On believing and hoping whether*

Aaron Steven White
University of Rochester

Abstract Theories of clause selection that aim to explain the distribution of interrogative and declarative complement clauses often take as a starting point that predicates like *think*, *believe*, *hope*, and *fear* are incompatible with interrogative complements. After discussing experimental evidence against the generalizations on which these theories rest, I give corpus evidence that even the core data are faulty: *think*, *believe*, *hope*, and *fear* are in fact compatible with interrogative complements, suggesting that any theory predicting that they should not be must be jettisoned.

Keywords: clause embedding, selection, interrogative, veridical, neg-raising, preferential

1 Introduction

Some of the most well-developed theories of clause selection aim to explain the distribution of embedded interrogative and declarative clauses—specifically, which lexical properties condition whether a predicate takes interrogative and/or declarative clausal complements (Hintikka 1975, Karttunen 1977a, Zuber 1982, Berman 1991, Ginzburg 1995, Lahiri 2002, Egré 2008, George 2011, Uegaki 2015, Theiler et al. 2017, 2019, Elliott et al. 2017, White & Rawlins 2018a, Roberts 2019; but see Mayr 2018). Following Grimshaw 1979, a canonical contrast in this literature is that between predicates like *think*, *believe*, *hope*, and *fear*—which are often judged worse with interrogative complements than with declaratives (1a) out of context—and *know*—which is often judged fine with both (1b).

(1) a. Jo {thinks, believes, hopes, fears} (*whether) Bo left.
    b. Jo knows (whether) Bo left.

This contrast is often taken to imply that *know* is compatible with both declarative and interrogative complements—i.e. *know* is a *responsive predicate* (Lahiri

* I am grateful to Judith Tonhauser as well as two anonymous reviewers for their extensive comments on this paper. I am also grateful to Kyle Rawlins for discussion of the work presented in §3.1; to to Hannah An for discussion of the work presented in §3.2; and to Will Gantt and Ben Kane for general discussion. All of the data and code used for this paper can be found at megaattitude.io. This work was supported by National Science Foundation grant BCS-1748969 *The MegaAttitude Project: Investigating selection and polysemy at the scale of the lexicon*. Word count: 7,973

2002)—while *think*, *believe*, *hope*, and *fear* are compatible with only declarative complements—i.e. they are *antirogative predicates*. This inference forms the basis for various generalizations that authors then attempt to derive. I discuss three influential proposals in §2.

The aim of this paper is to show that, not only are these generalizations (and thus their derivations) nonviable when evaluated at the scale of the entire English lexicon; even the inference from the contrast in (1) to the antirogativity of *think*, *believe*, *hope*, and *fear* is unlicensed based on an assessment of the broader empirical landscape. And, I argue, because these predicates' (purported) antirogativity forms the core data for all of these generalizations, the generalizations are left so frail without these data that they should be jettisoned altogether.

I first review prior lexicon-scale experimental evidence against one of these proposals in §3 and then present further evidence against a second based on existing lexicon-scale datasets. This evidence leaves open the possibility that the proposed generalizations might be saved by further constraining them. Using corpus examples drawn from a variety of genres, I suggest in §4 that any such constrained version must jettison the contrast in (1) as supporting evidence. This corpus study leaves open the question of what drives the contrast in (1). In §5, I present an analysis of verb-complementizer cooccurrence frequencies suggesting that one potential factor is the grammatical context of the clause-taking predicate—specifically, tense, aspect, and modality morphology on the verb.

2 Proposed generalizations

Egré (2008) proposes generalization $\mathbb{V}$.

$\mathbb{V}$ A predicate is responsive iff it is veridical.

WHERE a predicate V is *veridical* in a sentence NP V S iff one infers from NP V S that S.[1]

Think and *believe* are not veridical—from (2a), one does not infer that (3)—and so they should not be responsive according to $\mathbb{V}$. In contrast, *know* is veridical—from (2b), one infers that (3)—and so it should be responsive according to $\mathbb{V}$.

(2) a. Jo {thinks, believes} Bo left.
b. Jo knows Bo left.

1 Interpreting *predicate* as *predicate type* and *sentence* as *sentence type*, this definition of veridicality is consistent with a more common definition: that a predicate V is *veridical* iff NP V S entails S. For reasons discussed below, I avoid a definition in terms of entailment, which would imply a property of all possible uses of sentences containing a token of some predicate type.

(3) Bo left.

Thus, insofar as the inference is licensed from the contrast in (1) to the responsivity of *know* and the antirogativity of *think*, *believe*, *hope*, and *fear*, the predictions of $\mathbb{V}$ are correct for those predicates.

Theiler et al. (2017, 2019), following Zuber (1982), propose generalization $\mathbb{NR}$ with similar consequences.

$\mathbb{NR}$ A predicate is antirogative if it is neg-raising.

WHERE a predicate V is *neg(ation)-raising* in sentence NP not V S if one can infer from NP not V S that NP V not S

Think and *believe* are also neg-raising—from (4a), one can infer (4b)—and so they should not take interrogatives according to $\mathbb{NR}$.

(4) a. Jo doesn't {think, believe} Bo left.
b. Jo {thinks, believes} Bo didn't leave.

Importantly, since $\mathbb{NR}$ takes the form of a conditional—rather than a biconditional, like $\mathbb{V}$—it predicts nothing about the distributional properties of non-neg-raising predicates. Thus, insofar as the inference is licensed from the contrast in (1) to the antirogativity of *think* and *believe*, the predictions of $\mathbb{NR}$ are correct for those predicates. This outcome is positive for $\mathbb{NR}$: *hope* and *fear* are standardly assumed to be non-neg-raising (see Uegaki & Sudo 2019), but if $\mathbb{NR}$ took the form of a biconditional, it would predict *hope* and *fear* to be responsive, thereby conflicting with the inference from the contrast in (1) to the antirogativity of *hope* and *fear*.

In an attempt to extend Theiler et al.'s proposal to predicates like *hope* and *fear*, Uegaki & Sudo (2019) propose generalization $\mathbb{V}+\mathbb{P}$.

$\mathbb{V}+\mathbb{P}$ A predicate is antirogative if it is nonveridical and preferential

WHERE a predicate V is *preferential* iff focus in the scope of the predicate has a truth conditional effect.

Hope and *fear* are not veridical—from (5), one does not infer (3)—but they are preferential (see Uegaki & Sudo's Ex. 11), and so they should not take interrogative complements according to $\mathbb{V}+\mathbb{P}$.

(5) Jo {hopes, fears} Bo left.

Like $\mathbb{NR}$, $\mathbb{V}+\mathbb{P}$ takes the form of a conditional: it predicts nothing about the distributional properties of veridical predicates, like *know*, or nonpreferential predicates, like *think* and *believe*.

