Non-Reducibility with Knowledge *wh*:
Experimental Investigations

DRAFT - DO NOT CIRCULATE WITHOUT PERMISSION

Jonathan Phillips and B. R. George

**Abstract** Most accounts of knowledge *wh* seek to reduce it to knowledge *that*. That is, facts about answer knowledge are deemed sufficient to determine the truth value of knowledge *wh* statements. George (2011, 2013) argues against the possibility of such a reduction. On George’s proposal, standing in the knowledge *wh* relation to a question requires the agent to both know at least one answer and not believe any false answer.

George’s (2011, 2013) main empirical claim rests on contested judgments that have not been investigated in a systematic way. In order to evaluate George’s case against reducibility, we set out to experimentally assess the claim that competent speakers will not assent to the truth of knowledge *wh* ascriptions if the agent in question believes false answers, regardless of whether the agent possesses answer knowledge. In line with George’s (2011, 2013) proposal, we found that false belief defeats knowledge *wh*, even when it does not defeat knowledge *that* for answers, supporting the key non-reducibility claim. In contrast to George’s picture, however, speaker judgments tracked to the proportion of the agent’s beliefs that were false, not there mere presence of false beliefs.

1 *know wh* and the reducibility problem

In linguistic and philosophical work on question-knowledge ascriptions, a common point of departure is the assumption that this knowledge *wh* is in some sense reducible to knowledge of the facts that answer or resolve the interrogative object of *know*.

Roughly, the starting assumption is that the truth or falsehood (1-a) is supposed be determined, in principle, by the kinds of knowledge ascriptions made by sentences like (1-b), or more abstractly, by those with a general scheme something like what is given in (1-c):

\[
\text{(1) } \begin{align*}
\text{a. } & \text{ John knows where he can buy an Italian newspaper.} \\
\text{b. } & \text{ John knows that he can buy an Italian newspaper at PaperWorld.} \\
\text{c. } & \text{ John knows that he can(’t) buy an Italian newspaper at } x.
\end{align*}
\]

If this picture is right, then there is nothing special about the epistemology of knowledge *wh* – a theory of knowledge *that* and a theory of question-answering (or
question-resolution) should be expected to exhaust all the interesting data.

Formalizing a bit further, on the reductive picture, sentences like (1-a) are analyzed as some variant of (2), where \( p \) is a variable over propositions (or whatever other theoretical objects serve as potential answers):\(^1\)

\[
(2) \quad \exists p \text{ s.t. } p \text{ answers/resolves “Where can J. buy an Italian newspaper?” and John knows } p.
\]

More generally, we seem to have something like the following general picture, where \( \alpha \) names individual \( a \), and \( \pi \) is an interrogative:

\[
(3) \quad \text{Whether a knowledge ascription of the form “} \alpha \text{ knows } \pi \text{” is entirely determined by which (potential) answers for } \pi \text{ are among the things } a \text{ knows.}
\]

The thesis here is one of reducibility. It does not require that the analysis of know wh be stated directly in terms of such a reduction. Some authors who are committed to (3) would nevertheless encode this by writing a distinct lexical or meaning postulate for know wh that happens to make reference to the relation of propositional knowledge, in effect treating them as distinct words. What (3) requires is just that the world’s facts for know that suffice to determine its facts for know wh. The widespread (though often tacit) acceptance of (3) points to the central role of propositional knowledge in the analysis of question-knowledge.

Most accounts of know wh in the formal semantics literature are quite reducibility-friendly. This trend is seen in the early influential work of Karttunen (1977) and Groenendijk & Stokhof (1984), and continues through Beck & Rullmann (1999) and Lahiri (2002),\(^2\) and a wide variety of more recent sources.\(^3\) Recently, something close to the reducibility thesis seems to be endorsed by Klinedinst & Rothschild (2011), who take it as a virtue of their account that it does not require separate lexical entries for question-embedding uses, take it as a virtue of their account that “those verbs that embed both questions and that-clauses make the same semantic

---

1 See, e.g., Higginbotham (1996) for a representative statement of this idea.
2 Lahiri (2002) and Beck & Rullmann (1999) employ relatively flexible general frameworks that make it possible to implement nonreducibility. In both cases, this non-reducibility is somewhat marginal. For Lahiri it potentially arises from a special parameter of lexically determined domain restriction in answer quantification, and for Beck and Rullmann it is stipulated into the lexical entry for agree wh in an entirely ad hoc manner, and isn’t discussed in the more general remarks on the treatment of question embedding (indeed, it is not recognized as a form of non-reducibility at all, and is, curiously, described as a case of weak exhaustivity). In both cases, the nonreducibility of the theory can be noted on careful inspection, but is not remarked upon explicitly, and is not applied to know wh or closely related attitudes. See George (2011), Chemla & George, and Spector & Égré (2015) for some discussion of the non-reducibility of agree in these types of accounts.
3 Working in a very different semantic paradigm, Ginzburg (1996) nevertheless persists in analyzing knowledge wh in terms of knowledge of the facts that resolve the question.
contribution in all cases,” where this semantic contribution is implemented as a predicate of propositions.

In the philosophical literature, reducibility has been met with more skepticism. Early reservations regarding this picture were expressed by Ryle (1945), who, famously, argued that know how to is fundamentally different from know that. A more recent line of objection is offered by Schaffer (2007), who articulates and explicitly rejects a reductive picture. This rejection is based on specific assumptions about what constitutes an answer, and connects with Schaffer’s broader research program in the contrast-sensitivity of knowledge ascriptions. We will not discuss the details of either of these here, but the reader is referred to Stanley (2011) for one defense of reducibility against these arguments.

We take it that reducibility is an appealing point of departure because it is the most obvious way (though not, in principle, the only one) of capturing the important intuition that know wh and know that involve the same know (and similarly for forget, be certain (about), and so on). It says that there is little or no special work that needs to be done on the epistemology or semantics of know wh – given a sufficient theory of propositional knowledge and an adequate theory of answers, we can, on this reductive picture, regard the analysis of know wh as a relatively modest – perhaps almost trivial – exercise in combining these. If we reject reducibility, we are forced to contend with the possibility that know wh offers fresh surprises, not evident (and in principle not discoverable) from the analysis of know that. Challenges to reducibility thus have the potential to seriously complicate the study of knowledge ascriptions.

A recent challenge to these reductive accounts of know wh has come from George (2011, 2013), who argues for a difference in judgments about (4) between cases such as All True and Mixed below:

(4) Sue knows where she can buy an Italian newspaper.

(5) (All True) Sue is standing on the street near a store called Newstopia. Sue’s friend, Bob, a native of the city who is normally very well-informed and trustworthy, told her that she can buy an Italian newspaper at Newstopia. Having no reason to doubt this, Sue took Bob at his word. Bob was correct about Newstopia, which does sell Italian newspapers.

(6) (Mixed) Sue is standing on the street near two stores: one called PaperWorld, and another called Newstopia. Sue’s friend, Bob, a native of the city who is normally very well-informed and trustworthy, told her that she can buy an Italian newspaper at PaperWorld, and also at Newstopia. Having no reason to

---

4 George frames things in terms of two different people’s knowledge and beliefs, while we frame things in terms of two competing scenarios involving the same person. This distinction is relevant for certain aspects of presentation, and perhaps of George’s formulation of the notion of reducibility, but the key truth-conditional claims remain the same.
doubt this, Sue took Bob at his word. However, Bob completely misinformed Sue about PaperWorld, which does not sell newspapers, but actually just sells stationery and office supplies. Bob was correct about Newstopia.

