidating contexts.

These can make the inference-based test more demanding for the consultants than the truth- and coherence-based tests. This is one of the reasons for why tests based on judgments about entailment are quite unusual in fieldwork, and why Tonhauser and Matthewson (2016) rate such tests towards the lower end of a scale based on linguistic tests’ stability, replicability and transparency.

Case Study In light of this general concern about the practical feasibility of the inference-based test, We conducted an initial investigation with 3 non-linguist native speakers of English. Each consultant, in the beginning of their respective elicitation session, was told that they would see pairs of sentences and that their task was to determine whether the second sentence necessarily follows from (the truth of) the first.

To familiarize them with the basic structure of the task, the consultants were first presented with Chierchia and McConnell-Ginet’s (1990) textbook examples of entailment (45) and (46), as well as highly plausible inferences that are not entailment (43) and (44). For each pair of sentences, the consultants were asked the

¹⁶This second *if*-clause corresponds to a strong formulation of the inference-based test, which allows for the conclusion that a given predicate is not focus-sensitive, as opposed to the conclusion that we fail to find evidence that it might be. It is not clear whether Dretske would endorse this version. We will further discuss this issue later in section 4.2.3.

following question: “Suppose the first sentence is true. Does it necessarily follow that the second sentence is also true?” All 3 consultants correctly answered NO for (43) and (44) and YES for (45) and (46). They were also asked to provide explanations for their judgments of non-entailment, i.e., for NO judgments, along the lines of “It could be sunny and cold.”

- (43) a. Today is sunny.
b. Today is warm.
- (44) a. Lee kissed Kim passionately.
b. Lee kissed Kim several times.
- (45) a. Lee kissed Kim passionately.
b. Kim was kissed.
- (46) a. Bo will teach phonology on Thursday.
b. Bo will teach phonology.

Next, the consultants were presented with the inference-based test for the additive particle *too*. They were asked to read out the first sentence in the pair with stress or emphasis on the word in capital letters and then were asked whether the truth of the second sentence necessarily follows from the first one.

- (47) a. ALICE will visit Sue on Thursday, too.
b. Someone other than Alice will visit Sue.
- (48) a. Alice will visit Sue on THURSDAY, too.
b. Someone other than Alice will visit Sue.

While Consultants 1 and 3 correctly answered NO for (48), Consultant 2 initially judged (48) to be an entailment. However, when we followed up by asking whether (48a) is compatible with the sentence *nobody else will visit Sue*, which is equivalent to the negation of (48b), Consultant 2 correctly answered YES and agreed that (48b) in fact does not necessarily follow from (48a).

Having finished the two practice questions with *too* (and the follow-up discussion, if there was any), the consultants were presented with the broad context (40), repeated below in (49), and were told that this would be the general background throughout the remainder of the consultation session.

- (49) Prof. Smith is a professor in a linguistics department. The department will offer several courses in the next semester. Each course will be taught by exactly one professor on exactly one day of the week.
Lisa is a student in the department. She knows all the background information above, but she may or may not know the exact course schedule. That is, it is possible that she has full, no, or partial information about the course schedule.

Finally, the consultants were presented with pairs of sentences used in the inference-based test for clause-embedding predicates: (50) and (51) for *believe* and (52) and (53) for *want*.

- (50) a. Lisa believes that Prof. Smith will teach syntax on THURSDAY.
b. Lisa believes that Prof. Smith will teach syntax.
- (51) a. Lisa believes that Prof. SMITH will teach syntax on Thursday.
b. Lisa believes that Prof. Smith will teach syntax.
- (52) a. Lisa wants Prof. Smith to teach syntax on THURSDAY.
b. Lisa wants Prof. Smith to teach syntax.
- (53) a. Lisa wants Prof. SMITH to teach syntax on Thursday.
b. Lisa wants Prof. Smith to teach syntax.

In the case of *believe*, all 3 consultants judged both inferences to be entailments. In the case of *want*, only Consultant 2 judged (52) to be a non-entailment, commenting that it could be that Lisa gets Prof. Smith whether she likes it or not. All 3 consultants judged (53) to be an entailment.

Therefore, the inference-based test classifies *believe* as not focus sensitive for all 3 consultants, but it classifies *want* as focus sensitive for only 1 out of our 3 consultants.

