

ARTICLE

# A Corpus Study of “Know”: On The Verification of Philosophers’ Frequency Claims about Language

Nat Hansen<sup>1\*</sup>, J.D. Porter<sup>2</sup> and Kathryn Francis<sup>1</sup>

<sup>1</sup>Department of Philosophy, University of Reading and <sup>2</sup>Stanford University Literary Lab

\*Corresponding author. Email: [n.d.hansen@reading.ac.uk](mailto:n.d.hansen@reading.ac.uk)

(Received 29 May 2018; revised 9 January 2019; accepted 4 April 2019)

## Abstract

We investigate claims about the frequency of “know” made by philosophers. Our investigation has several overlapping aims. First, we aim to show what is required to confirm or disconfirm philosophers’ claims about the comparative frequency of different uses of philosophically interesting expressions. Second, we aim to show how using linguistic corpora as tools for investigating meaning is a productive methodology, in the sense that it yields discoveries about the use of language that philosophers would have overlooked if they remained in their “armchairs of an afternoon”, to use J.L. Austin’s phrase. Third, we discuss facts about the meaning of “know” that so far have been ignored in philosophy, with the aim of reorienting discussions of the relevance of ordinary language for philosophical theorizing.

**Keywords:** discourse markers; frequency; J.L. Austin; know; knowledge; linguistic corpora; meaning; ordinary language philosophy

## 1. Philosophers and observations about “ordinary” use

Observations about what we “ordinarily” or “typically” or “most commonly” mean by some expression of interest have been a staple of philosophical practice at least since the heyday of ordinary language philosophy in the 1950s.<sup>1</sup> For example, here are contemporary examples of philosophers making frequency claims about how people ordinarily talk about someone *knowing* something:

Also, it is worth keeping in mind that most of the time, outside of epistemology, when we consider whether somebody knows something, we are mainly interested in whether the person has the information, not in whether the person’s belief rises to the level of knowledge. Ordinarily we do not already assume that they have a true belief and just focus on whether or not their epistemic position suffices for knowing. Similarly, when we say that someone does not know something, typically we mean that they don’t have the information. (Bach 2005a: 62–63)

‘I know that such and such’ is far more commonly used in situations in which the obtaining of such and such is not in question and no one is in need of being assured

<sup>1</sup>For a sample drawn from JSTOR, see the Appendix.

of it. Think, for example, of the ‘I know’ of sharing a reaction to a piece of purported news, or the ‘I know’ of acknowledging a significant fact. (Baz 2012: 40)

Both Bach and Baz make frequency claims about knowledge claims or uses of “I know that such and such”, and use those observations in support of methodological claims about how best to investigate knowledge. Bach argues that because the knowledge claims of ordinary speakers aren’t usually about “whether or not their epistemic position suffices for knowing”, the use of such judgments is not a reliable guide to the nature of knowledge that philosophers care about, which *is* about what kind of epistemic position is sufficient for knowing. Baz, in contrast, argues that philosophers’ investigations of knowledge will be distorted if they ignore the most common uses of “I know that such and such”.

The central aim of this paper is to answer the following question: What would justify philosophers’ frequency claims about how people ordinarily talk? Tversky and Kahneman (1973) and a wave of subsequent research have made it clear that our frequency judgments sometimes do not track objective frequencies, so we should be skeptical of armchair claims about how common certain uses are in ordinary situations. Baz’s judgment that one particular use of “I know that such and such” is more frequent than another might be the result of that use being more easily recalled, rather than it being genuinely more frequent, for example.

Fortunately for the philosopher of language, however, there now exist resources to investigate objective frequencies of occurrences of linguistic expressions, namely *linguistic corpora*, which are organized bodies of text, purpose-built for answering linguistic questions, such as what the comparative frequencies of various expressions are (Bluhm 2016: 91).<sup>2</sup> Saebo (2004) sets out the reasons for supplementing and correcting armchair linguistic judgments with linguistic corpora:

It will be seen that a corpus can force us to revise old hypotheses and that it can inspire new ones. Corpora offer an insurance against noxious idealizations [covert simplifications of facts, hidden behind selected data], where relevant aspects are disregarded and core facts are missed, by laying bare the relation between raw and interpreted data so it is open for inspection; and they offer a constructive means of assisting the imagination, guiding the researcher towards facts which would otherwise not be thought of. (Saebo 2004: 200)