As George argues, Sue seems to know the same facts regarding question (7) (assuming, as seems reasonable in this case, her true beliefs on the basis of Bob’s reports all constitute knowledge), but, according to George’s characterization of knowledge \( wh \), (4) is true in the All True case not in the Mixed case.

(7) Where can one buy an Italian newspaper?

The issue, according to George, is that the difference in judgments for (4) in All True and Mixed is not explicable in terms of differences in any kind of answer-knowledge. Although Sue believes different answers in the two scenarios, these differences of belief involve only false answers, and so do not bear on Sue’s answer-knowledge, which remains the same.

If, as in the approach given in (2), knowledge of questions is determined entirely by answer-knowledge, then the type of difference described above cannot be explained. George proposes instead that knowledge \( wh \) needs to be analyzed non-reductively in terms of knowledge and something like belief: to stand in the knowledge \( wh \) relation to a question is to know at least one answer, and to not believe any false answer.

If this is right, it is a major challenge to the large segment of work on knowledge \( wh \) that has assumed that truth-conditions of knowledge \( wh \) ascriptions can be given in terms of answer knowledge, and has concerned itself in large part with debates over which notion(s) of answerhood are relevant.

In this paper, we concern ourselves with the empirical exploration of ordinary truth evaluations of the sort knowledge \( wh \) ascriptions which drove George’s ar-

---

5 For this reasoning to go through, we should also require that all of Sue’s beliefs on the topic of (7) be subjectively indistinguishable from knowledge, and to impose some other requirements along these lines. See George (2011, 2013) for some discussion of these details and a slightly different scenario.

6 Strongly exhaustive answers, of the sort employed in partition-based accounts like Groenendijk & Stokhof (1984), appear superficially helpful for these examples, but, as George notes, Sue does not in either case know whether any other stores sell Italian newspapers, so if knowledge of a strongly exhaustive answer is required, we should expect that (4) is false in both scenarios, and this does not appear to be the judgment.

7 Some additional care is needed here: for example, we may need to make some (probably inoffensive) supporting assumptions about the nature of answerhood, and we need to stipulate that what counts as an answer is not parameterized by the subject of the knowledge ascription. Again, the reader is referred to George (2011, 2013) for a more complete sketch of the issues.

8 There are numerous notions of answer that might be considered, of which George (2011) employs more than one, but the type of answer most relevant to all of our examples is that of a mention-some answer.
argument. George’s reported judgments are not as clear-cut as we might like: the
truth judgment of (4) in All True appears relatively unobjectionable, but although it
appears informally that many speakers reluctant to affirm (4) in Mixed, some also
seem reluctant to affirm its negation or declare it false.

The issue is further complicated by context-sensitivity: if Newstopia is much
more situationally salient than PaperWorld, then (4) appears less objectionable. It
seems plausible that, as George suggests, this could be the result of a general domain
restriction effect, in which case it would not compromise the basic claim about
non-reducibility, but the existence of this effect still makes informal comparison of
small numbers of individual judgments difficult to interpret.

In order to evaluate the empirical status of this challenge to reduction, along
with George’s positive claim about the analysis of knowledge wh, a more nuanced
empirical picture is needed. We take a first step in addressing this lacuna, and report
a series of experiments intended to assess native speaker truth-value judgments of
know wh.

2 An Initial Test

We began by testing whether ordinary truth assessments reflect George’s core judg-
ments for the type of scenario described in All True and Mixed. The aim of this
first pass was to establish a rough sense of the basic empirical status of George’s
reported intuitions, and the plausibility of the associated non-reducibility claims.

2.1 Methods

103 (53 females, age $M(SD) = 34.32(12.11)$) participants with American IP ad-
resses were recruited from Amazon’s Mechanical Turk (www.mturk.com) and paid
a small amount of money (< $0.50) in compensation for their time (Buhrmester
et al. 2011, Paolacci & Chandler 2014, Sprouse 2011). Additional demographic in-
formation about these participants can be found in the supplementary materials. For
this study, and all of the following, all stimuli, data, analyses, and other supporting
materials can be retrieved from: https://github.com/phillipsjs/Knowledge_wh

Participants read vignettes All True (in which Sue knows an answer and has
only true beliefs) and Mixed (in which Sue knows an answer and has relevant false
beliefs) in counterbalanced order. After reading each vignette, participants were
asked to indicate their agreement with (4) on a scale from 1 (“Completely disagree”)
to 7 (“Completely agree”).

(4) Sue knows where she can buy an Italian newspaper.

Finally, participants were asked to summarize the difference between the two stories
and then completed a series of demographic items including a question asking whether English was their native language.

2.2 Results

2 participants were excluded from the analyses because they reported that English was not their native language. For simplicity, we first consider only participants’ responses on the first trial. In All True, when Sue had only true beliefs about stores that sold Italian newspapers, participants tended to strongly agree with (4) ($M(SD) = 6.23(1.35)$). This is expected under George’s characterization of the facts, and under any analysis that assigns a mention-some reading. However, in Mixed, when Sue had one true and one false belief, participants agreed significantly less with (4) ($M(SD) = 4.88(2.16)$), $t(79.83) = 3.75$, $p < .001$, $d = .757$ (Fig. 1).

While this difference is again expected on George’s account, which claims (4) is untrue, it is not predicted by an analysis of (4) in which only Sue’s knowledge of answers is bears on the truth of the sentence (especially if, as seems appropriate for (4), the emphasis is on mention-some answers).

Additional analyses revealed that this pattern did not change when considering participants’ responses to both the first and second trials and was not affected by the order in which participants completed the tasks.

2.3 Discussion

The results of this study provide some initial evidence that native-speaker truth value judgments align with George’s claims insofar as speakers distinguish between the truth status of (4) in the cases with and without false beliefs. At the same time, participants’ level of agreement also differs from what would be expected for a clear false case. This is not necessarily a problem for George’s general non-reducibility claim, but it does count strongly against the specific sample implementations in George (2013) and George (2011), which both place the “no false beliefs” requirement among the core truth conditions (albeit while noting that this is preliminary and that presuppositional possibilities also require consideration). This intermediate agree-

---

9 Throughout, we focus on first-trial results, and conduct only one or two trials per participant. These decisions are motivated by a desire to minimize demand characteristics of the study which might bias participants’ responses by causing them to focus on the aspects of the experiment that vary between trials.

10 A paired-samples analysis of participants’ responses to both conditions again revealed a difference in participants’ agreement ratings in the two conditions $t(100) = 3.94$, $p < .001$, $d = .529$, and a comparison of linear mixed-effect models (Bates et al. 2012, Jaeger 2008, Gelman & Hill 2006) revealed that there was no main effect of the order in which the items were completed, $\chi^2(1) = 0.17$, $p = .682$, and no Condition*Order interaction effect, $\chi^2(1) = 2.51$, $p = .113$. 