The results of our initial investigation suggest that the inference-based test is challenging, at least for non-linguist consultants. In particular, consultants may fail to consider the relevant invalidating contexts and consequently judge a non-entailment to be an entailment. When this happens, the inference-based test will wrongly classify a focus-sensitive predicate as not focus sensitive. That is, the result of the inference-based test can be a false negative.

4.2.2 Focus sensitive predicates validating both inferences?

Above we identified one way in which the test may yield false negatives (i.e., classifying a focus sensitive predicate as not focus sensitive): the consultant may fail to consider the relevant invalidating context and wrongly judge a non-entailment to be an entailment. This can be seen as a type of performance errors. Below, when we discuss another potential way in which the test can yield false negatives, we will set this type of errors aside. That is, we assume that our consultants can reliably determine whether an inference is an entailment or not.

The issue of potential false negatives we demonstrate in the following applies not only to the inference-based test, but also to the truth-based test. To see the generality of the issue, it helps to first briefly return to the truth-based test. Recall the definition of a focus-sensitive clause-embedding predicate (12), repeated in (54).

(54) **Definition of a focus-sensitive clause-embedding predicate**

A clause-embedding predicate P is focus sensitive iff there exist a context C and two clauses S and S' that are only different in terms of the placement of focus such that (i) $\lceil x Ps S \rceil$ and $\lceil x Ps S' \rceil$ have different truth values in C and (ii) the difference in the truth values cannot be attributed to factor(s) independent from the use of P .

The truth-based test compares the truth values of the following sentences in a context C and classifies a predicate as focus sensitive if the two sentences have different truth values in C (and it fails to be conclusive if the truth values are the same).

- (55) a. Lisa Ps that Peter will teach syntax on THURSDAY.
 b. Lisa Ps that PETER will teach syntax on Thursday.

As discussed before, a major limitation of the truth-based test is that there is no general recipe to construct the relevant context. For now we will set this limitation aside, and we will ask whether for any focus-sensitive predicate, the test can in principle correctly identify it as focus sensitive. It turns out that the answer is NO. Consider the following predicate B' , defined to be exactly the same as *believe*, except that it always returns FALSE when the set of focus alternatives for the embedded clause is {Peter will teach syntax on Thursday, Peter will teach semantics on Thursday, Peter will teach phonology on Thursday, ... }.

It is easy to verify that B' is indeed focus sensitive: Suppose that Lisa believes Peter will teach syntax on Thursday. According to the definition of B' , (56a) is TRUE whereas (56b) is FALSE.

- (56) a. Lisa $B's$ that Peter will teach syntax on THURSDAY.
 b. Lisa $B's$ that Peter will teach SYNTAX on Thursday.

However, since *believe* is not focus sensitive and B' is equivalent to *believe* in the case of (55a) and (55b), these two sentences will always have the same truth values no matter which context is used and therefore the truth-based test, which is based on the pair (55a) and (55b), can never correctly classify B' as focus sensitive.

At this point, a natural objection is that the truth-based test should not be tied to the specific sentence pair (55a) and (55b). Rather, the test should be based on any pair of sentences $\lceil x Ps S \rceil$ and $\lceil x Ps S' \rceil$, where S and S' only differ in the placement of focus. Indeed, when the truth-based test is formulated in this way, it can in principle correctly classify any focus-sensitive predicate P as focus sensitive. However, this would lead to another practical problem: not only is there no general recipe to construct the relevant context, there is no general recipe to construct the relevant pair of sentences, either. That is, in order to

correctly handle cases like B' , the truth-based test needs to be formulated in a way that makes it even more difficult to be applied in a principled and uniform way.

But is such a formulation really necessary? That is, do we really need to worry about cases like B' ? The definition of B' is very ad hoc. It only exhibits focus sensitivity for a specific part of a specific embedded clause. This makes the meaning of B' so unnatural that it is hard to imagine that it would actually be attested in a natural language. Therefore, even though cases like B' can technically be a problem for the formulation of the truth-based test based on a specific sentence pair (55a) and (55b), we suspect that they are not attested in natural languages and therefore in practice we can safely ignore this problem.

More generally, we conjecture the following semantic universal for focus-sensitive predicates in natural languages (57). The intuition behind it is that a focus-sensitive predicate attested in natural language should be focus sensitive in a uniform way, in that it should exhibit focus sensitivity regardless of which particular clause it embeds.

(57) **Conjectured semantic universal T :**

For any predicate P attested in natural languages, if P is focus sensitive then $\lceil x P s S \rceil$ and $\lceil x P s S' \rceil$ have different truth conditions, for any S and S' that differ only in terms of the placement of focus. In particular, if P is compatible with finite clauses, (55a) and (55b) have different truth conditions.