In this paper, we investigate philosophers’ claims about the frequency of talk about knowledge using linguistic corpora, with a focus on the *Corpus of Contemporary American English*, which is “composed of more than 520 million words in 220,225 texts” and which is “evenly divided between five genres of spoken, fiction, popular magazines, newspapers, and academic journals”.<sup>3</sup> By scouring COCA and other

<sup>2</sup>An early example of a corpus study of [know] is Ludlow (2005), which uses the results of Google searches involving the various modifiers that can combine with [know] as a way of challenging claims about what sorts of arguments exist in the logical form of [know]. More recently, Vetter (2014) uses corpus data to challenge philosophers’ informal judgments about the meaning of disposition ascriptions (“disposed to”), Fischer *et al.* (2015) looks at the distribution of the perception verbs “appears”, “looks”, and “seems” in a corpus for evidence of what those expressions mean, Liao *et al.* (2016) cites corpus data about the relative frequency with which aesthetic adjectives occur with “for-” phrases in support of an argument that aesthetic adjectives behave differently than relative adjectives, and Andow (2015) uses corpora to investigate the changing frequency of the expression “intuition” as used by philosophers. While this paper was in press, Pinillos and Nichols (2018), Sytsma *et al.* (2019), and Meija-Ramos *et al.* (2019), all which make use of corpora to investigate philosophical questions, were published.

<sup>3</sup>Davies (2008); <http://corpus.byu.edu/coca/>.

corpora, we can improve on armchair methods of judging the frequency of expressions of philosophical interest (such as “know”).

## 2. How frequent is “know”?

How do people use “know” when they are speaking and writing (and not doing philosophy)? As Jennifer Nagel observes, “know” is one of the most commonly used verbs in spoken English:

In spoken English (as measured most authoritatively, by the 450-million-word *Corpus of Contemporary American English*),<sup>4</sup> ‘know’ and ‘think’ figure as the sixth and seventh most commonly used verbs, muscling out what might seem to be more obvious contenders like ‘get’ and ‘make’. Spoken English is deeply invested in knowing, easily outshining other genres on this score. In academic writing, for example, ‘know’ and ‘think’ are only the 17th and 22nd-most popular verbs, well behind the scholar’s pallid friends ‘should’ and ‘could’. To be fair, some of the conversational traffic in ‘know’ is coming from fixed phrases, like – *you know* – invitations to conversational partners to make some inference, or – *I know* – indications that you are accepting what conversational partners are saying. But even after we strip out those formulaic uses, the database’s randomly sampled conversations remain thickly larded with genuine references to knowing and thinking.<sup>5</sup>

Nagel uses these frequency facts to back up the idea that investigating “know” and what it (presumably) refers to, namely *knowledge*, is of central human importance.<sup>6</sup> But the fact that “know” occurs relatively frequently in human speech and writing only lends support to the epistemologist’s hope that “know” and *knowledge* are of widespread concern if most or some significant proportion of those occurrences of “know” are being used in a way continuous with the epistemologist’s use. Nagel acknowledges this assumption when she points out that some of the occurrences of “know” in the corpus she is relying on (COCA) are *not* being used in a way that is continuous with the epistemologist’s use – namely, the discourse marker use (what Nagel calls “fixed phrases”) of “you know” and “I know”.

### 2.1 A terminological note, and some comments on Nagel’s frequency claims

When Nagel refers to the frequency of a word like “know”, she is not just talking about the frequency of a particular string (k\_n\_o\_w), but to all forms of the common base form, or *lemma*, for the verb “know”, namely:

- know
- knew
- known
- knows
- knowing
- knowed (this is the regularized version of the irregular past tense “knew”)

<sup>4</sup>The size of COCA has grown since Nagel wrote this passage – it currently contains more than 520 million words.

<sup>5</sup><https://blog.oup.com/2014/09/what-commuters-know-vs-i/>.