6
Figure 1: Boxplots of participants’ agreement rating with the knowledge ascription in the All true and Mixed conditions. Colored dots depict individual participants’ responses, the box plot “hinges” show the 25th and 75th percentiles, and the horizontal black bars indicate participants’ median response.

Agreement rating does not naturally fall out of a simple truth-conditional account. The question we face is one of whether it belongs to the semantic/conventional meaning at all. If the reduced agreement results from a semantic presupposition or similar conventional not-at-issue meaning, or if it represents a borderline case associated with conventional but vague or somehow degree-sensitive truth-conditions, then, plausibly, we should consider this a problem for the lexical/semantic/conventional meaning of know, and a challenge to a reducibility-friendly semantics for treatment of know wh. Alternatively, the degraded agreement might be the result of an unsatisfied conversational implicature: that is, it might be that a reductive semantics is fine, but that entirely extasemantic considerations of rational communicative behavior produce this effect, at least given normal contextual assumptions. The problem of teasing these alternaties apart is not simple, but a number of the experiments below take some first steps towards narrowing down the possibilities.
3 A More Robust Test

We next replicated and expanded on this initial study in two substantial ways. First, because of the intermediate agreement ratings in the previous study, we wanted to get a better sense of the truth-status of know wh ascriptions in cases where the agent has both true and false beliefs. To this end, we decided to include a comparison case in which the agent’s beliefs were all false and also look at negated knowledge wh ascriptions, which we hoped would shed some light on whether the judgments of reduced truth we saw in the first test were the result of at-issue falsehood, or some other status such as presupposition failure or vagueness-associated borderline falsehood.

Second, we looked at a few closely related kinds of where questions, including in particular infinitival where questions: in addition to the where ... can ... type of mention-some question given in (8-a) we considered infinitival wh questions like (8-b) (which play a central role in the treatment of reducibility for know how ascriptions in Stanley (2011)), and where ... should ... questions like (8-c), which are often described as providing the closest paraphrases of these infinitival questions.

(8) a. Sue knows where she can buy an Italian newspaper.
   b. Sue knows where to buy an Italian newspaper.
   c. Sue knows where she should buy an Italian newspaper.

3.1 Methods

989 (362 females, age $M(SD) = 29.98(9.77)$) participants with American IP addresses were recruited from Amazon’s Mechanical Turk (www.mturk.com) and paid a small amount of money ($< 0.50) in compensation for their time. Additional demographic information about this sample can be found in the suppentary materials.

Participants were randomly assigned either to read the vignettes as described in Section 2.1 or to read vignette All False in which Sue’s belief about where to buy an Italian newspaper was straightforwardly false.

(9) All False Sue is standing on the street near a store called Paperworld. Sue’s friend, Bob, a native of the city who is normally very well-informed and trustworthy, told her that she can buy an Italian newspaper at Paperworld. Having no reason to doubt this, Sue has always assumed that Bob was right. However, Bob completely misinformed Sue about Paperworld, which does not sell newspapers, but actually just sells stationery and office supplies.

This additional condition provided a clearly false baseline against which to compare agreement ratings in the other two conditions. After reading the vignette, participants
were either asked to rate their agreement with one of three forms of a knowledge *wh* ascription (8-a)–(8-c) or were asked to rate their agreement with the negated form of one of these three knowledge *wh* ascriptions (10-a)–(10-c).

(10)  
a. Sue doesn’t know where she can buy an Italian newspaper.  
b. Sue doesn’t know where to buy an Italian newspaper.  
c. Sue doesn’t know where she should buy an Italian newspaper.

Finally, participants were asked to provide a summary of the vignettes they read and then completed a series of demographic items including a question asking whether English was their native language.

### 3.2 Results

22 participants were excluded from the analyses because they reported that English was not their native language. Participants’ first-trial responses were analyzed by comparing linear mixed-effects models (Jaeger 2008, Gelman & Hill 2006), which revealed a main effects of False Beliefs, $\chi^2(2) = 549.73$, $p < .001$ and a False Belief$\times$Negation interaction effect $\chi^2(2) = 29.345$, $p < .001$ (Fig. 2). For simplicity and clarity, we focus here on illustrating the critical patterns with simpler pairwise-comparisons. The complete analyses are provided in the supplementary materials.

We first consider participants’ agreement ratings with the the standard (non-negated) knowledge *wh* ascriptions (8-a)–(8-c). Participants agreed with the knowledge *wh* ascription significantly more when all of Sue’s beliefs were true ($M(SD) = 6.24(1.20)$) than when Sue had both true and false beliefs ($M(SD) = 4.24(1.81)$), $t(213.87) = 10.00$, $p < .001$, $d = 1.285$. Additionally, participants agreed with the knowledge *wh* ascription in the *Mixed* case significantly more than the case in which Sue had only false beliefs ($M(SD) = 1.97(1.46)$), $t(205.08) = 12.08$, $p < .001$, $d = 1.432$.

This pattern both replicates and extends the findings of the initial test in Section 2: we again observed that participants’ agreement ratings were sensitive to all of the agent’s beliefs across a number of different forms of knowledge *wh* ascriptions (as predicted by George). Comparing agreement ratings in the cases of *Mixed* true and false beliefs to the new *All False* baseline condition, we also observed that

---

11 The overall model for participants’ first-trial responses took the form Knowledge $\sim$ Beliefs $\times$ Negation $+ (1|\text{questionModality})$. Main effects and interactions were tested for by comparing minimally differing models as suggested by Barr et al. (2013). In these initial analyses, agreement ratings with the negated ascription were reverse coded to make them more directly comparable. As in the previous study, including participants’ second-trial responses did not qualitatively affect the patterns observed.
participants agreed significantly more in the Mixed cases than they did in cases where all of the agent’s beliefs were false.

Participants’ agreement ratings with the negated knowledge wh ascriptions (10-a)—(10-c) exhibited a corresponding pattern. Participants agreed significantly less with the negated knowledge wh ascription when all of Sue’s beliefs were true ($M(SD) = 2.84(2.09)$) than when Sue had both true and false beliefs ($M(SD) = 3.50(2.04)$), $t(219.52) = -2.41, p = .017, d = 0.322$. Additionally, participants agreed with the negated knowledge wh ascriptions in the Mixed case significantly less than in the All False case ($M(SD) = 5.66(1.60)$), $t(205.08) = 12.08, p < .001, d = 1.232$. While this pattern revealed that participants’ agreement ratings for negated knowledge wh ascriptions were also sensitive to the agent’s false beliefs, it is worth noting that there was somewhat less of a difference between the All True and Mixed cases for the negated statements than for non-negated statements, giving rise to a significant Belief*Negation interaction effect $\chi^2(1) = 15.59, p < .001$. 

Figure 2: Boxplots of participants’ agreement rating with the non-negated (left) and negated (right) knowledge ascriptions in the All True (red) and Mixed (green) and All False (blue) conditions for all three forms of knowledge wh statements.
3.3 Discussion

These results reinforce our initial observation, and suggests that it occurs across a wide variety of ways of making a question explicitly or implicitly modal. The intermediate agreement ratings for the mixed case already suggested that something other than simple at-issue falsehood might be at work, and the the failure to pattern with the (clearly false) all false case further supports this position. The question remains one of whether this is a semantic phenomenon, or a conversational implicature. If the ordinary sentences are true but have an unsatisfied implicature in the mixed case, we would expect the negated sentences to at least as bad as if they were simply false in this case, so the fact that they are judged a bit better than the negated sentences in the all true case offers some preliminary evidence that a conversational implicature is not at work (we take up this issue more directly in Section 5). In sum, the these first two studies provide relatively clear empirical support for a difference in truth-evaluations of the sort suggested by George (2011, 2013), and accordingly, for the non-reducibility of knowledge wh.