In what follows, I assume what I take to be the most charitable possible interpretation of each generalization: that each concerns inferences drawn on particular uses of any particular sense of a predicate that falls under the generalization. This interpretation is strictly weaker than the one commonly assigned to generalizations of this form—wherein (i) *predicate* and *sentence* in the definitions of $\mathbb{V}$, $\mathbb{NR}$, and $\mathbb{V}+\mathbb{P}$ are interpreted as *predicate type* and *sentence type*, respectively; and (ii) predicate types are associated with an inferential property, such as veridicality or neg-raising, if all relevant uses of the predicate in sentences of the relevant type have the property. This approach is useful, not only because counterexamples to the weak interpretation are always counterexamples to the more common (strong) interpretation (but not vice versa), but also because it can be hard to pin down whether the strong version of the generalization given in a particular paper is intended in the first place. This shows up in two contexts.

First, it is not always clear how to interpret lexical generalizations when a predicate has potentially multiple senses. Generally, predicate sense is invoked to save a generalization. For instance, at least Egré explicitly argues that *tell* has two distinct senses—one veridical and one nonveridical—and that only the veridical sense is responsive (see also Spector & Egré 2015). In contrast, other arguments rely on senses remaining constant. For instance, *say* with an infinitival complement is nonveridical and preferential under Uegaki & Sudo's criterion yet responsive, counter to the predictions of $\mathbb{V}+\mathbb{P}$. To save $\mathbb{V}+\mathbb{P}$, Uegaki & Sudo are forced to stipulate that *say* only has a single nonpreferential sense—presumably, the same one that shows up with finite complements—and that that preferentiality comes from the infinitival complement.[2] In light of these strategies, it seems reasonable to interpret these generalizations as specific to whatever senses fall under them, rather than assuming that the generalization holds of a predicate if any sense of that predicate falls under it—even one that is plausibly related via a rule of regular polysemy.

Second, because at least neg-raising inferences are defeasible, a predicate sense might fall under a generalization, such as $\mathbb{NR}$, only in contexts where it is triggered. This interpretation is explicitly invoked by Theiler et al. to save their generalization in the face of examples like (6), wherein *believe*—a predicate that can trigger neg-raising inferences—appears to be compatible with an interrogative complement.

(6) You won't believe who called (*in ages)!

2 I call this move a stipulation because it requires unmotivated predicate-specific carve-outs: if the infinitival complement gives rise to preferentiality, it is unclear why *decide* with an infinitival complement should not be similarly preferential under Uegaki & Sudo's definition—especially following current proposals about the classification of infinitives (Wurmbrand 2014). But if the definition of preferentiality were modified to include *decide*, it would create a counterexample to $\mathbb{V}+\mathbb{P}$, since *decide* is nonveridical but responsive.

They argue (their Footnote 11) that this compatibility is only possible when the neg-raising inference is not triggered. Evidence for the lack of a neg-raising inference in (6) comes from the fact that strong negative polarity items (strong NPIs; Zwarts 1998), like *in ages*, cannot occur in the complement (Gajewski 2007). For this move to work, it must be that particular uses of a predicate (sense) in some context are relevant to $\mathbb{NR}$—not the fact that the predicate can trigger the relevant inferences in some contexts. Thus, the weak interpretation of at least $\mathbb{NR}$ seems required. Absent an argument that (6) contains a veridical sense of *believe*, such a move would similarly be required to save $\mathbb{V}$, for which (6) is a counterexample.

3 Experimental evidence against the generalizations

Under the interpretation of $\mathbb{V}$, $\mathbb{NR}$, and $\mathbb{V}+\mathbb{P}$ just discussed, the explanation for the contrast in (1) must be that the default contexts against which the default senses of *think*, *believe*, *hope*, and *fear* are judged imply antirogativity. Assuming that, for a particular sentence, these default senses and contexts remain constant across acceptability and inference judgments, all three generalizations are testable given quantitative measures of veridicality, neg-raising, responsivity, and preferentiality.

Large-scale datasets from which such measures can be derived currently exist for veridicality (the MegaVeridicality dataset; White & Rawlins 2018b), neg-raising (the MegaNegRaising dataset; An & White 2020), and responsivity (the MegaAcceptability dataset; White & Rawlins 2016, 2020)—enabling tests of $\mathbb{V}$ and $\mathbb{NR}$—but to my knowledge, a sufficiently large-scale datasets measuring preferentiality does not yet exist. As such, I focus only on $\mathbb{V}$ and $\mathbb{NR}$ in this section. I first discuss existing quantitative evidence against $\mathbb{V}$ in §3.1. I then turn to a new analysis that provides evidence against $\mathbb{NR}$ in §3.2. Because all three measures are continuous in nature, I begin each subsection by using statistical simulations to cash out the relevant generalization's predictions under different distributional assumptions and then compare the results of these simulations against the observed data.

3.1 Evidence against $\mathbb{V}$

$\mathbb{V}$ predicts that one should find a perfect positive correlation across predicates (modulo measurement error) between measures of veridicality and responsivity. This prediction is visualized in Figure 1 (left), with points and correlations based on a simulation assuming 500 predicates. This simulation assumes (i) that each licit category (*nonveridical antirogative* and *veridical responsive*) is distributed multivariate normal and contains an equal number of predicates; (ii) that the variance for each dimension is equal within and across categories; (iii) that there are no correlations among dimensions within a category; and (iv) that the classes are

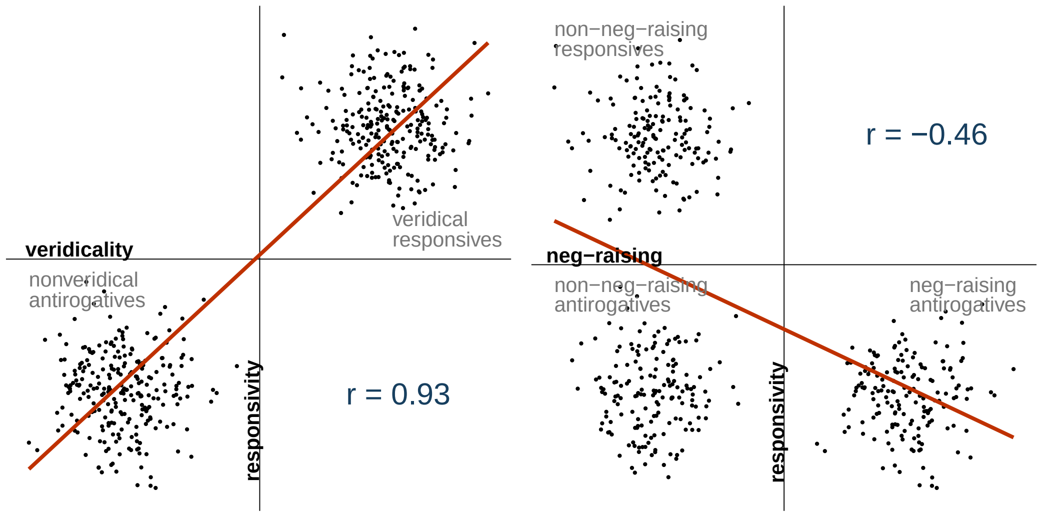

Figure 1 Simulated predictions of $\mathbb{V}$ (left) and $\mathbb{NR}$ (right).

extremely clearly delineated—in Figure 1 (left), the distance between the class centroids is 10 times the intraclass standard deviation along each dimension.