<sup>6</sup>Similar observations are made by Pinillos (2012: 193) and Michael Hannon: <https://talkinghumanities.blogs.sas.ac.uk/2018/05/29/whats-point-knowledge-post-truth-era/>. Behind these observations is the assumption “that lexical frequency will correlate with cultural prominence” (Roque *et al.* 2015: 6). But that assumption can’t be correct as it stands, since it’s implausible that “the” – by far the most common word in English – has any special cultural prominence (other than to analytic philosophers of language). Thanks to Chris Kennedy for this observation.

In COCA, it is possible to search for all of the forms of the lemma for “know” by putting the expression in brackets: [know]. For that reason, from now on we will adopt the convention, when talking about all of the verb forms of “know”, of putting the expression in brackets.<sup>7</sup> So, for example, [be] is the most common verb lemma (with forms including “is”, “was”, “be”, “were”, “am”, etc.) followed by [have], and then [do]. The raw counts for all of the forms of the top 12 most frequent verb *lemmas* in COCA are the following:<sup>8</sup>

1. [be] 20,345,475
2. [have] 6,474,715
3. [do] 3,408,919
4. [say] 2,538,688
5. [go] 1,530,661
6. [will] 1,435,282
7. [can] 1,344,061
8. [get] 1,317,745
9. [know] 1,213,282
10. [would] 1,209,395
11. [make] 1,133,929
12. [think] 983,539

Note that these counts are for COCA as a whole, which includes both transcripts of speech and written texts (newspapers, magazines, fiction, and academic texts). In the spoken part of COCA, [know] is the sixth most frequent verb. Nagel observes that the relative frequency of [know] is much greater in spoken English than it is in academic texts, which comes across clearly in [Figure 1](#).

What would explain this substantial difference in the frequency of [know] in speech vs. academic writing? As Nagel points out, some of the occurrences of “know” in speech come from “formulaic” “fixed phrases”, like “you know” and “I know”, which are what linguists would call *pragmatic markers* or *discourse markers*, which, you know, probably don’t occur as frequently in academic writing.<sup>9</sup>

The fact that discourse uses of [know] occur more frequently in spoken than in written English prompts two further questions, which we can use the corpus to try to answer:

- Q1: How frequent are discourse marking uses of [know] in speech, compared to academic writing?
- Q2: Once the discourse marking uses of [know] are excluded, how far does [know] drop down the list of the most frequently occurring verbs in English?

If it turns out that the frequency of [know] in speech is mostly due to the frequency of discourse marking uses, that should weaken the appeal of the epistemologist’s idea that the

<sup>7</sup>Baz (2012: 12 and *passim*) frequently refers to “know” and its “cognates” when he intends to refer to “know”, “knows”, “known”, etc. But a cognate of a word is just something that shares the same etymological origin. “Shirt” and “skirt” are cognates, for example, both being derived from the Old English “skyrt” (a tunic). It is therefore preferable, when investigating the various forms of “know”, to refer to the different forms of the lemma [know], rather than to “know” and its cognates.

<sup>8</sup>For a different list of lemmatized verbs ordered by frequency, see [http://ucrel.lancs.ac.uk/bncfreq/lists/5\\_2\\_all\\_rank\\_verb.txt](http://ucrel.lancs.ac.uk/bncfreq/lists/5_2_all_rank_verb.txt), which is based on the *British National Corpus* ([know] is #12 on that list).

<sup>9</sup>For an overview of theories of discourse markers, see Schiffrin (2001).

Section of COCA	Occurrences of [know]	Section size (M)	Occurrences of [know] in section (%)
Spoken	553,191	109.4	0.51
Academic	68,700	103.4	0.07

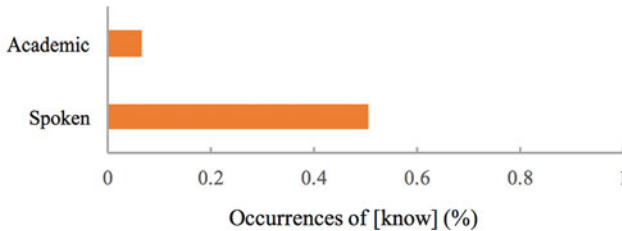


Fig. 1. Comparison of frequencies of [know] between the spoken and academic sections of COCA

frequency of [know] tells us something important about the centrality of *knowledge* in ordinary people's lives. By looking at random samples of occurrences of [know] that occur in transcribed speech, we can get a rough idea of the proportion of those occurrences that are discourse marking uses, as opposed to what Nagel calls "genuine references to knowing".