This conjectured universal, if correct, will ensure that the truth-based test based on a specific pair of sentences (55a) and (55b) can in principle correctly classify any focus-sensitive predicate attested in natural languages as focus sensitive.

Now we return to the inference-based test. Recall its general structure as follows:

(58) a. **Premise 1:**
Lisa P s that Peter will teach syntax on THURSDAY.

b. **Conclusion:**
Lisa P s that Peter will teach syntax.

(59) a. **Premise 2:**
Lisa P s that PETER will teach syntax on Thursday.

b. **Conclusion:**
Lisa P s that Peter will teach syntax.

The consultants are asked whether Premise 1 entails the Conclusion and whether Premise 2 does. If the answers are different, the predicate is focus sensitive according to the test, and if the answers are the same, the predicate is not focus sensitive.¹⁷

Technically, a predicate P can be focus sensitive and still validate the two inferences (58) and (59). For instance, consider a predicate B'' , which is defined to be almost exactly the same as *believe*, except that it will return FALSE when the set of focus alternatives of the embedded clause is {Peter will teach syntax on Monday, Peter will teach syntax on Tuesday, ... }.

In a situation where Lisa believes that Peter will teach syntax on Thursday, Premise 1 is FALSE (because the alternative set matches the special clause in the definition of B'') and Premise 2 is TRUE. This shows that B'' is focus sensitive.

Given that *believe* is upward-entailing and not focus sensitive, both (60) and (61) are entailments.

(60) a. **Premise 1:**
Lisa believes that Peter will teach syntax on THURSDAY.

b. **Conclusion:**
Lisa believes that Peter will teach syntax.

(61) a. **Premise 2:**
Lisa believes that PETER will teach syntax on Thursday.

b. **Conclusion:**
Lisa believes that Peter will teach syntax.

¹⁷In the next section, we will discuss variants the inference-based test that differ in the conclusion they would draw when the answers are the same.)

In the case of Premise 2 and Conclusion, B'' is equivalent to *believe*, and therefore (62) is an entailment.

- (62) a. **Premise 2:**
 Lisa B'' 's that PETER will teach syntax on Thursday.
 b. **Conclusion:**
 Lisa B'' 's that Peter will teach syntax.

Finally, given the definition of B'' , Premise 1 is always FALSE and therefore (63a) trivially entails (63b).

- (63) a. **Premise 1:**
 Lisa B'' 's that Peter will teach syntax on THURSDAY.
 b. **Conclusion:**
 Lisa B'' 's that Peter will teach syntax.

In sum, even though B'' is focus sensitive, it validates both inferences. Therefore, the inference-based test would wrongly classify it as not focus sensitive.

Again, to what extent should we be worried about this? We observe that in this case focus sensitivity comes from a highly idiosyncratic use of the focus alternatives: B'' is doing something special only to some particular sets of focus alternatives, whereas canonical focus sensitive predicates such as *hope* and *advise* use focus alternatives in uniform ways. This suggests that while technically predicates like B'' can lead to false negatives, their focus sensitive meanings may be too artificial to be actually lexically realized, and therefore in practice they may not present a real challenge to the inference-based test.

More generally, in parallel with (57), we tentatively propose the following semantic universal (64).¹⁸

- (64) **Conjectured semantic universal I:**
 For any predicate P attested in natural languages, if P is focus sensitive then $\lceil x P s A B_F \rceil$ does not entail $\lceil x P s A \rceil$, for any A and B . In particular, if P is compatible with finite clauses, it invalidates (58).

If this conjecture is true, then predicates like B'' do not actually exist in natural languages, and we do not need to worry about this type of false negatives.

Of course, it is ultimately an empirical question whether the conjectured semantic universals (57) and (64) in fact hold. And if they do, it would also be interesting to see whether these universals can be theoretically derived based on some more fundamental assumptions about focus sensitivity and lexicalization. Even though addressing these questions is beyond the scope of the current paper, we note that all the predicates that have been suggested to be focus sensitive that we know of are consistent with (57) and (64), which lends some initial support to the conjectures.

Finally, we acknowledge that even though there is some parallel between (57) and (64) and to the best of our knowledge that they are both correct, they may not be perfectly comparable. As discussed above, the intuition behind (57) is rather straightforward: if a predicate is focus sensitive, it should exhibit focus sensitivity regardless of which clause it embeds. It is unclear how to similarly motivate (64). Also, as we will see in the next section, even assuming (64) is correct, the inference-based test is not fully immune to false negatives.