To give a sense for how sensitive the correlation is to these assumptions, Figure 2 (left) plots the simulated correlation when varying the proportion of veridical predicates and how distinguishable the categories are: the larger the ratio of distance between the category centroids to category standard deviation, the more distinguishable the categories. The main take-away from this plot is that, even in the worst case scenario where the categories are very hard to distinguish—i.e. the distance-standard deviation ratio is 0.5—the average correlation is around zero, with the lower bound of the confidence interval never dipping below -0.08.

White & Rawlins (2018b) test the predictions of $\mathbb{V}$, alongside the predictions of a closely related generalization that Egré attributes to Hintikka 1975 (see also Berman 1991, Ginzburg 1995). The fact that they simultaneously test both generalizations requires additional complexity in how they report their analysis that is not relevant here, and so I focus only on a replication of their analysis that assesses the correlation between their measures of veridicality and responsivity.

Using a veridicality judgment task to construct their MegaVeridicality dataset, White & Rawlins obtain a quantitative measure of veridicality for all English verbs that are acceptable with a declarative complement—operationalized as an average acceptability judgment of 4 out of 7 or better for (7a) or (7b) in White & Rawlins's (2016, 2020) MegaAcceptability dataset. A total of 517 verbs in MegaAcceptability fit this criterion.

(7) a. Someone {thought, knew, ...} that something happened.
b. Someone was {told, worried, ...} that something happened.

White & Rawlins's veridicality judgment task prompt is exemplified in (8) for *know*.[3]

(8) Someone knew that a particular thing happened.
Did that thing happen? *yes*, *maybe or maybe not*, *no*

To obtain a measure of each predicate's veridicality from responses to prompts such as (8), White & Rawlins apply an ordinal mixed model-based normalization procedure like that used by White & Rawlins (2020). This procedure, which produces more positive values for predicates that receive more *yes* responses and more negative values for predicates the receive more *no* responses, adjusts for biases in how particular participants use the ordinal scale and shows high correlation with mean by-participant z-scores (see White & Rawlins 2020: Appendix C).

White & Rawlins then derive a measure of responsivity for each verb from MegaAcceptability using the normalized acceptability scores described by White & Rawlins (2016). They first derive a measure of interrogative-taking by taking the maximum average acceptability over both polar (9a) and constituent (9b) interrogatives. The idea behind this measure is to treat a predicate as interrogative-taking insofar as it is good with either polar or constituent interrogatives.[4]

(9) a. Someone (was) {thought, told ...} whether something happened.
b. Someone (was) {thought, told, ...} which thing happened.

They derive an analogous measure of declarative-taking by taking the maximum acceptablility over both the overt (10a) and covert (10b) complementizer variants.

(10) a. Someone (was) {thought, told ...} that something happened.
b. Someone (was) {thought, told ...} something happened.

Finally, to derive the measure of responsivity, they take the minimum of interrogative- and declarative-taking measures for each predicate. The idea behind this responsivity measure is to treat a predicate as responsive only to the extent that it is good with both interrogatives and declaratives: to the extent it is degraded with either, taking the minimum sets the responsivity measure to the less acceptable of the two.

One thing this measure of responsivity does not account for is that it is sensitive to variability in the measure of declarative-taking, since White & Rawlins' method for

3 Their task manipulates the matrix polarity, but only the positive polarity items, such as (8), are relevant here, given the definition of veridicality in $\mathbb{V}$.

4 It is important to include constituent interrogatives in order to capture many emotive predicates, since these predicates are known to be degraded with polar interrogative complements (cf. Sæbø 2007).

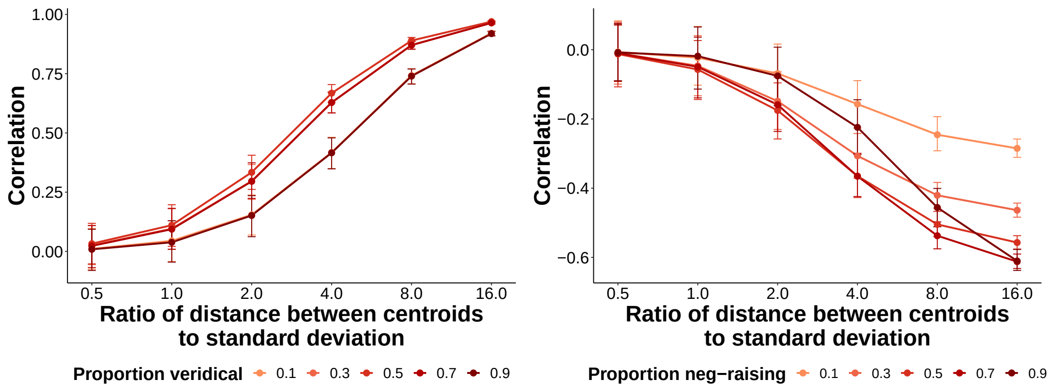

Figure 2 Correlations under different simulation assumptions for $\mathbb{V}$ (left) and $\mathbb{NR}$ (right). Error bars shows 95% confidence intervals.

selecting predicates based on acceptability with declaratives allows predicates that might be quite middling with declaratives. This situation is potentially problematic because the generalizations discussed in §2 presuppose that a predicate is acceptable with declaratives; but it is largely unavoidable, since there is no natural acceptability threshold distinguishing grammatical and ungrammatical items in MegaAcceptability (see the histograms in White & Rawlins 2020: Appendix C). To deal with this, White & Rawlins weight their analyses by acceptability with a declarative—predicates that are less acceptable with declaratives receiving less weight. To derive this weight, they z-score the measure of declarative-taking across only predicates in the MegaVeridicality and then apply the normal cumulative distribution function to those values. This method ensures that predicates that are the "most" declarative-taking have weights near 1 and that predicates that are the "least" declarative-taking have weights near 0.