A random sample drawn from COCA of 500 occurrences of [know] in spoken English yielded 254 discourse marking uses – over 50% of all occurrences of [know].<sup>10</sup> Here are two examples of discourse marking uses of [know]:

(1) And I'm hopeful that some of these viable candidates like Rubio, **you know**, maybe Walker, maybe Bush will be able to capitalize on this as an opportunity to show, **you know**, what Donald Trump really is and actually stand up for what the party is pushing forward which is brighter future for America.<sup>11</sup>

(2) Well, Dr. Drew, what bothers me the most is when Anahita said that innocent people need attorneys, too. I mean, she could have nip this in the butt and just said, "**You know what**, distance myself from my husband. **I don't know – you know**, I am in no way a part of this, and let the investigation conduct on it's own".<sup>12</sup>

In contrast, in a random sample of 500 occurrences of [know] drawn from academic writing, only eight were discourse marking uses, all of which appeared in direct quotations of speech, such as the following:

(3) "I've got a very addictive personality, I was told, **ya know**, maybe that has something to do with it". (Daniulaityte *et al.* 2006)

That aligns with the following observation made by Erman (2001):

[Discourse markers] are all restricted to spoken language (or mimetic dialogue). (Erman 2001: 1339)

<sup>10</sup>All of the random samples discussed in this paper were selected using the "find sample" function on COCA, which selects a random sample from a specified set of examples drawn from the corpus – for example, you can select a random sample of occurrences of [know] drawn from the spoken part of the corpus.

<sup>11</sup>2015, "Donald Trump". Source: SPOK: CNN.

<sup>12</sup>2014, "Hot Car Child Death Case". Source: SPOK: CNN.

The adjusted ranking of the most common verbs in the spoken component of COCA, excluding discourse marker uses from the [know] count, therefore runs as follows:

1. [be] 6,055,987
2. [have] 1,664,956
3. [do] 1,276,348
4. [go] 671,571
5. [say] 573,359
6. [get] 487,694
7. [think] 483,586
8. [will] 397,824
9. [can] 342,526
10. [would] 304,046
- 11? [know] – estimated *non-discourse markers* 296,411–247,928<sup>13</sup>
- 12? [see] – estimated *non-discourse markers* 253,297–247,032<sup>14</sup>
- 13? [make] 250,314
- 14? [want] 249,865

It's important not to put too much weight on the significance of the downgrading of [know] from #6 to #11–14 on this ranking, but Nagel's claim that [know] occurs more frequently than [get] and [make] depends on counting uses of [know] that are discourse markers, which are not, by Nagel's own lights, genuine references to knowing. But what are "genuine references to knowing"?

### 3. "Genuine references to knowing"

Philosophers have analyzed the existence of several different types of knowledge associated with different types of complements of the verb [know]:

- propositional knowledge, associated with a that-clause, or a sentential complement ("I know that it's sunny"; "I know it's sunny")
- knowledge-wh, associated with an embedded question expressed by a wh-expression ("I know when the grocery store closes", "I know who killed Kennedy", "I know how to cold brew coffee")
- objectual knowledge, associated with a noun phrase complement ("I know her"; "I know Wyeth")

<sup>13</sup>Using confidence interval calculations based on our sample, we can be 95% confident that the number of all classifications of [know], excluding discourse marking uses of [know], is within this range (in the spoken portion of COCA). In the sample of 500 occurrences of [know] drawn from the spoken English part of COCA (which contains 553,191 occurrences of [know] total), 50.8% of the occurrences of [know] were discourse markers: margin of error = 4.38%; confidence interval = 46.4–55.2%. Note that according to Cumming and Maillardet (2006), this 95% confidence interval corresponds to an 83.4% capture percentage. That is, there is a 83.4% chance that a repetition of this experiment would produce a mean that falls within the original confidence interval (see Cumming and Maillardet (2006) for a description of capture percentages). Thanks to Shen-yi Liao for discussion of this issue.