4 A Digression on Knowledge that

In this section, we introduce the first in a series of experiments intended to differentiate a genuine semantic non-reducibility effect from other possible explanations of the observed effects presented in Sections 2 and 3.

So far, we have seen that knowledge wh ascriptions have a different status in scenarios without false belief and in scenarios with false belief. We have suggested that this as evidence that knowledge wh is not reducible to knowledge that. Of course, this conclusion is only warranted if the knowledge that facts are stable across the different scenarios. If, in the kinds of examples discussed, the status of answer-knowledge ascriptions is affected by the presence of false beliefs, then a traditionally reductive description of the facts may well be consistent with the empirical data presented here.

It seems to us intuitively clear that the key knowledge that claims are simply true in both scenarios. However there is, in general, reason to think that false beliefs about related propositions can sometimes render otherwise good knowledge that ascriptions untrue (cf. Goldman (1976)), so this alternative explanation is worth addressing explicitly. To this end, we conducted an experiment to assess the truth of knowledge that claims in the situations described by vignettes All True, Mixed, and All False.
4.1 Methods

161 (41 females, age $M(SD) = 31.17(10.27)$) participants with American IP addresses were recruited from Amazon’s Mechanical Turk (www.mturk.com) and paid a small amount of money (< $0.50) in compensation for their time. Additional demographic information can be found in the supplementary materials.

The methods for this study are similar to those in Section 3 except that participants were instead asked to rate their agreement with knowledge *that* ascriptions. In the cases where all of Sue’s beliefs were true and where Sue’s beliefs were both true and false, participants rated their agreement with (11-a). Participants who read the case in which Sue’s belief was simply false instead rated their agreement with (11-b), since the this vignette did not mention Newstopia.

(11) a. Sues knows that she can buy an Italian newspaper at Newstopia.
    b. Sue knows that she can buy an Italian newspaper at Paperworld.

After rating their agreement, participants were asked to provide a summary of what they read and then completed series of demographic items including a question asking whether English was their native language.

4.2 Results

3 participants were excluded from the analyses because they reported that English was not their native language. An analysis of the remaining participants’ agreement ratings with (11-a) revealed that there was no significant difference in their agreement ratings when Sue had only true beliefs ($M(SD) = 5.94(1.22)$) and when Sue had both true and false beliefs ($M(SD) = 5.64(1.86)$), $t(71.01) = 0.830$, $p = .409$, $d = 0.184$. For comparison, participants did agree significantly disagree with (11-b) (i.e., with a mean agreement rating below 4) when Sue had only false beliefs ($M(SD) = 2.86(1.99)$), $t(82) = −5.244$, $p < .001$ (Fig. 3).

4.3 Discussion

While we saw above in Sections 2 and 3 that false beliefs reduce truth judgments for knowledge *wh* ascriptions, we find no analogous effect for knowledge *that* ascriptions. This suggests that we have a genuine challenge to reductive accounts: the different scenarios support the same facts regarding knowledge *that*, suggesting that knowledge *that* does not suffice to determine knowledge *wh*. 
Figure 3: Boxplots of participants’ agreement rating with the knowledge that ascription in the All True, Mixed and All False conditions.

5 The Naive Relevance-Implicature Approach

George’s key claims are framed in terms of simple truth and untruth, and do not directly address the distinction between falsehood of at-issue content and failure of semantic presuppositions. Our data, in contrast, show a number of gradability effects, and in particular show intermediate agreement ratings for most of the key cases. These intermediate agreement ratings make it tempting to try to exclude this effect from the semantic or conventional meaning entirely – to say that (4) is simply semantically true in both the All True scenario and the Mixed scenario, but that in the latter it is infelicitous as the result of some more or less Gricean cooperative considerations (cf. Grice (1975)).

(4) Sue knows where she can buy an Italian newspaper.

That is, we might say that the requirement that Sue not believe any false answers is not any part of the semantic or conventional meaning of knowledge w/h ascriptions, but is simply a conversational implicature.

12 George is primarily concerned with arguing that there is a semantic nonreducibility, and less with whether this nonreducibility is located in the at-issue content, the semantic presuppositions, or elsewhere.
There are, of course, numerous kinds of Gricean explanations, and we cannot attempt to evaluate all of them here. However, one kind of explanation along these lines seems especially intuitive. A natural reply to the judgments reported so far is to claim that what is varying is the relevance of the knowledge facts to the implied problem-at-hand, with the Mixed scenario yielding degraded relevance and so degraded felicity.

To flesh this out a bit, one of the main purposes served by knowledge reports is to identify people as information resources who can provide a solution for some problem. If (as is strongly suggested by the choice of example sentences and vignettes) the problem at hand is the problem of procuring an Italian newspaper, then, according to this line of reasoning, reporting that somebody knows where Italian newspaper can be bought should only be felicitous if this knowledge enables them to obtain Italian newspapers and to reliably direct others to places where they can buy such newspapers. Such an ascription would then be infelicitous in the Mixed case, arguably explaining the reported judgment.

In order to directly assess this line of analysis, we constructed a scenario in which knowledge is ascribed to a person who, although they know of a place where Italian newspapers can be bought, is not a useful resource for solving the newspaper-obtaining problem at hand for reasons unrelated to false beliefs. We did this by varying whether the person to whom knowledge is ascribed shares a language in common with the person looking for a newspaper. If usefulness as a problem-solving resource is driving a relevance implicature effect responsible for these judgments, we should expect false beliefs and lack of a shared language to produce similar degraded truth judgments. If the judgments reported above are instead part of the conventional semantics of know wh, we should expect the two cases to yield very different truth assessments.

5.1 Methods

257 (84 females, age $M(SD) = 29.53(9.12)$) participants with American IP addresses were recruited from Amazon’s Mechanical Turk (www.mturk.com) and paid a small amount of money (< $0.50) in compensation for their time. Additional demographic information can be found in the supplementary materials.

All participants read a vignette about a woman named Sue who wanted to know where she could buy an Italian newspaper and about a man named Bob who held beliefs about two stores which either did or did not sell Italian newspapers. To directly manipulate the relevance of the knowledge facts to the problem-at-hand (Sue’s finding an Italian newspaper), we randomly assigned participants to read either a vignette in which Bob and Sue could not communicate because they did not share a common language (12) or a vignette in which they could communicate
because Bob and Sue did share a common language (13).

(12) Sue, who speaks only Italian and English, needs to buy an Italian newspaper. She is standing on the street near two stores: one called PaperWorld, and one called Cellulose City. Sue sees a man named Bob nearby. Bob is a native of the city who is normally very well-informed and trustworthy. Bob speaks only English and Hungarian. He believes that PaperWorld and Cellulose City sell Italian newspapers. He would be happy to tell this to Sue if she asked and he understood her. Moreover, Bob should be able to understand Sue’s question, since they can speak the same language.