Figure 3 (left) plots the measure of responsivity against the measure of veridicality, with transparency corresponding to the declarative-taking weight. The line shows a linear regression weighted by declarative-taking weight. The corresponding correlation weighted by declarative-taking is -0.08 (95% CI = $[-0.18, 0.01]$). This correlation violates the predictions of $\mathbb{V}$, suggesting that $\mathbb{V}$ is false as a generalization about the lexicon as a whole.

3.2 Evidence against $\mathbb{NR}$

$\mathbb{NR}$ predicts a negative correlation between measures of neg-raising and responsivity. But in contrast to $\mathbb{V}$, which takes the form of a biconditional, $\mathbb{NR}$ does not predict that

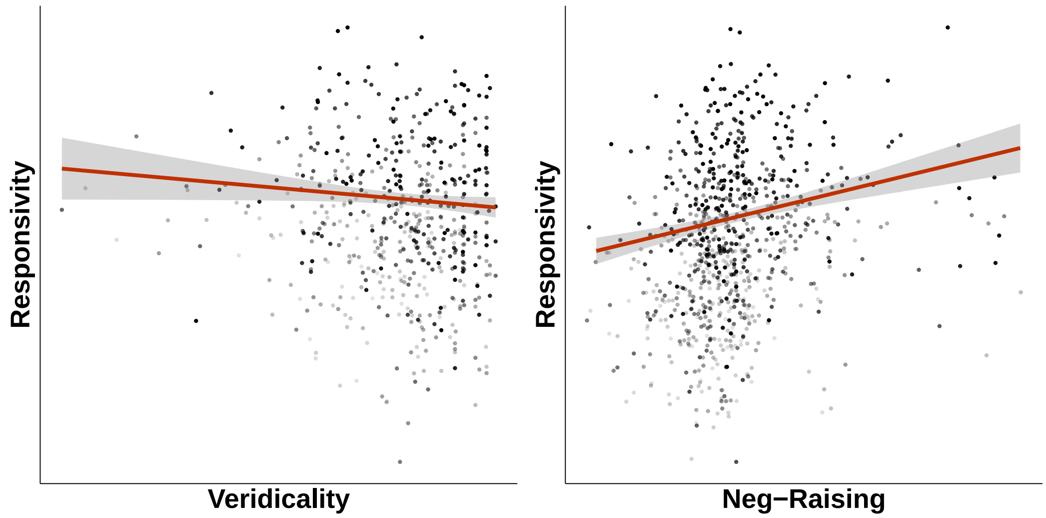

Figure 3 Correlations between veridicality (left), neg-raising (right), and responsivity.

the correlation is perfect, since it takes the form of a conditional. The reason for this is visualized in Figure 1 (right), which employs similar distributional assumptions to Figure 1 (left). The predicted correlation is highly dependent on the ratio of the responsives to antirogatives among the non-neg-raisers. In Figure 1 (right), a ratio of 1 is assumed, and insofar as the ratio is high, the correlation will be higher and vice versa. Indeed, a correlation of 0 is technically compatible with $\mathbb{NR}$—the case where there are either no responsives or no neg-raisers—but it would leave $\mathbb{NR}$ with effectively no predictive power.

To give a sense for how sensitive the correlation is to these assumptions, Figure 2 (right) plots the simulated correlation when varying the proportion of neg-raising predicates and how distinguishable the categories are. For the x-axis, the distance between the centroid of the non-neg-raising responsives and the centroid of the neg-raising antirogatives is used to compute the ratio. These plots do not vary the ratio of responsives to antirogatives among the neg-raisers, leaving that ratio constant at 1. Analogous to the left plot, the main take-away from this plot is that, even in the worst case scenario where the categories are very hard to distinguish—e.g. where the distance-standard deviation ratio is 0.5—the correlation is around zero with the upper bound of the confidence interval never rising above 0.09.

To test $\mathbb{NR}$, I deploy White & Rawlins's methodology in conjunction with An & White's (2020) MegaNegRaising dataset. An & White derive a measure of neg-raising using a likelihood judgment task, wherein participants are asked to judge how likely a speaker is to mean one thing if they say another. For instance, to assess whether *think* is neg-raising in the present tense with a first person subject, participants were asked to respond to (11) with a slider.

(11) If I were to say *I don't think that a particular thing happened*, how likely is it that I actually mean that *I think that that thing didn't happen*?

After showing that this method tracks neg-raising judgments reported in the literature well (their Appendix C), An & White apply their method to all English verbs acceptable with a declarative complement under White & Rawlins's declarative-taking criterion, varying both the subject (*first* v. *third*) and tense (*past* v. *present*). The subject-tense variants for the item in (11) are exemplified in (12).

(12) {I {didn't, don't}, A particular person {didn't, doesn't}} think that a particular thing happened.

To obtain a measure of each predicate's neg-raising with each combination of person and tense from responses to prompts such as (11), White & Rawlins apply a mixed model-based normalization procedure (see An & White 2020: Appendix D). To obtain a single measure of neg-raising for each predicate, I take the maximum of these normalized neg-raising judgments across all four subject and tense variants for a particular predicate, with the assumption that a predicate is neg-raising if it is neg-raising with at least one combination of subject and tense.

Figure 3 (right) plots White & Rawlins' measure of responsivity against this measure of neg-raising, with transparency corresponding to the declarative-taking weight. The line shows a linear regression weighted by declarative-taking weight. The corresponding correlation weighted by declarative-taking is 0.19 (95% CI = [0.12, 0.27]). This correlation violates the predictions of $\mathbb{NR}$, suggesting that $\mathbb{NR}$ is false as a generalization about the lexicon as a whole.

4 Corpus evidence against the generalizations

Having established that at least $\mathbb{V}$ and $\mathbb{NR}$ fail when applied to the lexicon as a whole, I now address the possibility that they might be rescued by constraining their statement further—e.g. in the way $\mathbb{V}+\mathbb{P}$ is constrained to preferential predicates. Such a constraint might serve to exclude peripheral classes of predicates, while retaining the core contrast in (1). On the basis of corpus examples, I argue that even this core contrast cannot be retained and that all three generalizations should be jettisoned.

I select examples from two corpora: the Corpus of Contemporary American English (COCA; Davies 2008) and the iWeb Corpus (Davies 2018). COCA is a corpus that, at the time of my search, contains over 500 million words spread over five genres: spoken, fiction, popular magazines, newspapers, and academic texts; and iWeb is a corpus containing over 14 billion words of web text (blogs, news articles, etc.). To search for examples, I use each corpus' web interface to query for *think*, *believe*, *hope*, and *fear* in any morphological form, followed by *whether*—e.g. for *think*: `[THINK] [WHETHER]`.[5] From the results of this search, I select examples that I (a native speaker of English acquired in Southern California) take to be perfectly acceptable in context.[6]

I break discussion of these examples into two sections based on which generalizations they are relevant to. In §4.1, I discuss *think* and *believe*, showing that neither $\mathbb{V}$ nor $\mathbb{NR}$ can explain these examples. I then turn in §4.2 to *hope* and *fear*, showing that $\mathbb{V}+\mathbb{P}$ cannot explain these examples.