<sup>14</sup>There is also a discourse marking use of [see] – "you see" – that is similar to the discourse marker "you know". We sampled 500 random occurrences of [see] taken from the 255,270 total occurrences of [see] in the spoken part of the corpus, and found 10 examples of discourse marking uses (2%). Using confidence interval calculations based on our sample, we can be 95% confident that the number of all likely classifications of [see] as non-discourse markers in the spoken portion of COCA, is within the range given in the list: margin of error = 1.23%; confidence interval = 0.77–3.23%. Thanks to an anonymous referee for asking about the frequency of other discourse markers besides [know]."

Propositional knowledge has received an enormous amount of attention from philosophers (for an overview, see Ichikawa and Steup 2017). Very briefly, philosophers have been concerned with how to characterize knowledge as a relation between a subject and a proposition: is the relation one of justified true belief? Is the relation context sensitive in some way? Is it sensitive to the interests of the knower? Is it unanalyzable? We will take a close look at propositional knowledge in the second part of this paper, when we come to evaluate the Bach and Baz frequency claims mentioned in §1. In this section, we will examine how frequent cases of talk about propositional knowledge are in relation to other types of knowledge-talk.

Knowledge-wh has been analyzed by some philosophers as being a form of propositional knowledge: “S knows wh-” is truth-conditionally equivalent to “there is a proposition *p* such that *x* knows that *p*, and *p* (truly) answers the indirect question of the *wh*- clause” (Brogaard 2009: 439, summarizing the views of Higginbotham 1996; Bach 2005*b*; Braun 2006). Schaffer (2007) and Brogaard (2009) argue against the reduction of knowledge-wh to propositional knowledge. A special case of knowledge-wh that has been the subject of a great deal of philosophical debate is knowledge-how. Stanley and Williamson (2001) and Stanley (2011) offer a detailed (but contentious) account of knowledge-how and its relation to semantic theories of embedded questions and knowledge-wh.<sup>15</sup>

Objectual knowledge is a relation between a subject and a non-propositional object. Roughly, someone knows an object when they stand in some appropriate psychological relation (sometimes called *acquaintance*) to the relevant object.<sup>16</sup> While it is the “default view” that the [know] that appears in propositional knowledge and knowledge-wh involves the same meaning, facts about cross-linguistic variation indicate that the [know] of objectual knowledge and the [know] of propositional knowledge and knowledge-wh have different meanings (Goddard and Wierzbicka 1994: 31–2; Stanley 2011: 36; Nagel 2014: 6; Jary and Stainton 2017: 482). In Table 1, for example, French, Spanish and German each use one verb for propositional knowledge and knowledge-wh (*savoir*, *saber*, *wissen*), and another verb for objectual knowledge (*connaître*, *conocer*, *kennen*). That kind of cross-linguistic variation is evidence that English [know] can have different meanings.<sup>17</sup>

In terms of frequency counts, the ambiguity of English [know] is important, because the frequency of [know] in English will lump together uses that would be distinguished in languages like French, Spanish and German. In the *Corpus de Español* (Web/Dialects), for instance, the Spanish word for propositional knowledge and knowledge-wh, [saber], ranks 15th among verb lemmas, while the word for objectual knowledge, [conocer], ranks 25th.<sup>18</sup> If combined, the counts for the two Spanish verbs for knowledge would rank 10th, similar to the rank of [know] in English. This comparison should be taken with a large grain of salt, since it is not possible to make a one-to-one comparison of English and Spanish, especially when it comes to something like a ranking of the most common verbs. For example, Spanish has two verbs for the English [be] (*ser* and *estar*), and both are in the top 10 Spanish verb lemmas, while no Spanish word directly translates to the English [get], which in Spanish

<sup>15</sup>Objections to Stanley and Williamson’s “intellectualist” theory of knowledge-how are set out in Rumfitt (2003), Glick (2013), Fridland (2015) and Brownstein and Michaelson (2016).

<sup>16</sup>For a survey of competing conceptions of the appropriate psychological relation of acquaintance, see Hasan and Fumerton (2014: §2).

<sup>17</sup>English isn’t alone in this respect; Russian *znaht’*, for example, is similarly applicable to both objectual and propositional knowledge/knowledge-wh.

<sup>18</sup>Davies (2002); <http://www.corpusdelespanol.org/web-dial/>.