4.2.3 Non-upward-entailing predicates

If the conjectured semantic universal (64) holds, assuming that consultants can correctly recognize entailments and non-entailments, the inference-based test can always correctly identify a focus-sensitive predicate if it is Upward Entailing (UE). To see why, we first note that if a predicate P is UE, by definition the following inference (65) is an entailment.

- (65) a. **Premise 0:**
 Lisa P s that Peter will teach syntax on Thursday.
 b. **Conclusion:**
 Lisa P s that Peter will teach syntax.

¹⁸Note that this conjecture is uni-directional. It does not say that all predicates invalidating (58) are focus sensitive.

Given the conjectured semantic universal (64), if this UE predicate is focus sensitive, it will invalidate (58), repeated below in (66).

- (66) a. **Premise 1:**
Lisa *P*s that Peter will teach syntax on THURSDAY.
b. **Conclusion:**
Lisa *P*s that Peter will teach syntax.

Assuming that consultants can correctly judge (66) to be a non-entailment, their judgments for (65) and (66) will be different. Therefore, the inference-based test will correctly classify *P* as focus sensitive.

However, when the predicate is non-UE, the inference-based test may yield yet another type of false negatives. Consider the following predicate *D*. (Its meaning loosely resembles the English predicate *deny*, but we are not committed to this being a fully adequate analysis of *deny*. For our current purposes, what matters is that the focus-sensitive meaning of *D* seems reasonable and therefore it is conceivable that *D* may be lexicalized in some natural language.)

- (67) A *D*s φ iff A says something that commits them to $\neg\varphi$ in response to a QUD that corresponds to the focus alternatives of φ

First, in the following scenario (68), Premise 1 is TRUE but the Conclusion is not. Therefore, (69) is not an entailment.

- (68) Mary asks Lisa: “When will Peter teach syntax? I heard that he will do it on Thursday?”
Lisa: “No. It will be on Tuesday.”

- (69) a. **Premise 1:**
Lisa *D*s that Peter will teach syntax on THURSDAY.
b. **Conclusion:**
Lisa *D*s that Peter will teach syntax.

Now consider a different scenario (70).

- (70) Mary asks Lisa: “Is it going to be Peter who will teach syntax on Thursday?”
Lisa: “All I know is that Peter can’t possibly teach syntax on Thursday because he has a conflict. I am actually not sure whether syntax will be taught on Thursday. And I think it is possible that Peter will teach syntax.”

In this scenario, Premise 2 is TRUE and Conclusion is not.¹⁹ Therefore, (71) is not an entailment, either.

- (71) a. **Premise 2:**
Lisa *D*s that PETER will teach syntax on Thursday.
b. **Conclusion:**
Lisa *D*s that Peter will teach syntax.

Given that the answers to the 2 test questions are the same, the inference-based test will classify *D* as not focus sensitive. But in reality, *D* is focus sensitive, because in scenario (68), Premise 1 (69a) is TRUE but Premise 2 (71a) is not. All in all, this means that the inference-based test yields a false negative for *D*. More generally, for non-UE focus-sensitive predicates, the answers to the 2 test questions may be both NO. In such cases, the inference-based test will yield false negatives.

In light of this problem, one may question whether the inference-based test has been formulated appropriately, especially regarding what to conclude when the answers to the 2 test questions are both NO. Below we discuss two ways to reformulate the test and suggest that neither is fully satisfactory.

First, according to a radical reformulation of the test, the result only depends on whether Premise 1 entails the Conclusion. Specifically, the test will classify the predicate as focus sensitive iff Premise 1 does

¹⁹Note that this is a place where *D* may be different from English *deny*. In the case of *deny* (at least for some speakers) Premise 2 (i.e., *Lisa denies that PETER will teach syntax on Thursday*) strongly implies that Lisa agrees/believes that syntax will be taught on Thursday. If this implication is an entailment, then Premise 2 is not TRUE in scenario (70) and hence (70) will not be an invalidating context.

not entail the Conclusion. This version of the test correctly classifies D as focus sensitive, but the crucial question is whether it will generally yield correct results.