4.1 *Thinking* and *believing whether*

As discussed in §2, both $\mathbb{V}$ and $\mathbb{NR}$ predict that *think* and *believe* should not take interrogative complements. But *think* is attested with interrogative complements in transcribed speech (13), periodical text (14), and internet text (15).

(13) a. The image of having the members of one branch of government standing up[...]cheering and hollering while the court[...]has to sit there, expressionless[...]is very troubling. And it does cause you to **think whether** or not it makes sense for us to be there.

b. [...]the righteousness is unbelievable and people[...]will have to **think whether** they want four more years of that.

(14) a. When Jan Brown completed her safety briefing for the passengers, she tried to **think whether** she had covered everything.

b. I'm trying to **think whether** I'd have been a star today or not.

(15) a. [O]ften, when listening to some other players (especially beginners) I start to **think whether** there's an unwritten law for guitarists to never play an interval bigger than the major third.

b. [...]he wanted a domain that was memorable, brandable, keyword-rich, and relatively short. That's tough and he started to **think whether** it was worthwhile to look into other TLDs.

5 I focus on polar (*whether*) interrogatives because, following Egré, I assume that polar interrogative selection is more constrained than constituent interrogative selection.

6 An anonymous reviewer reports that they find most of the examples cited below acceptable, but that others are not as good for them. An exploration of such idiolectal variation in this area would be interesting, but it is out of scope for the current paper.

In at least a subset of these examples, one might attempt to save $\mathbb{V}$ by arguing that *think* is in fact veridical in these contexts (see Spector & Egré's 2015 approach to predicates like *tell*). For example, *think* in (14a) might be paraphrased using *remember*, and *think* in (14b) might be paraphrased using *figure out*, both of which are veridical and non-neg-raising. But with respect to veridicality, this move is somewhat suspect for (13) and (15), wherein reasonable paraphrases seem to involve the predicate *consider*. Now, *consider that* may well have a veridical sense—roughly, *consider the fact*, in contrast to *consider the possibility*—but that sense does not appear to be the one active in at least (13b), wherein the speaker seems to be encouraging consideration of a possible desire.

These examples may be less problematic for $\mathbb{NR}$, since proponents might argue that *think* is not neg-raising in (13)-(15)—similar to how *believe* is not neg-raising in (6), discussed in §2 and duplicated below. Remember that Theiler et al. argue that the compatibility of *believe* with an interrogative complement is only possible because no neg-raising inference is triggered in (6), evidenced by the fact that strong NPIs are not licensed in the complement.

(6) You won't believe who called (*in ages)!

Before moving forward, I would like to take a brief aside to patch up what I take to be a flaw in Theiler et al.'s argument: neg-raising never occurs with interrogative complements, so it is unclear why one should expect it to occur here. A proponent of $\mathbb{NR}$ might argue that this data point is another piece of evidence in their favor. But if this data point is indeed evidence for the generalization, it is quite weak, because it does not disassociate a predicate's propensity to license neg-raising in a particular context from other syntactic and semantic factors that might block the inference, such as the form of the complement.

With this in mind, I propose that a better test is to minimally modify sentences like (6) to remove any effect of the form of the complement while retaining all other aspects of the context—e.g. by simply converting an interrogative, as in (6), to a declarative, as in (16). When we do this for (6), we get a result consistent with Theiler et al.'s argument: (16) is bad.

(16) *You won't believe that someone called in ages!

I refer to this as the *interrogative-to-declarative test*. Applying this test to corpus examples like (17) yields contexts where both *think whether* and neg-raising *think* are possible: the strong NPI *until...* is fine in (18) with low attachment to *help*.[7]

7 For those having difficulty obtaining the low attachment reading, consider a context wherein the speaker's manager is questioning why they, as team leader at a help desk, were serving only one customer at a time, when the help desk was staffed with three people.

(17) I was **thinking whether** there was a way to[...]help more than one person.

(18) I wasn't thinking there was a way to help more than one person (at a time) until Jo got back from lunch.

This form of argument is weak, since the *think* in (17) still might be a different *think* from the one in (18), though such a polysemy is otherwise unmotivated. To posit a regular polysemy of this form, it would be necessary to see other predicates showing similar behavior. *Believe* is such a predicate, and it is compatible with interrogative complements—seen in (6) and corroborated by (19).

(19) a. [...]I didn't believe the Bible growing up, I wasn't a Christian growing up, I struggled to **believe whether** I could trust the Scriptures[...]
b. We can choose to **believe whether** the word of God is true[...]or not.
c. I am torn between **believing whether** or not Jagex can detect the RSBot client.

But unlike (6), at least some of these examples pass the interrogative-to-declarative test. For instance, in the context of (20a), the strong NPI *either* in (20b)—analogous to (19b)—is clearly good.

(20) a. We can choose not to believe the teachings of the Buddha are true.
b. We can choose not to believe the word of God is true either.

One property that all of the sentences in (19) share is that, in context, they apparently either do not trigger an *opinionatedness inference*—i.e. the inference that a speaker who uses *NP believe S* endorses $[\![\text{NP}]\!]$ BELIEVE $[\![\text{S}]\!] \vee [\![\text{NP}]\!]$ BELIEVE $\neg[\![\text{S}]\!]$ (Bartsch 1973)—or they only do so vacuously: in (19a), an explicit belief is stated about the trustworthiness of the Bible before the use of *believe whether*, assymetrically entailing the inference; in (19b), a choice of belief has apparently not been made and thus the speaker's cohort is presumably not yet opinionated; and in (19c), in being torn, the speaker explicitly states that they do not have a firm belief. The opinionatedness inference is implicated in the derivation of the neg-raising inference in some approaches, so this fact could then be used to save $\mathbb{NR}$ by denying that the conditions for neg-raising are met in context. But beyond substantially weakening the predictive power of that generalization, this move is not possible for all sentences.

Consider (21).

(21) [...]Richard Stavin, a former veteran federal prosecutor[...]declared[...] "[...]I believe it was an organized, orchestrated effort on the part of certain individuals within Washington, D.C. to keep a hands-off policy towards MCA[...]" **Believing whether** certain individuals within Washington D.C. had an MCA policy is not the same as proving there was such a policy.