(13) Sue, who speaks only Italian and English, needs to buy an Italian newspaper. She is standing on the street near two stores: one called PaperWorld, and one called Cellulose City. Sue sees a man named Bob nearby. Bob is a native of the city who is normally very well-informed and trustworthy.

Bob speaks only Cantonese and Hungarian. He believes that PaperWorld and Cellulose City sell Italian newspapers. He would be happy to tell this to Sue if she asked and he understood her. However, Bob would not be able to understand Sue’s question, since they can’t speak the same language.

For each of these two conditions, participants were randomly assigned to read a case in which Bob’s beliefs about where to buy an Italian newspaper were all true (14-a), were both true and false (14-b), or were all false (14-c) (as in the previous studies).

(14) a. Bob is right about both stores: PaperWorld and Cellulose City both sell Italian newspapers.
   b. Bob is right about PaperWorld. However, Bob is mistaken about Cellulose City, which does not sell Italian newspapers. (It is actually a stationery shop.)
   c. Bob is completely mistaken about PaperWorld and Cellulose City, neither of which sells Italian newspapers. (They are actually both stationery shops.)

After reading the vignette, participants rated their agreement with (15) on a scale from 1 (‘Completely disagree’) to 7 (‘Completely agree’).

(15) Bob knows where Sue can buy an Italian newspaper.

Subsequently, participants completed an item that was designed to measure the extent to which we were successful in manipulating the relevance-implicature (16).
Because Sue needs an Italian newspaper, she calls her friend Mary, to ask her what she should do, and happens to mentions that she sees Bob nearby. Mary, who knows that Bob likes to keep up with the latest news from Italy, tells Sue

“Bob knows where you can buy an Italian newspaper.”

Given everything you know about Bob, how useful will Mary’s information be to Sue?

Participants indicated their agreement on a scale from 1 (‘Not at all useful’) to 7 (‘Very useful’). Participants also completed two comprehension questions that simply tested whether they thought that Bob would understand Sue’s question if she asked it in English and whether they thought it was likely that Sue would succeed in finding an Italian newspaper, given her situation. Finally, participants answered a series of demographic items including whether English was their native language.

5.2 Results

8 participants were excluded from the analyses because they indicated that English was not their native language. An additional 16 participants were excluded for failing to correctly answer the first comprehension question.13

First, in order to ensure that we successfully manipulated the relevance of the knowledge ascription by changing whether Bob and Sue shared a common language, we considered participants’ judgments of how useful Mary’s information was for Sue as indicated by their rating of (15) (Fig. 4). A 2 (Language: Shared vs No Shared) x 3 (Belief: All True vs Mixed vs All False) Analysis of Variance (ANOVA), revealed a main effect of whether Bob and Sue shared a language, $F(1,227) = 42.04, p < .001 \eta^2_p = .156$, and a main effect of false beliefs, $F(2,227) = 56.11, p < .001 \eta^2_p = .327$. Most importantly, though, we also observed a Language*False Belief interaction, $F(2,227) = 42.04, p < .001 \eta^2_p = .077$, which suggests that the impact of the truth of Bob’s beliefs on judgments of usefulness differed when Sue and Bob shared a language and when they did not. Further investigation of this interaction effect revealed the predicted pattern of responses: When Sue and Bob shared a language, the usefulness of Mary’s information was significantly affected by whether Bob’s beliefs were all true ($M(SD) = 6.38(1.00)$) or were both true and false ($M(SD) = 5.15(1.51)$), $t(67.84) = 4.269, p < .001, d = 0.954$.

The second comprehension question was not used to exclude participants because of an unforeseen lack of a clarity about what the “correct” answer was in a couple of the conditions. Regardless, the qualitative pattern of results remains the same even when not adopting any exclusion criteria.
However, when Sue and Bob did not share a language, the usefulness of Mary’s information was no longer significantly affected by whether Bob’s beliefs were all true ($M(SD) = 3.93(2.32)$) or were both true and false ($M(SD) = 3.46(1.80)$), $t(73.34) = 0.991, p = .325, d = 0.222$. Unsurprisingly, when Bob’s beliefs were all false, Mary’s information was never considered useful, regardless of whether Bob and Sue shared a language ($M(SD) = 2.36(1.76)$) or not ($M(SD) = 2.26(1.52)$), $t(69.23) = 0.256, p = .799, d = 0.060$.

Figure 4: Boxplots of participants’ ratings of the usefulness of the statement “Bob knows where you can buy an Italian newspaper” when Bob and Sue shared a language or did not share a language for the All True, Mixed and All False conditions.

The overall pattern of usefulness judgments allows us to appropriately test whether the difference in participants’ agreement ratings with the knowledge $wh$ ascriptions can be adequately explained by the relevance of the knowledge facts to the implied problem-at-hand. If this pragmatic line of analysis is correct, participants’ agreement ratings with (15) should show a similar pattern such that the truth of Bob’s beliefs affects participants’ agreement with (15) when Bob and Sue share a language, but does not affect their agreement when Sue and Bob do not share a language, since the information is equally unhelpful in these latter cases.

To investigate whether this is the case, we next analyzed participants’ agreement ratings with (15) (Fig. 5). A 2 (Language: Shared vs No Shared) x 3 (Belief:
All True vs Mixed vs All False) ANOVA revealed a main effect of the truth of Bob’s false beliefs, $F(2, 227) = 256.17, p < .001 \eta^2_p = .693$, and a marginal effect of whether Bob and Sue shared a language, $F(1, 227) = 3.58, p = .060 \eta^2_p = .016$. Critically, however, we did not observe a Language*False Belief interaction effect, $F(2, 227) = 0.26, p = .773 \eta^2_p = .002$, which suggests that, unlike judgments of usefulness, agreement ratings with the the knowledge $wh$ ascription were not differentially impacted by the falsehood of Bob’s beliefs when Sue and Bob shared a language, compared to when they did not. Even when Bob and Sue did not share a language, participants’ agreed with the knowledge $wh$ ascription less when Bob had both true and false beliefs ($M(SD) = 5.21(1.69)$), than when Bob had all true beliefs ($M(SD) = 6.63(0.67)$), $t(49.34) = 4.89, p < .001, d = 1.111$. Similarly, when Bob and Sue shared a language, participants’ agreed with the knowledge $wh$ ascription less when Bob had both true and false beliefs ($M(SD) = 4.90(1.63)$) than when Bob had all true beliefs ($M(SD) = 6.15(1.10)$), $t(68.39) = 4.89, p < .001, d = 0.899$. Moreover, comparing the two cases in which Bob has both true and false beliefs revealed that there is no significant difference in participants’ agreement ratings with (15) when Bob and Sue share a language and when they do not $t(76.71) = −0.82, p = .417, d = 0.184$, suggesting that agreement with the knowledge $wh$ ascription in the Mixed cases was not significantly affected by the relevance of the knowledge facts to the implied problem at-hand.¹⁴

5.3 Discussion

We think that these results count strongly against a relevance-implicature analysis of the key judgments for knowledge $wh$ ascriptions. When the fact of Bob’s answer-knowledge is rendered irrelevant to Sue’s problem by a factor other than Bob’s false beliefs (here a lack of a shared language) truth-judgments were not diminished in the same way, even though judgements of usefulness were. Thus, it would appear that false beliefs have some significance that goes beyond diminishing the relevance or utility of the underlying knowledge facts and that, unlike relevance or utility diminishers, false beliefs affect truth values.