According to our previously conjectured semantic universal (64), repeated below in (72), all focus sensitive predicates attested in natural languages invalidate (58), i.e., Premise 1 does not entail the Conclusion. This means that all focus sensitive predicates will be correctly classified as such according to the radical formulation of the test, i.e., the test will not lead to false negatives.²⁰

(72) **Conjectured semantic universal I:**

For any predicate P attested in natural languages, if P is focus sensitive then $\lceil x P s A B_F \rceil$ does not entail $\lceil x P s A \rceil$, for any A and B . In particular, if P is compatible with finite clauses, it invalidates (58).

The remaining question is whether the test will yield false *positives*, i.e., it will incorrectly classify a non-focus sensitive predicate as focus-sensitive. This hinges on whether the following assumption (73), which is the other direction of our previously conjectured one (64), also holds.

(73) **The other direction of Conjecture I:**

For all predicates attested in natural languages, if $\lceil x P s A B_F \rceil$ does not entail $\lceil x P s A \rceil$, for any A and B , then P is focus sensitive. In particular, if P is compatible with finite clauses and it invalidates (58) then it is focus sensitive. [not endorsed]

If (73) also holds, then any predicate classified to be focus sensitive by the test is in fact focus sensitive. That is, the test will not yield false positives.

But how plausible is (73)? We first note that it is technically possible to define a non-focus-sensitive predicate that invalidates (58). Consider D' , defined as follows (74), which is not focus sensitive (since its meaning makes no use of focus alternatives).

(74) $A D'$'s φ iff A says something that commits them to $\neg\varphi$.

Now consider scenario (68) again, repeated below (75).

(75) Mary asks Lisa: “When will Peter teach syntax? I heard that he will do it on Thursday?”
Lisa: “No. It will be on Tuesday.”

In this scenario, Premise 1 is TRUE and the Conclusion is not.

(76) a. **Premise 1:**

Lisa D' 's that Peter will teach syntax on THURSDAY.

b. **Conclusion:**

Lisa D' 's that Peter will teach syntax.

Hence, D' invalidates the inference despite not being focus sensitive, and the radical reformulation of the test would wrongly classify it as focus sensitive.

Therefore, the radical reformulation of the test relies on the correctness of (73). If (73) is correct, then D' is not attested in any natural language. But, this seems too strong a prediction. While D' might not be the correct analysis of *deny* and hence might be unattested in English, it is unclear that it will be unattested in any language. At the very least, this suggests that the correctness of (73), which the radical reformulation of the inference-based test is based on, is far from obvious. Consequently, it is possible that this version of the test can yield false positives.

The second way to reformulate the inference-based test is a conservative one. Concretely, similar to the original test, the conservative formulation will classify a predicate as not focus sensitive if both answers to the 2 test questions are YES, classify it as focus sensitive if the two answers differ. However, when both answers are NO, the conservative version will remain inconclusive (whereas the original test will classify the predicate as not focus sensitive).

²⁰Of course, this assumes that our conjectured semantic universal (64) is correct, and that consultants can correctly recognize non-entailments. The crucial point is that the radical formulation is not worse than the previous version in these respects.

Technically, in light of examples like D and D' above, when both inferences are non-entailments, staying agnostic about whether the predicate is focus sensitive seems to be all one can do. In principle, if the predicate is focus sensitive, i.e., the two premises differ in truth conditions, then there must exist a conclusion sentence C' such that one premise entails it and the other does not. Therefore, when the answers to the 2 test questions are both NO, in principle the researchers cannot and should not conclude anything. Instead, they should try a different conclusion sentence and repeat the process until the answers to the 2 test questions are different. Then, they can safely conclude that the predicate is focus sensitive. Crucially, however, this method faces the same problem as the truth- or coherence-based test do. This is so because there is no general recipe for determining which other conclusion sentences to test. As a result, when the answers remain the same when researchers try different conclusion sentences, they will not know whether it is because the predicate is not focus sensitive, or because they have not used the relevant conclusion sentence. Thus, we conclude that the conservative formulation of inference-based test faces the same problem as the truth- or coherence-based tests do.

In sum, we acknowledge that the non-upward entailing predicates pose a genuine problem for the inference-based test, and that two possible attempts to remedy the problem are ultimately unsatisfactory.

4.3 Overall assessment and recommendation

In sum, in light of the various limitations discussed above, we conclude that the inference-based test is not a test researchers can solely rely on. Meanwhile, despite its various limitations, we still think it is worth adding the inference-based test to the researchers' toolbox.