Here, the author takes as common ground that Stavin has an opinion, then asserts that having said opinion is not the same as proving that opinion to be true. Beyond being problematic for $\mathbb{NR}$, this example is also problematic for $\mathbb{V}$: the content of (21) draws a contrast between the nonveridicality of belief and the veridicality of proof, and so it would be contradictory to interpret *believe* as veridical in this context.

On the basis of these examples, attesting that *think* and *believe* do indeed take interrogative complements, I argue that both $\mathbb{V}$ and $\mathbb{NR}$ should be jettisoned.

4.2 *Hoping* and *fearing whether*

$\mathbb{V} + \mathbb{P}$ predicts that *hope* and *fear* should not take interrogative complements. Indeed, *hope* and *fear* are two of the central cases discussed as evidence for $\mathbb{V} + \mathbb{P}$ by Uegaki & Sudo. But *hope* is attested with interrogative complements in both speech transcripts (22) and internet text (23).

(22) This Trump/Carson boom really has people like Bush, Walker, Rubio, and others wondering and **hoping whether** history will repeat itself and whether Republicans will return back to focusing on the establishment choices but it's all about outsider candidates right now.

(23) a. I was **hoping whether** you are able to guide me[...]
b. I have done a quite a bit of research on using a Limited Co but was **hoping whether** someone with more experience could confirm my understanding of a few points[...]

A potential worry with at least the sentences in (23) is that they seem paraphrasable using *hope that*—evoking Karttunen's (1977b) observation that declarative and polar interrogative complements appear interchangeable for responsives like *doubt*.

(24) Jo doubts that Bo left. $\approx$ Jo doubts whether Bo left.

While such a paraphrase may be available, this availability cannot imply that the interrogative is somehow different from the one found under other predicates: note that, in (22), *hope* is coordinated with *wonder*, which is unambiguously interrogative-taking. Thus, even if one argues that *hope that* and *hope whether* are in some sense interchangeable, this explanation cannot presume that the interchangeability is due to the interrogative embedded under *hope* having a distinct semantics.

Similar examples can be found for *fear*.

(25) a. Interstellar space is so vast that there is no need to **fear whether** stars in the Andromeda galaxy will accidentally slam into the Sun.

b. I **fear whether** this test would run safely on the oxygen sensor as it has a lot of drawback when compared with the others.
c. [...]I **fear whether** I'll have use of my arms/hands by age 55 or 60.
d. I know parents who seriously **fear whether** their children will ever hold a meaningful job.

Example (25a) might be explained under Mayr's (2018) proposal that interrogative embedding for some predicates, like *be certain*, is licensed by downward-entailing contexts, since the *no* scoping over *fear* in (25a) creates such a context. But this move is not available for at least (25b)-(25d), where no downward-entailing operators outscope *fear*. Further, running the interrogative-to-declarative test on these cases does not yield veridicality inferences about the content of *fear*'s complement, suggesting that veridicality is not licensing these cases: from (26a), one does not infer that the test would (not) run safely; from (26b), one does not infer that the speaker will not (or will) have use of their arms/hands; and from (26c), one does not infer that the parents' children will (not) ever hold a meaningful job.

(26) a. I fear that this test would run safely on the oxygen sensor as it has a lot of drawback when compared with the others.
b. [...]I fear that I won't have use of my arms/hands by age 55 or 60.
c. I know parents who seriously fear that their children will ever hold a meaningful job.

On the basis of these examples, attesting that *hope* and *fear* do indeed take interrogative complements, I argue that $\mathbb{V} + \mathbb{P}$ should be jettisoned.

5 Grammatical correlates of polar interrogative-taking

In light of the corpus evidence discussed in §4, an important question still remains: if *believe*, *think*, *hope*, and *fear* are in fact compatible with interrogative complements, what drives the contrast in (1)? I cannot hope to answer this question here—it seems likely that a vast confluence of syntactic, semantic, and pragmatic factors as well as task effects are at play—but I can attempt to speak to the relevance of one collection of possible factors: tense, aspect, and modality (TAM) morphology on the verb.

The astute reader may notice that, in contrast to the canonical (1), many of the predicates in the corpus examples in §4 occur in something other than simple past or simple present—some occurring in the progressive and others occurring within an infinitival complement. This pattern, if it holds, raises the possibility that TAM morphology interacts with the clause-embedding predicate to license interrogatives. Indeed, if it furthermore *only* holds for *believe*, *think*, *hope*, and *fear*, proponents of $\mathbb{V}$, $\mathbb{NR}$, and $\mathbb{V}+\mathbb{P}$ might argue that the corpus examples in §4 plausibly involve some distinct use of each predicate and thus are not truly counterexamples.

I show here that such an argument is unlikely to be successful: polar interrogative-taking is indeed more frequent when the a predicate has certain kinds of TAM morphology, but this correlation is not confined to *believe*, *think*, *hope*, and *fear*; it is a trend observed across the lexicon. Beyond casting doubt on an approach that attempts to explain away the examples in §4 by positing ambiguity—or at least by positing ambiguity that is confounded with TAM morphology—this finding may suggest a reason why the compatibility of *believe*, *think*, *hope*, and *fear* with interrogatives has gone unnoticed: TAM morphology is not systematically controlled for in any prior work (that I am aware of).

5.1 Corpus search

To investigate whether interrogative selection correlates with particular kinds of TAM morphology, I obtain the frequency with which verbs occur in the patterns in (27) using the COCA search interface.

(27)
a. *present:* {V_{PRES}, do(es) NOT V_{BARE}} (PRONOUN) {that, whether}
b. *past:* {V_{PAST}, did NOT V_{BARE}} (PRONOUN) {that, whether}
c. *perfect:* HAVE (NOT) $\text{V}_{\text{PAST-PART}}$ (PRONOUN) {that, whether}
d. *progressive:* BE (NOT) $\text{V}_{\text{PRES-PART}}$ (PRONOUN) {that, whether}
e. *modal:* MODAL (NOT) V_{BARE} (PRONOUN) {that, whether}
f. *infinitival:* (NOT) to (NOT) V_{BARE} (PRONOUN) {that, whether}

In these patterns, *that* and *whether* must be part-of-speech-tagged as complementizers and small caps elements have the interpretations in (28).

(28)
a. NOT: *not*, *n’t*
b. PRONOUN: *them*, *themself*, *themselves*, *her*, *herself*, *him*, *himself*, *us*, *ourself*, *ourselves*, *it*, *itself*
c. $\text{V}_{\{\text{PRES, PAST, PAST-PART, PRES-PART, BARE}\}}$: any {present, past, past participle, present participle, bare} form of a verb
d. HAVE: any form of the auxiliary *have*
e. BE: any form of the auxiliary *be*
f. MODAL: any modal auxiliary (e.g. *can*, *could*, *might*, etc.)