**Table 1.** Cross-linguistic evidence that objectual [know] and propositional [know] / [know]-wh have different meanings.

English	French	Spanish	German
I know that it's sunny	Je sais que c'est ensoleillé	Sé que hace sol	Ich weiss, dass es sonnig ist
I know who did it	Je sais qui l'a fait	Yo sé quién lo hizo	Ich weiss, wer es getan hat
I know what time it is	Je sais à quelle heure il est	Sé qué hora es	Ich weiss, wie spät es ist
I know Wyeth	Je connais Wyeth	Conozco a Wyeth	Ich kenne Wyeth
I know Prague	Je connais Prague	Conozco Praga	Ich kenne Prag
I know the killer*	Je connais le tueur	Conozco al asesino	Ich kenne den Mörder

\*Note that “I know the killer” is ambiguous in English between an objectual reading and a “concealed question” reading, on which it means “I know who the killer is” (Heim 1979). As Heim points out (p. 51), when a noun phrase object is permitted with the verb “wissen” in German, it unambiguously gets a concealed question reading, as in “Ich weiss den Mörder schon” (I already know who the killer is) (Engelen 2010: 160). See §4, below, for further discussion of concealed questions.

might be expressed with *llevar*, *tener* or *traer*. The upshot of these cross-linguistic facts is that the [know] that is of primary interest to epistemologists (propositional knowledge and know-wh), is not as frequent as it may appear from a quick inspection of English corpora because a substantial portion of the overall occurrences of [know] are (a) objectual occurrences of [know] (which have a different meaning), and (b) discourse marking uses of [know], which are not “genuine references to knowing”.

#### 4. How to determine relative frequencies of different types of “genuine references to knowing”

Suppose we wanted to determine what proportion of occurrences of [know] refer to propositional knowledge and and knowledge-wh, and what proportion refer to objectual knowledge. How would we do that? One possibility would be to see if it is possible to get a rough sense of the relative frequency of what Nagel calls “genuine references to knowing” by counting the frequency of [know] + complement combinations that characterize knowledge that, knowledge-wh and objectual knowledge. For example, searching for verb forms of [know], followed by a noun, yields a list of examples of what look like objectual knowledge:

1. know people
2. know things
3. knew things
4. knew people
5. know god
6. know shit
7. know John
8. know Jack
9. know women
10. know right



But even a cursory inspection of the examples that contribute to generating this ranking reveals that some of the occurrences of [know] + noun are actually examples of propositional knowledge, such as the following, where there is no explicit “that”-clause, but only a sentential complement:

- (4) I didn’t know people still talked that way.<sup>19</sup>  
 (5) And I know women will continue to pay close attention to all of this.<sup>20</sup>

For the same reason, while searching for “[know] + that” reliably yields examples of propositional knowledge, such a search will miss all occurrences of propositional knowledge that lack a “that”-clause.<sup>21</sup> In a random sample of 100 occurrences of [know], the number of examples of propositional knowledge was roughly divided between those with explicit “that”-clauses (12 occurrences) and those without (15 occurrences). A search that targets explicit “that”-clauses could therefore be missing more than 50% of examples of propositional knowledge.”

Searches for one or two types of words, then, prove inadequate for reliably distinguishing between different forms of [know]. What about more complicated grammatical structures? For instance, propositional uses of [know] should in theory involve clausal complements – the part of the sentence which follows the “that”, as in:

- (6) You know that **I think of him like a son.**<sup>22</sup>

To test this idea, we used Stanford’s CoreNLP parsing program to detect the presence of clausal complements in a sample of 100 occurrences of sentences featuring [know] that we had already tagged as propositional, objectual, and so forth.<sup>23</sup> One potential advantage with such an approach is that clausal complements are detectable even when the word “that” is missing. For example, in the original version of the example above, the speaker did not say “that”, but CoreNLP still identifies “he didn’t get along fine” as a clausal complement governed by “know”. As a result, this parsing might help us identify propositional knowledge claims that simpler searches would miss. Table 2 gives the results produced by the parser, for sentences that we hand coded as propositional.