Our recommendation is to use the inference-based test as a first step in the investigation. Instead of using the inference-based test to draw a definitive conclusion about whether a predicate is focus sensitive, researchers can use it mainly to elicit relevant invalidating contexts. Such contexts can then be used to construct truth-based tests, which are less demanding for consultants, more replicable, and can help determine the focus sensitivity status of a non-UE predicate.

For instance, consider the hypothetical predicate D in the previous section, whose definition is repeated below (77):

- (77) A D s φ iff A says something that commits them to $\neg\varphi$ in response to a QUD that corresponds to the focus alternatives of φ

Using the inference-based test as a starting point, under the assumption that the consultant can correctly identify (78) as a non-entailment, the researcher can elicit an invalidating context such as (68), repeated below in (79).

- (78) a. **Premise 1:**
 Lisa D s that Peter will teach syntax on THURSDAY.
 b. **Conclusion:**
 Lisa D s that Peter will teach syntax.
- (79) Mary asks Lisa: "When will Peter teach syntax? I heard that he will do it on Thursday?"
 Lisa: "No. It will be on Tuesday."

The researcher can then use this context (79) to construct a truth-based test. That is, they can ask the consultant whether (80) is true in (79) and compare it with (78a) (which has already been judged true because (79) is an invalidating context for (78)).

- (80) Lisa D s that PETER will teach syntax on Thursday.

Given the semantics of D (77), the consultant would judge (80) to be not true in (79). Therefore, the researcher will correctly classify D as focus sensitive.

In addition, note that once the researcher has elicited the invalidating context (68) from a consultant, they can apply the corresponding truth-based test to other consultants, even including those who initially judge the inference (78) to be an entailment. The results would provide further information on whether such consultants have a semantics for D that is different from (77), or the inference (78) is just so highly plausible that it is hard to come up with an invalidating context.

Now let us consider the predicate D' , whose definition is repeated below (81):

(81) A D' 's φ iff A says something that commits them to $\neg\varphi$.

Using the inference-based test as a starting point, the researcher can also elicit (79) as an invalidating context for (82), i.e., in context (79), (82a) is true but (82b) is not.

(82) a. Premise 1: Lisa D' 's that Peter will teach syntax on THURSDAY.
b. Conclusion: Lisa D' 's that Peter will teach syntax.

However, this time, when the researcher uses (79) to construct a truth-based test by asking whether (83) is true in (79), the consultant would judge it to be true as well. Since both (82a) and (83) are true in (79), the researcher will not incorrectly classify D' as focus sensitive.

(83) Lisa D' 's that PETER will teach syntax on Thursday.

Combining the inference- and truth-based tests by using the inference-based test as an initial step to elicit relevant contexts for the truth-based test can make the two tests complement each other. On the one hand, the inference-based test can make the combined test generally applicable. On the other, the truth-based test can yield more stable and replicable results, and the correctness of the results does not hinge on whether the predicate is UE.²¹

5 Summary and outlook

A number of analyses of the syntactic and semantic behaviours of clause-embedding predicates make crucial use of the property of focus sensitivity (e.g., Villalta, 2008; Romero, 2015; Uegaki and Sudo, 2019; Wehbe and Flor, 2022). However, although the analytical intuition behind this property—that the predicate is sensitive to the focus structure of the complement—is relatively straightforward, there are challenges associated with the construction of a concrete empirical test for focus sensitivity which can be applied to any arbitrary predicate in a cross-linguistic setting. In this paper, we have critically examined three types of empirical test for focus-sensitivity: the truth-based test, the coherence-based test, and the inference-based test, and discussed their advantages and limitations. The truth-based test presupposes a construction of a suitable context against which the truth value of target sentences can be judged, which is not possible a priori in a situation where the researcher is yet to uncover the precise lexical semantics of the target predicate. The coherence-based test suffers from a confound arising from the ambivalent nature of the denial expression used in the test. The outcome of the test varies depending on whether the denial expression targets falsity or infelicity, which makes it difficult to formulate the test in a cross-linguistically general manner. The inference-based test is free from these challenges, and can constitute a test format that is general with respect to both predicates and languages. However, the inference-based test faces further challenges associated with practical difficulty with detecting inferences/non-inferences as well as with non-upward-entailing predicates. As a way to mitigate these issues with the individual tests, we have given a methodological recommendation of combining the inference-based test and the truth-based test: the inference-based test can be used as the first step, after which the truth-based test can be used with respect to the (potentially) inference-breaking contexts elicited in the first step.