I use transitive (containing PRONOUN) and intransitive variants of each pattern so that both transitive and intransitive verbs are captured, and I use negative (containing NOT) and positive variants of each pattern so that negation can be explicitly controlled for in the analysis. Controlling for the presence of negation is potentially important in light of preliminary evidence that downward-entailing contexts may play a role in licensing polar interrogative-taking (Mayr 2018). Of course, this string pattern-based approach has limitations in terms of capturing such environments generally, in that it fails to capture cases where a downward-entailing operator does not occur string-near the verb.[8] But due to the fact that no sufficiently large dataset labeled for downward-entailing contexts exists and that even state-of-the-art automated systems for detecting such environments are not yet sufficiently accurate for current purposes, I leave a more general corpus study for future work.

I take up to the top 1,000 most frequent verb lemma types in each pattern, treating the transitive and intransitive variants, the negative and positive polarity variants, and the *that* and *whether* variants of each pattern as distinct, obtaining the 1,000 most frequent matches for each variant. Not all patterns yield 1,000 attested verb lemma types, and for every pattern, the least frequent matches occur only once. I then subset the results of this search to only verbs that occur in the MegaVeridicality dataset to make the corpus frequency results as comparable as possible to the experimental results described in §3. This subsetting leaves a total of 440 verb types in the dataset.

5.2 Statistical analysis

To assess the extent to which each context listed in (27) is correlated with polar interrogative-taking v. declarative-taking, I calculate the number of times each verb occurs in each context with *whether* (1) and with *that* (0). For instance, in the intransitive positive variant of (27b), *ask* occurs 1,695 times with *that* and 1,325 times with *whether*; *wonder* occurs 93 times with *that* and 1,359 times with *whether*; and *believe* occurs 9,478 with *that* and 0 times with *whether*.

I then fit a mixed effects logistic regression with a binomial likelihood predicting *whether* (1) v. *that* (0), with fixed effects for CONTEXT (*present*, *past*, *perfect*, *progressive*, *modal*, *infinitival*) as in (27), POLARITY (*negative*, *positive*), and their interaction, as well as by-verb random intercepts. The inclusion of random intercepts for VERB ensures that high frequency verbs do not have an outsized influence on the fixed effects estimates, since their baseline preferences for *whether* or *that* complements are accounted for by those intercepts. The interaction of CONTEXT and POLARITY is reliable in a likelihood ratio test ($\chi^2(5) = 4820.5$, $p < 0.001$) and so I investigate the coefficients of the model containing fixed effects for CONTEXT, POLARITY, and their interaction.

8 This string-based method has additional, (I believe) less important limitations. For example, it excludes any cases of transitive verbs taking anything besides a pronoun as an object as well as any cases where an adverbial intervenes between the lexical verb and the complementizer. I know of no reason to believe that these exclusions are themselves correlated with TAM, in which case they do not present a problem for the current analysis.

In a reference coding with *present* and *positive* as the reference levels for CONTEXT and POLARITY, respectively, the fixed intercept is reliably negative ($\beta = -6.89$, $z = -32.62$, $p < 0.001$); and the simple effects of *progressive* ($\beta = 0.82$, $z = 14.88$, $p < 0.001$), *modal* ($\beta = 1.50$, $z = 58.93$, $p < 0.001$); and *infinitival* ($\beta = 2.39$, $z = 126.19$, $p < 0.001$) are reliably positive, while the simple effects of *past* ($\beta = -1.51$, $z = -48.30$, $p < 0.001$) and *perfect* ($\beta = -2.20$, $z = -36.02$, $p < 0.001$) are reliably negative. This pattern suggests that, among positive polarity items, there is a substantial bias for declarative-taking across predicates, but that, when accounting for that overall bias, a verb takes polar interrogative complements significantly more frequently when it is in the progressive, within an infinitive, or under a modal than when it is in the simple present and significantly less frequently than when it is in the simple past and perfect.

Further, a similar pattern does not hold among the negative polarity contexts. With the same coding, there is a reliably positive simple effect of POLARITY ($\beta = 3.18$, $z = 122.93$, $p < 0.001$); and the interactions of POLARITY with *progressive* ($\beta = -2.40$, $z = -16.05$, $p < 0.001$), *modal* ($\beta = -1.23$, $z = -27.04$, $p < 0.001$), and *infinitival* ($\beta = -4.37$, $z = -29.60$, $p < 0.001$) are reliably negative, while the interactions with *past* ($\beta = 1.21$, $z = 25.45$, $p < 0.001$) and *perfect* ($\beta = 1.59$, $z = 15.89$, $p < 0.001$) are reliably positive. The simple effect of negation suggests that, as predicted by Mayr (2018), a verb in the present tense takes polar interrogatives more frequently when it is negated than when it is not, and the interactions suggest that the TAM-associated pattern observed with positive polarity are distinct from those observed with negation.

5.3 Discussion

These results suggest that the pattern observed in §4 for *think*, *believe*, *hope*, and *fear* is not specific to those predicates: it appears to be part of a more general pattern that progressive, modal, and infinitival morphology tend to correlate with a higher relative frequency of polar interrogative-taking. Further, this correlation does not appear to be driven by other potentially important factors, such as whether the embedding predicate occurs in a downward-entailing context—at least one created by negation close to the predicate. It remains unclear, however, (i) whether downward-entailing contexts are confounded with TAM morphology; and if not, (ii) why exactly TAM morphology is correlated with interrogative-taking in the first place.

Insofar as the pattern among positive items is in fact driven by TAM, one possibility is that the relevance of TAM to interrogative-taking might be at least partially linked with event structural properties such as stativity, durativity, and telicity (White & Rawlins 2018a; see also Kratzer 2006, Moulton 2009, Bogal-Allbritten 2016). For instance, the fact that polar interrogative complements are significantly more frequent when a verb is in the progressive than when it is in the present may suggest the relevance of eventivity to interrogative selection. More would need to be said here, however, since it is unclear how best to fit the facts about modal and infinitival contexts into this picture.

6 Conclusion

The proposals discussed here pursue the laudable goal of associating selection with independently motivated lexical properties, such as veridicality, neg-raising, and preferentiality. This means that the generalizations on which these proposals rest are predictive and thus falsifiable. I have presented evidence from both lexicon-scale experimental data and examples attested in corpora that these generalizations are, if not outright falsified, so frail that they should be jettisoned.

Importantly, though, this finding does not imply that proposals wherein clause selection is a function of semantic selection (Grimshaw 1979) should be abandoned wholesale. It remains a live possibility that some alternative set of lexical properties might be found that predict clause-selection—e.g. event structural properties, such as stativity, durativity, and telicity.