Pragmatic accounts clearly deserve further investigation, and no one experiment will systematically rule out the possibility that the judgments that we’ve been

¹⁴ A similar result is obtained when analyzing the data by comparing linear mixed models to test for three-way interaction which should be expected if the two different questions (knowledge $wh$ agreement vs. usefulness) are differentially sensitive to the interaction effect between the truth of Bob’s beliefs and whether Bob and Sue shared a language. As predicted, we find evidence for the three-way interaction when comparing a model that includes the critical three-way interaction term (Judgment $\sim$ Question * Language * Beliefs + (1|Subject)) to one that does not contain that term (Judgment $\sim$ Question+Language + Question+Beliefs + Language+Beliefs + (1|Subject)), $\chi^2(1) = 15.742, p < .001$
discussing might have some Gricean explanation. Still, we think the above results help to put the burden of proof on the advocate of a specific Gricean account. What we have seen so far is that the predictions of the most intuitive form of a Gricean approach (when applied in a straightforward manner) were not confirmed.

6 Who Knows How

Above, we focused on a know where ascription that is familiar from the literature as an example of a mention-some reading for an embedded question (that is, a reading where partial information without false beliefs does not suffice to compromise the truth of the ascription). It is, naturally, important to see whether this effect extends to other kinds of knowledge wh. In this section we consider know who ascriptions (as another relatively widely-studied kind of knowledge wh) and know how ascriptions (which have received particular attention in discussions of reducibility in the philosophical literature, as in, e.g., Ryle (1945), Stanley (2011)).
6.1 Methods

364 (125 females, age \( M(SD) = 30.80(9.74) \)) participants with American IP addresses were recruited from Amazon’s Mechanical Turk (www.mturk.com) and paid a small amount of money (< $0.50) in compensation for their time. Additional demographic information can be found in the supplementary materials. Participants either completed the “knows how” or “knows who” version of the study.

In the “knows how” version, participants read a vignette that described a man named Jason who needs to know how to get to a hardware store because of a leaking pipe in his house (17).

(17) Jason recently moved to a new town, and while he knows his way around the town, he doesn’t know where the hardware store is. He needs to find a hardware store because there’s a leaking pipe in his house. When Jason asks his neighbor where he can find a hardware store, his neighbor tells him that there’s one four miles northeast and on the left side of the street. But when Jason asks his coworker where he can find a hardware store, his coworker tells him that there’s one six miles south and on the right side of the street. Jason knows that his neighbor and his co-worker have lived in the town for about the same amount of time.

After participants read this vignette they were either told that all of Jason’s beliefs were true (18), that they were both true and false (19), or they were all false (20).

(18) There is a hardware store four miles northeast and on the left side of the street. There is also a hardware store six miles south and on the right side of the street.

(19) There is a hardware store four miles northeast and on the left side of the street. There is no hardware store six miles south and on the right side of the street.

(20) There is no hardware store four miles northeast and on the left side of the street. There is also no hardware store six miles south and on the right side of the street.

After reading one of the three versions of this vignette, participants rated their agreement with (21).

(21) Jason knows how to get to a hardware store.

The “knows who” version was similarly structured. Participants read a vignette that described a woman named Amy who needs to know who has a MiniDisplayPort adapter so that she can connect her computer to a projector. Amy’s friend Bert
tells her that two different coworkers have the adapter she needs. Similar to the the “knows how” version, Amy’s beliefs could have all been true, have been both true and false, or all have been false. Participants were randomly assigned to one of these three conditions. We also included an additional condition to ensure that this vignette correctly elicited a mention-some reading. In this version, Amy has only true beliefs about who has a MiniDisplayPort adapter, but is ignorant of an additional coworker who also has the adapter she needs. After reading the vignette, all participants rated their agreement with (22).

(22) Amy knows who can lend her an adapter.

In both versions of the study, participants rated their agreement on a scale from 1 (‘Completely disagree’) to 7 (‘Completely agree’). Additionally, participants who read about Amy answered a comprehension question which asked them to indicate which of Amy’s coworkers actually had the adapter that Amy needed. Finally, all participants answered a series of demographic items including whether English was their native language.

6.2 Results

16 participants were excluded because they indicated that English was not their native language. An additional 21 participants were excluded because they did not correctly answer the comprehension question in the “knows who” version. An analysis of the mention-some test for the “knows who” version revealed that participants mean agreement with (22) was significantly above the midpoint of 4 ($M(\text{SD}) = 5.67(1.37)$), $t(44) = 8.19$, $p < .001$, $d = 1.22$ suggesting that a mention-some reading was clearly available in the knows who case. Accordingly, we proceeded to consider whether knowledge $wh$ truth assessments were affected by truth of all of the agents’ beliefs.

The primary 2 (Knowledge-$wh$ ascription: who vs how) x 3 (Beliefs: All True vs Mixed vs All False) analysis of variance of the remaining 282 participants’ agreement ratings revealed that their agreement was significantly influenced by whether the subject had false beliefs $F(2,276) = 130.33$, $p < .001$ $\eta_p^2 = .482$ (Fig. 6). Most importantly, participants agreed with the knowledge ascriptions more when the subjects beliefs were all true ($M(\text{SD}) = 5.61(1.72)$) than when they were

---

15 In George (2011, 2013), the argument for non-reducibility depends both on the judgment of untruth with false-beliefs, and on the availability of mention-some or similar truth-conditions in the absence of such false beliefs. Most of our other are very similar to classic mention-some examples from the literature, and had, in our judgment, clearly available mention-some truth-conditions. This example, however, was farther from established mention-some examples, so, as a precaution, we decided to confirm the availability of a mention-some reading experimentally.
both true and false ($M(SD) = 4.15(1.89)$), $t(184.18) = 5.50$, $p < .001$, $d = 0.802$. Moreover, in the “knows who” Mixed condition, participants’ agreement was also significantly lower ($M(SD) = 4.64(1.96)$) than their agreement in the additional “knows who” mention-some test, where the agent was simply ignorant of relevant answer knowledge rather than having a false belief $t(72.64) = -2.81$, $p = .006$, $d = 0.610$.

We additionally observed an unpredicted main effect such that participants were overall more willing to agree with the knowledge ascription in “knows who” version ($M(SD) = 4.46(2.34)$) than in the “knows how” version ($M(SD) = 3.39(2.12)$), $F(1,276) = 27.10$, $p < .001$ $\eta^2_p = .089$. While this difference was not anticipated, it is relatively unsurprising given the number of differences between the “knows who” and “knows how” vignettes, and more importantly, we did not observe a significant interaction effect $F(1,276) = 2.85$, $p = .060$ $\eta^2_p = .020$, suggesting that the presence of false beliefs had at least a somewhat similar impact of both types of knowledge wh.