Through examining the methodological challenges associated with the testing of focus sensitivity, we hope to have contributed to the ongoing discussion on the methodology of controlled semantic data collection in the cross-linguistic context (Matthewson, 2004; Tonhauser and Matthewson, 2016). In particular, our contribution lies in setting one of the precedents for an in-depth *phenomenon-specific* investigation of the semantic data collection methodology. We believe such phenomenon-specific investigations are particularly

²¹To be clear, we are not claiming that the inference-based test should be the only way to obtain relevant contexts for the truth-based test. In many cases, a researcher may well have good intuition about the meaning of a predicate and what kind of contexts may show its focus sensitivity, and they can certainly use such contexts for the truth-based test. Our point is merely that the inference-based test provides a principled way to elicit such contexts from consultants, and therefore is a useful first step, especially when the researcher does not have good intuition about the focus sensitivity status of a predicate in the target language. For instance, a researcher who has D' , which is not focus sensitive, in their native language may fail to entertain the semantics of D a priori.

relevant as the field increasingly relies on cross-linguistic observations in a diverse set of phenomena, where the quality of data assessment methods is crucially important.²²

We would like to conclude this paper by comparing the general methodological position that underlies our proposal with the methodological positions that are most commonly taken in semantic fieldwork and in experimental semantics, respectively. We believe our position is relatively unique in that it combines practices from semantic fieldwork with ones from experimental semantics. Although elicitation of judgements concerning entailment is not commonly used in semantic fieldwork (cf. Matthewson, 2004), it is commonly used in current experimental semantics and more generally for collecting training and test data used in computational linguistics (e.g., Dagan et al., 2006; White and Rawlins, 2018b; Degen and Tonhauser, 2022). In this sense, our position is in line with a large body of work in experimental semantics. At the same time, however, our methodology drastically differs from standard practices in experimental semantics and is closer to fieldwork semantics in emphasizing the importance of *qualitative* judgments. In our methodology, the inference-based test is an initial prompt that facilitates a discussion between the researcher and the consultant about the precise nature of the consultant’s judgment about inference (or lack thereof). If the consultant judges the inference to hold, the researcher may ask the consultant whether they have considered certain specific contexts. If the consultant judges the inference not to hold, the researcher may ask in which concrete contexts the inference breaks. Such qualitative judgments are typically not part of the data gathered in experimental semantics, and in this respect, our methodology aligns more with fieldwork semantics.

Our general methodological choice is rooted in our belief that the traditional semantic notion of inference/entailment is a useful primitive notion that can be insightfully investigated in cross-linguistic data collection (in line with experimental semantics). At the same time, it would be counterproductive to treat native speaker consultants as ‘subjects’ who only provide raw judgments because—we believe—they can provide qualitative data that are theoretically informative about what underlies those judgments (contra standard practice in experimental semantics; and also contra some claims in the fieldwork semantics literature (see Louie, 2015)). We hope to have shown that cross-linguistic investigation of focus-sensitivity is a domain where this hybrid methodological position is particularly suitable.

References

- Beaver, D. I. and B. Z. Clark (2007). *Sense and Sensitivity: How Focus Determines Meaning*. Wiley-Blackwell.
- Chierchia, G. and S. McConnell-Ginet (1990). *Meaning and Grammar: An Introduction to Semantics*. MIT Press.
- Dagan, I., O. Glickman, and B. Magnini (2006). The pascal recognising textual entailment challenge. In J. Quiñero-Candela, I. Dagan, B. Magnini, and F. d’Alché Buc (Eds.), *Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment*, Berlin, Heidelberg, pp. 177–190. Springer Berlin Heidelberg.
- Deal, A. R. (2015). Reasoning about equivalence in semantic fieldwork. In M. R. Bochnak and L. Matthewson (Eds.), *Methodologies in semantic fieldwork*, pp. 157–174. Oxford University Press, USA.
- Degen, J. and J. Tonhauser (2022). Are there factive predicates? an empirical investigation. *Language* 98(3), 552–591.
- Djävrv, K. (2023). Knowing and believing things: what DP-complements can tell us about the argument structure and the composition of (factive) attitudes. *Journal of Semantics* 40(2–3), 179–233.
- Dretske, F. I. (1972). Contrastive statements. *The Philosophical Review* 81(4), 411–437.
- Egré, P. (2008). Question-embedding and factivity. In F. Lihoreau (Ed.), *Grazer Philosophische Studien* 77, pp. 85–125. Leiden, The Netherlands: Brill.
- Grimshaw, J. (1979). Complement selection and the lexicon. *Linguistic Inquiry* 10(2), 279–326.