References

An, Hannah & Aaron White. 2020. The lexical and grammatical sources of neg-raising inferences. *Proceedings of the Society for Computation in Linguistics* 3(1). 220–233. http://dx.doi.org/10.7275/yts0-q989.

Bartsch, Renate. 1973. Negative transportation" gibt es nicht. *Linguistische Berichte* 27(7).

Berman, S.R. 1991. *On the semantics and logical form of wh-clauses*. Amherst, MA: University of Massachusetts at Amherst dissertation.

Bogal-Allbritten, Elizabeth A. 2016. *Building Meaning in Navajo*: University of Massachusetts, Amherst dissertation.

Davies, Mark. 2008. The Corpus of Contemporary American English (COCA). https://www.english-corpora.org/coca/.

Davies, Mark. 2018. The iWeb Corpus. https://www.english-corpora.org/iWeb/.

Egré, Paul. 2008. Question-embedding and factivity. *Grazer Philosophische Studien* 77(1). 85–125. http://dx.doi.org/10.1163/18756735-90000845.

Elliott, Patrick D, Nathan Klinedinst, Yasutada Sudo & Wataru Uegaki. 2017. Predicates of relevance and theories of question embedding. *Journal of Semantics* 34(3). 547–554. http://dx.doi.org/10.1093/jos/ffx008.

Gajewski, Jon Robert. 2007. Neg-raising and polarity. *Linguistics and Philosophy* 30(3). 289–328. http://dx.doi.org/10.1007/s10988-007-9020-z.

George, Benjamin Ross. 2011. *Question Embedding and the Semantics of Answers*: University of California Los Angeles dissertation.

Ginzburg, Jonathan. 1995. Resolving questions, II. *Linguistics and Philosophy* 18(6). 567–609. http://dx.doi.org/10.1007/BF00983299.

Grimshaw, Jane. 1979. Complement selection and the lexicon. *Linguistic Inquiry* 10(2). 279–326.

Hintikka, Jaakko. 1975. Different Constructions in Terms of the Basic Epistemological Verbs: A Survey of Some Problems and Proposals. In *The Intentions of Intentionality and Other New Models for Modalities*, 1–25. Dordrecht: D. Reidel.

Karttunen, Lauri. 1977a. Syntax and semantics of questions. *Linguistics and Philosophy* 1(1). 3–44. http://dx.doi.org/10.1007/BF00351935.

Karttunen, Lauri. 1977b. To doubt whether. In *The CLS Book of Squibs*, Chicago: Chicago Linguistic Society.

Kratzer, Angelika. 2006. Decomposing attitude verbs.

Lahiri, Utpal. 2002. *Questions and Answers in Embedded Contexts*. Oxford University Press.

Mayr, Clemens. 2018. Predicting polar question embedding. In Robert Truswell, Chris Cummins, Caroline Heycock, Brian Rabern & Hannah Rohde (eds.), *Proceedings of Sinn und Bedeutung 21*, 863–880. Edinburgh: Semantics Archive.

Moulton, Keir. 2009. *Natural Selection and the Syntax of Clausal Complementation*: University of Massachusetts, Amherst dissertation.

Roberts, Tom. 2019. I can't believe it's not lexical: Deriving distributed veridicality. *Semantics and Linguistic Theory* 29. 665–685. http://dx.doi.org/10.3765/salt.v29i0.4634.

Spector, Benjamin & Paul Egré. 2015. A uniform semantics for embedded interrogatives: An answer, not necessarily the answer. *Synthese* 192(6). 1729–1784. http://dx.doi.org/10.1007/s11229-015-0722-4.

Sæbø, Kjell Johan. 2007. A Whether Forecast. In Balder D. ten Cate & Henk W. Zeevat (eds.), *Logic, Language, and Computation* Lecture Notes in Computer Science, 189–199. Berlin, Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-540-75144-1_14.

Theiler, Nadine, Floris Roelofsen & Maria Aloni. 2017. What's wrong with believing whether. *Semantics and Linguistic Theory* 27. 248–265. http://dx.doi.org/10.3765/salt.v27i0.4125.

Theiler, Nadine, Floris Roelofsen & Maria Aloni. 2019. Picky predicates: Why *believe* doesn't like interrogative complements, and other puzzles. *Natural Language Semantics* http://dx.doi.org/10.1007/s11050-019-09152-9.

Uegaki, Wataru. 2015. *Interpreting Questions under Attitudes*: Massachusetts Institute of Technology dissertation.

Uegaki, Wataru & Yasutada Sudo. 2019. The *hope-wh puzzle. *Natural Language Semantics* 27(4). 323–356. http://dx.doi.org/10.1007/s11050-019-09156-5.

White, Aaron Steven & Kyle Rawlins. 2016. A computational model of S-selection. *Semantics and Linguistic Theory* 26. 641–663. http://dx.doi.org/10.3765/salt.v26i0.3819.

White, Aaron Steven & Kyle Rawlins. 2018a. Question agnosticism and change of state. In Robert Truswell, Chris Cummins, Caroline Heycock, Brian Rabern & Hannah Rohde (eds.), *Proceedings of Sinn und Bedeutung 21*, 1325–1342. Edinburgh: Semantics Archive.

White, Aaron Steven & Kyle Rawlins. 2018b. The role of veridicality and factivity in clause selection. In Sherry Hucklebridge & Max Nelson (eds.), *Proceedings of the 48th Annual Meeting of the North East Linguistic Society*, 221–234. Amherst, MA: GLSA Publications.

White, Aaron Steven & Kyle Rawlins. 2020. Frequency, acceptability, and selection: A case study of clause-embedding. *Glossa* 5(1). 1–105. http://dx.doi.org/10.5334/gjgl.1001.

Wurmbrand, Susi. 2014. Tense and aspect in English infinitives. *Linguistic Inquiry* 45(3). 403–447. http://dx.doi.org/10.1162/ling_a_00161.

Zuber, Richard. 1982. Semantic restrictions on certain complementizers. In S. Hattori & K. Inoue (eds.), *Proceedings of the 12th International Congress of Linguists*, 434–436. Tokyo.

Zwarts, Frans. 1998. Three types of polarity. In Fritz Hamm & Erhard W. Hinrichs (eds.), *Plurality and Quantification*, vol. 69 Studies in Linguistics and Philosophy, 177–238. Dordrecht: Springer. http://dx.doi.org/10.1007/978-94-017-2706-8_5.

Aaron Steven White
511A Lattimore Hall
Department of Linguistics
University of Rochester
500 Joseph C. Wilson Blvd.
Rochester, NY 14627
aaron.white@rochester.edu