As Table 2 reveals, targeting clausal complements governed by “know” is an improvement over the simpler searches, but not by much – it still misses about 40% of the uses hand coded as propositional. As a few examples show, these are not particularly unusual or complicated sentences, but they have features, like anaphoric “[know] this” or “[know] that”, or a [know] parenthetical (“as every parent knows”), that block the application of a purely syntactic criterion as a way of identifying propositional occurrences of [know]:

- (7) But now she knows this: EeDee came down from the highsky in a silver ship to bring no good to the World.<sup>24</sup>

<sup>19</sup>Houser (2015).

<sup>20</sup>PBS NewsHour for February 12, 2014. Source: SPOK: PBS.

<sup>21</sup>Searching for “[know] + that” will occasionally turn up examples of anaphoric uses of “that” which refer back to knowledge-wh or objectual knowledge.

<sup>22</sup>Delinsky (2015).

<sup>23</sup>Specifically, we used the “coreNLP” package in CRAN (Manning *et al.* 2014). Sentences containing [know] were annotated and parsed for dependencies. The CoreNLP program is available here: <https://stanfordnlp.github.io/CoreNLP/>. Thanks to Sebastian Schuster for suggesting this approach.

<sup>24</sup>Mason (2015).

**Table 2.** Presence and absence of clausal complements for sentences hand coded as cases of propositional knowledge.

	Has Clausal Complement	Does Not Have Clausal Complement
Propositional	15	11

- (8) However, as every parent knows, kids are great at finding loopholes and at persuading us to do for them what they do not want to do for themselves.<sup>25</sup>
- (9) Did you know that?<sup>26</sup>

The propositional nature of a knowledge claim does not require the presence of a clausal complement governed by “know”, or, based on our examination of the CoreNLP parses, any other obvious features that can be identified purely syntactically.

A further complication in trying to use syntactic criteria to identify types of knowledge is the existence of “concealed questions”. As mentioned in the note accompanying Table 1, sentences that combine certain attitude verbs (like [know]) with nominal complements can be ambiguous between objectual knowledge readings and readings on which they are equivalent to knowledge-wh readings, as in the following example:

- (10) Kim knows the governor of California. (Frana 2017: 1)

Sentence (10) can be understood as saying either that *Kim is acquainted with the governor of California* (which can be the case even if she doesn’t recognize that the person she’s acquainted with is the governor of California), or as saying that *Kim knows who the governor of California is* (which doesn’t require being acquainted with that person).

Heim (1979) proposes “paraphrasability by a wh-clause” as a heuristic for distinguishing concealed question readings of [know] + noun phrase constructions from objectual readings, and she also provides an entailment test for distinguishing the two readings.<sup>27</sup> The following argument is valid on the objectual reading of [knows] + NP, and invalid on the concealed question reading:

1. Kim knows the governor of California.
2. The governor of California is a supporter of high speed rail.
3. Kim knows a supporter of high speed rail.

If Kim is acquainted with the governor of California, and he is a supporter of high speed rail, then Kim is also acquainted with a supporter of high speed rail. But if Kim knows who the governor of California is, it doesn’t follow from the fact that the governor of California is a supporter of high speed rail that Kim knows who a supporter of high speed rail is. Given that the ways of distinguishing a concealed question from a case of objectual knowledge require assessing paraphrasability and entailment, there will be no purely syntactic test for distinguishing the two uses of [know], and

<sup>25</sup>Cortes (2015).

<sup>26</sup>Lupoff (2015).

<sup>27</sup>The entailment test is clearly described in Frana (2017: 2), and we are borrowing her exposition here, with only slight modification to her examples. She also provides an overview of different semantic treatments of concealed questions.

classification will depend on semantic and pragmatic judgments about which reading is more plausible given the surrounding context of each occurrence.

One of the major hurdles in any quantitative humanities project is “operationalizing”, described by literary critic Franco Moretti as “the process whereby concepts are transformed into a series of operations – which, in their turn, allow [us] to measure all sorts of objects” – in other words, figuring out a way to tie a concept like “propositional knowledge” to something we can algorithmically find and count (Moretti 2013). Identifying and counting the different uses of [know] proves difficult to operationalize. The most reliable approach is simply to take a representative sample and hand code each occurrence of [know].<sup>28</sup>

## 5. Lessons from a hand-coded sample of 500 occurrences of [know]

We hand coded a random sample of 500 occurrences of [know]. Our classifications of the sample, in decreasing order of frequency, go as follows (Figure 2):<sup>29</sup>

1. Discourse markers 164 (32.8%).
2. Propositional knowledge 113 (22.6%).
3. Knowledge-wh 101 (20.2%).
4. Knowledge-pp (prepositional phrase) 52 (10.4%).
5. Objectual knowledge 37 (7.4%).
6. Concealed questions 18 (3.6%).
7. Miscellaneous 15 (3.0%).