---

Figure 6: Boxplots of participants’ agreement ratings with the knowledge wh ascription in the All True, Mixed and All False conditions for knows how (left) and knows who (right).
6.3 Discussion

These results indicate that the effect we observe is replicated in a variety of types of wh questions, including the who questions that are often used as the standard case in work on wh questions semantics, and the how questions that have received special attention in philosophical work on reducibility (Ryle 1945, Stanley 2011). Given the similarity of the pattern observed here to the pattern in the four preceding studies, there is good reason to think that the effect we’ve demonstrated throughout applies quite generally to knowledge wh. Taken together, these five studies provide robust support for there being some difference in truth-evaluations as suggested by George (2011, 2013), and thus for the non-reducibility of knowledge wh.

7 Proportional Knowledge wh

While the results seen so far provide support for George’s proposed difference in truth-evaluations in the All True and Mixed cases, they diverge in important respects from what one would expect given George’s truth-conditional account. In particular, George’s account predicts substantially stronger disagreement than we observed with the knowledge wh ascription in cases where the agent has both true and false beliefs.

One way of accounting for this discrepancy would be to propose an alternative explanation for participants’ diminished agreement ratings that does not involve outright semantic untruth — that is, one could agree that George’s intuitions are widely shared but argue that George was wrong to account for the intuitive difference in the semantics of knowledge wh. Thus far, however, our attempts to test the most obvious versions of such proposals have been unsuccessful (see Sections 3, 4 & 5). An alternative way of accounting for participants’ higher than expected agreement with knowledge wh ascriptions in the Mixed cases would be to argue that George’s analysis was correct in providing a semantic/conventionalized account of the role for false beliefs in knowledge wh, but wrong in the particulars of the role that they were given in George’s (2011, 2013) analysis. We now want to consider this second possibility.

While the previous studies were informative in demonstrating that false beliefs affect truth assessments of knowledge wh ascriptions, they are not particularly informative as to how false beliefs affect participants’ truth assessments. What would be helpful in assessing this alternative proposal is an additional study that may shed more light on the role of false beliefs in knowledge wh. As a first step to addressing this issue, we designed a final test that allowed us to come closer to parametrically varying the proportion of the agent’s beliefs that were false, which should allow us to begin measuring how participants’ truth assessments of knowledge
wh ascriptions vary as function of the increasing presence of false beliefs.

7.1 Methods

171 (46 females, age \(M(\text{SD}) = 29.61(9.83)\)) participants with American IP addresses were recruited from Amazon’s Mechanical Turk (www.mturk.com) and paid a small amount of money (< $0.50) in compensation for their time. Additional demographic information can be found in the supplementary materials.

All participants read a vignette in which a woman named Sue was told where to buy a newspaper by her friend Bob (23).

(23) Sue is standing on the street near three stores: one called PaperWorld, one called Cellulose City, and one called Newstopia. Sue’s friend, Bob, a native of the city who is normally very well-informed and trustworthy, told her that PaperWorld, Newstopia, and Cellulose City all sell Italian newspapers. Having no reason to doubt this, Sue has always assumed that Bob was right.

Participants were told that Bob was either correct about all three stores (24), correct about two of the three stores (25), correct about only one of the stores (26), or incorrect about all of the stores (27).

(24) Bob was correct about all three stores, which do in fact all sell Italian newspapers.

(25) However, Bob completely misinformed Sue about PaperWorld, which does not sell newspapers, but is actually a stationery shop. Bob was correct about Newstopia and Cellulose City.

(26) However, Bob completely misinformed Sue about PaperWorld and Cellulose City, which do not sell newspapers, but are actually both stationery shops. Bob was correct about Newstopia.

(27) However, Bob completely misinformed Sue about all three stores, none of which sell Italian newspapers. PaperWorld and Cellulose City, are actually both stationery shops, and Newstopia is an ironically misnamed shop that sells T-shirts with obnoxious political slogans.

After reading the vignette, participants rated their agreement with (4), as in previous studies.

(4) Sue knows where she can buy an Italian newspaper.

Finally, participants were asked to summarize the story they read and then to complete a series of demographic items including a question asking whether English
was their native language.

7.2 Results

3 participants were excluded from the analyses because they reported that English was not their native language. Agreement ratings from the remaining 168 participants revealed an effect of how many of the agent’s beliefs were false $F(3,164) = 52.74$, $p < .001 \eta^2 = .491$ (Fig. 7). Specifically, participants agreed most strongly with (4) when Sue’s beliefs were all true ($M(SD) = 6.17(1.39))$, less when one of Sue’s beliefs was false ($M(SD) = 5.31(1.69)$), even less when two of Sue’s beliefs were false ($M(SD) = 3.77(2.01)$), and least of all when Sue’s beliefs were all false ($M(SD) = 1.88(1.60)$). In each case, the presence of an additional false belief significantly reduced participants’ agreement ratings ($p$’s < .05; $d$’s > .56).

![Figure 7: Agreement rating with the knowledge $wh$ ascription as a function of the portion of the agent’s beliefs that were false.](image)

7.3 Discussion

As these results demonstrate, truth assessments of knowledge $wh$ ascriptions are sensitive not only to the mere presence of false beliefs but also to the proportion of the agent’s beliefs that are false. Moreover, participants on average disagreed
with the knowledge \textit{wh} ascription when the agent had proportionally more false beliefs than true beliefs. Such results both provide strong support for some form of nonreducibility and require an account of knowledge \textit{wh} with the resources to handle the sort of proportionality effect demonstrated here. We discuss two approaches to this in the next section.

8 Theoretical Implications and Interpretation

The observations presented here point to a semantic/conventional impact of false beliefs on the truth of knowledge \textit{wh} ascriptions. This is at odds with typical reductive accounts of knowledge \textit{wh}, and appears problematic for reducibility generally.

Thus, our observations seem to support a claim of non-reducibility. In broad terms, this non-reducibility involves the kinds of factors (besides answer-knowledge) proposed by George (2011, 2013). For a variety of types of questions, we find that we can get something like mention-some truth conditions in the absence of false beliefs, together with diminished agreement in the case where there are additional false beliefs. This seems to be a genuine challenge to answer-based reducibility, as the introduction of false beliefs does not reduce the agreement for knowledge \textit{that} ascriptions for answers.

Moreover, we find that the diminished agreement pattern seen in these cases is not found in cases where there are only true beliefs, but where relevance or utility of the statement is otherwise compromised, suggesting that the obvious conversational implicature treatment cannot straightforwardly account for these effects.

Together, these observations would appear to support non-reducibility. The epistemology of knowledge \textit{wh} seems not to be reducible to knowledge \textit{that}, at least not in terms of something like the reduction concept at work in available accounts.

It is much less clear what the correct positive account of knowledge \textit{wh} should look like. In particular, George’s descriptive generalization (one of the few contenders explicitly designed to provide for this type of reducibility) fails to capture many of our observations. For George, what is required for knowledge \textit{wh} (at least with regard to these types of questions) is knowledge of a single answer (here a \textit{mention-some answer}), without belief in any false answer. As George (2011) formalizes things, knowledge \textit{wh} ascriptions for which these truth-conditions are not satisfied are simply false, although George discusses the possibility of making the “no false beliefs” requirement a semantic presupposition.