²²We thank a reviewer for encouraging us to emphasize this point.

- Guerzoni, E. (2007). Weak exhaustivity and whether: A pragmatic approach. In *Proceedings of Semantics and Linguistic Theory (SALT) XVII*, pp. 112–129. Ithaca, NY: CLC Publications.
- Harner, H. J. (2016). *Focus and the Semantics of Desire Predicates and Directive Verbs*. Ph. D. thesis, Georgetown University.
- Heim, I. (1992). Presupposition projection and the semantics of attitude verbs. *Journal of Semantics* 9(3), 183–221.
- Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. MIT Press.
- Klein, E. (1975). Two sorts of factive predicates. *Pragmatics Microfiche, vol. 1, fiche 1*, B5–C14.
- Lahiri, U. (1991). *Embedded Interrogatives and Predicates that Embed Them*. Ph.D. thesis, MIT.
- Louie, M. (2015). The problem with no-nonsense elicitation plans (for semantic fieldwork). In M. R. Bochnak and L. Matthewson (Eds.), *Methodologies in Semantic Fieldwork*. Oxford University Press.
- Major, T. (2021). *On the nature of “say” complementation*. Ph. D. thesis, UCLA.
- Matthewson, L. (2004). On the methodology of semantic fieldwork. *International Journal of American Linguistics* 70(4), 369–415.
- Mayr, C. (2019). Triviality and interrogative embedding: Context sensitivity, factivity and neg-raising. *Natural Language Semantics* 27, 227–278.
- Özyıldız, D. (2021). *The event structure of attitudes*. Ph. D. thesis, UMass Amherst.
- Roberts, C. (2012). Information Structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5, 1–69.
- Roberts, T. (2021). *How to make believe: Inquisitiveness, veridicality, and evidentiality in belief reports*. Ph. D. thesis, UCSC.
- Romero, M. (2015). Surprise-predicates, strong exhaustivity and alternative questions. In *Proceedings of SALT*, Volume 25, pp. 225–245.
- Rooth, M. (1996). Focus. In S. Lappin (Ed.), *The Handbook of Contemporary Semantic Theory*, pp. 271–298. Oxford: Blackwell.
- Rooth, M. E. (1985). *Association with focus*. Ph. D. thesis, UMass Amherst.
- Saebø, K. J. (2007). A whether forecast. In B. ten Cate and H. Zeevat (Eds.), *TbiLLC 2005*, pp. 189–199. Springer-Verlag Berlin Heidelberg.
- Selkirk, E. (1996). Sentence prosody: Intonation, stress, and phrasing. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory*. Blackwell Publishers.
- Theiler, N., F. Roelofsen, and M. Aloni (2019). Picky predicates: Why believe doesn’t like interrogative complements, and other puzzles. *Natural Language Semantics* 27(2), 95–134.
- Tonhauser, J. and L. Matthewson (2016). Empirical evidence in research on meaning. The Ohio State University and University of British Columbia.
- Uegaki, W. and Y. Sudo (2019). The *hope-wh puzzle. *Natural Language Semantics* 27(4), 323–356.
- Villalta, E. (2008). Mood and gradability: An investigation of the subjunctive mood in Spanish. *Linguistics and Philosophy* 31(4), 467–522.
- von Stechow, K. (1999). NPI licensing, Strawson entailment, and context dependency. *Journal of Semantics* 16(2), 97–148.

- Wehbe, J. and E. Flor (2022). Focus-sensitivity and homogeneity in attitude predicates. In *Proceedings of NELS 52*. Amherst: GLSA.
- White, A. S. (2021). On Believing and Hoping Whether. *Semantics and Pragmatics* 14(6), 1–18.
- White, A. S. and K. Rawlins (2018a). Question agnosticism and change of state. In *Proceedings of Sinn und Bedeutung 21*, pp. 1325–1342.
- White, A. S. and K. Rawlins (2018b). The role of veridicality and factivity in clause selection. In S. Hucklebridge and M. Nelson (Eds.), *Proceedings of the 48th Annual Meeting of the North East Linguistic Society*, Amherst, MA, pp. 221–234. GLSA Publications.
- Zuber, R. (1982). Semantic restrictions on certain complementizers. In *Proceedings of the 12th International Congress of Linguists, Tokyo*, pp. 434–436.