Contrary to the formal implementation sketched in George (2011), we find that these cases produce intermediate agreement judgments, distinct from the judgments in cases of clear falsehood, and, as we saw in Section 7, the level of agreement appears, tentatively, to deteriorate by steps as the proportion of known answers to false beliefs shifts – an effect which is completely unexpected on the simple universal
quantification characterization offered by George.

This type of gradable agreement is not easy to directly model in a classical two-valued semantics (or even one supplemented with presupposition failure as an additional option). If we adopted an account with gradable truth, it would be relatively straightforward to characterize things descriptively. The truth-degree of this type of knowledge \( wh \) ascription would scale with the portion of believed answers that are known true rather than false: if all believed answers are known true answers, then the ascription will be fully true, if all believed answers are false (or if no answers are believed), the ascription will be fully false, and for intermediate proportions we will have intermediate truth degree, with higher degrees of truth as the known true answers account for a larger proportion of the believed answers. The details will of course depend on which theory of gradable truth one adopts, but the general shape of the path forward within such a formalism is clear. Such an account would be non-reductive, but would be able to capture the observations above.

Barring the explicit inclusion of gradable truth in our basic characterization of knowledge \( wh \), a more circuitous analysis will be needed. One natural account of gradable truth judgments like those seen in Section 7 is to treat them as borderline cases within an otherwise more-or-less classical system. We might, following many popular semantic treatments of gradability, say that there is some threshold sufficient for truth, determined by a contextual or otherwise flexible parameter. This parameter might further be to some extent underspecified or be open to some level of accommodation.

The exact choice of formal implementation is not critical here, but, to get a sense of how this should work, consider a canonical gradable predicate like \( tall \). Such a predicate admits borderline cases – cases of people or things that we would be reluctant to call either clearly tall or clearly not tall. It also exhibits context-sensitivity: the criteria for clear inclusion in the extension of \( tall \) depend, among other things, on the comparison class or standard of height that is relevant to the conversation at hand.

One kind of way of handling these phenomena in a bivalent semantics is to analyze \( tall \) as something like \( taller \ than \ h \), where the exact value of the height variable \( h \) is not clearly fixed by the lexicon, but is allowed to vary with conversational context. It might be that the public context does not fully resolve a consensus

---

16 True answers that are believed but not strictly known will complicate things in the usual ways, and the right descriptive generalization here requires further investigation, as our vignettes were specifically designed to avoid the complications raised by such cases.

17 Further work would, of course, be necessary to determine the extent to which such an account could meet the goal, endorsed in George (2011), of uniformity across different attitudes, if such uniformity is deemed desirable.

18 Related options include variations on conspicuously tall among the members of \( X \), where \( X \) is some contextually specified comparison set, and there are of course many other options, using different
choice of $h$, in which case, when the choice of $h$ is not sufficiently finely-resolved to determine whether a particular object is tall, we should expect speakers to express some amount of uncertainty or partial agreement.

The details of tall are of course considerably more complex, but this notion of sensitivity to a contextually supplied threshold, domain, or standard of comparison is found in many analyses of many phenomena, and it allows us to analyze intermediate truth/agreement judgments as representing hearer’s reluctance, but not flat-out unwillingness, to consider a value of this contextual variable that would make the sentence under consideration true.

George gave the truth-conditions of (28) as the conjunction of (29) and (30-a), but what our observations above suggest are that we might do better to analyze it by conjoining (29) with something like (30-b), (30-c), or (30-d). That is, it seems preferable to analyze it in terms of some fuzzier, more standard-dependent requirement on avoiding false belief, rather than the absolute prohibition on false beliefs advocated by George.

(28) John knows where he can buy an Italian newspaper.
(29) There is $x \ s.t. \ J. \ knows \ he \ can \ buy \ an \ Italian \ newspaper \ at \ x$.
(30)  
  a. For every $x \ s.t. \ J. \ believes \ he \ can \ buy \ an \ Italian \ newspaper \ at \ x$, he in fact can.
  b. For many $x \ s.t. \ J. \ believes \ he \ can \ buy \ an \ Italian \ newspaper \ at \ x$, he in fact can.
      (Where the amount that constitutes many is in part determined by conversational context.)
  c. For enough $x \ s.t. \ J. \ believes \ he \ can \ buy \ an \ Italian \ newspaper \ at \ x$, he in fact can.
      (Where the amount that constitutes enough is in part determined by conversational context.)
  d. For more than $n\%$ of the $x \ s.t. \ J. \ believes \ he \ can \ buy \ an \ Italian \ newspaper \ at \ x$, he in fact can.
      (Where $n$ is a contextually supplied threshold.)

The intermediate agreement would then reflect differences among speakers about how to set the standard involved in the evaluation of (30-b), (30-c), or (30-d), as well as uncertainty on the part of individual speakers regarding how to set the standard.

These types of gradability and context-sensitivity are widespread, and not in any way special to knowledge $wh$. For our purposes, what is important is that, given a contextually supplied threshold, any of the criteria given in (30) and indeed any types of contextual parameter in different ways.
criterion which, like them, make reference to false beliefs, would be problematic for reducibility.

9 Open Problems

Our present empirical investigation suggests that truth judgments for knowledge \( wh \) are affected by false beliefs, and that this gives rise to a pattern of judgments incompatible with standard reduction accounts, and with reducibility generally. Still, we have left many issues unresolved.

It is not clear exactly what the role of false belief is in these examples. The results in Section 7 help to flesh out the picture some, but the various partly degraded truth-value judgments are, as already noted, difficult to account for with the tools typically employed in the formal semantics of question embedding. A full positive account of knowledge \( wh \) is still needed, and more theoretical and empirical work will be needed to flesh out such an account.

We have also not looked beyond \( know \) to other attitudes. Numerous other propositional attitude predicates can embed questions, and others may give rise to non-reducibility effects (George (2011) suggests \( forget \) as a likely candidate, and Chemla & George present experimental evidence for nonreducibility \( agree \)). Similar effects with other attitudes are possible and deserve to be explored. It would also be desirable to produce some general theory of question-embedding, and of the relationship between propositional and question-oriented uses of an embedder. Ideally, such an account would identify what is uniform across different embedders, and what degrees of freedom are available: if reducibility must be rejected, we should have some story of what takes its place as the link between attitudes \( that \) and attitudes \( wh \).

Many accounts for this are available, but few provide a reasonable level of uniformity while allowing for non-reducibility effects. Lahiri (2002) offers one account that allows for something like non-reducibility, but, as George (2011) discusses, it seems ill-suited to the effects discussed here. George’s own attempt, meanwhile, makes predictions that are cast into doubt by the observations about proportionality in Section 7. It remains to be seen whether George’s account can be revised to adequately address these cases,\(^{19}\) and of course the possibility of an entirely new account deserves serious consideration.

\(^{19}\) George’s account also faces some difficulties with \( tell \)-like embedders, and in particular with the kinds of readings reported by Klindedinst & Rothschild (2011).
References


Bates, Douglas, Martin Maechler & Ben Bolker. 2012. lme4: Linear mixed-effects models using s4 classes.


Lahiri, Utpal. 2002. *Questions and answers in embedded contexts*. Oxford University
Paolacci, Gabriele & Jesse Chandler. 2014. Inside the turk understanding mechanical turk as a participant pool. *Current Directions in Psychological Science* 23(3). 184–188.