The most frequent occurrence of [know] in the sample (roughly 1/3) were discourse markers (we will discuss discourse markers in detail below, in §5.1). Occurrences of propositional knowledge (which feature either a “that” clause or a sentential complement of [know]) were roughly as frequent as occurrences of know-wh (which includes “[know]-if”). Occurrences of objectual knowledge are roughly as frequent as [know] + a prepositional phrase (“known as”, “know about”, for example). Least frequent were concealed question occurrences of [know]. Finally, the “miscellaneous” category includes both occurrences of [know] that couldn’t be disambiguated, given the context provided by the corpus (that includes cases of objectual vs. concealed question readings), and what seem like idiomatic uses of [know] (“for all I know”). The straightforward exercise of coding a random sample of [know] raises a couple of philosophically interesting questions:

- What is the significance of the relative preponderance of discourse marking uses of [know] for our understanding of the meaning of [know] in general?
- How should we understand the category of [know] + prepositional phrase? This is a category that philosophers have not devoted any attention to analyzing, even though it occurs roughly as frequently as objectual knowledge.

### 5.1 Discourse marker occurrences of [know] and meanings of [know]

Discourse markers are expressions, such as “I mean”, “well”, and “like”, whose function is to “[monitor] discourse and conversation in various ways” (Erman 2001: 1339), and

<sup>28</sup>For a methodologically similar combination of linguistic corpora and qualitative judgment/coding, see the investigation of verbs of perception in Roque *et al.* (2015).

<sup>29</sup>The sample is available to download here: [https://semanticsarchive.net/Archive/jhiMTE0Z/Corpus\\_samples\\_know\\_hand\\_tagged.xlsx](https://semanticsarchive.net/Archive/jhiMTE0Z/Corpus_samples_know_hand_tagged.xlsx).

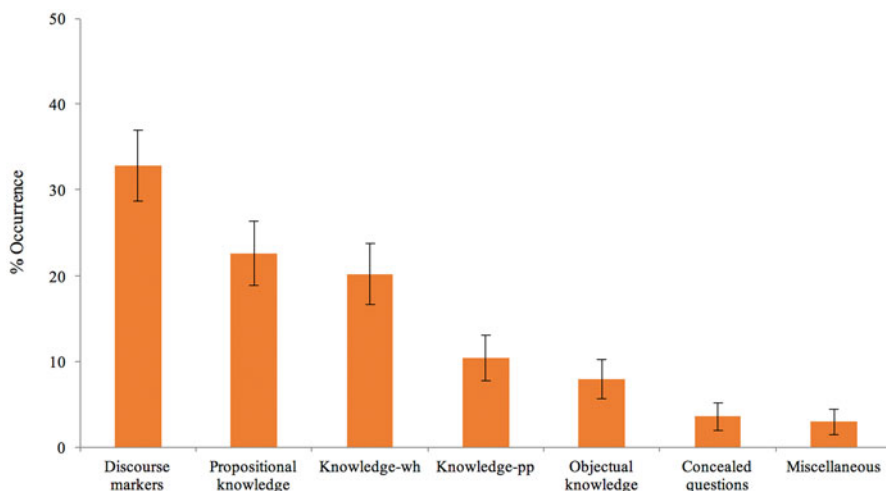


Fig. 2. Percentage of overall occurrences of [know] for each classification (from a random sample of 500 occurrences of [know]). Error bars represent the 95% confidence interval of the proportion.

which are both “syntactically optional in the sense that removal of a DM [discourse marker] does not alter the grammaticality of its host sentence”, and “contribute nothing to the truth-conditions of the proposition expressed by an utterance” (Schourup 1999: 231–2). The following example illustrate these features of discourse markers:

(11) Zelda: Are you from Philadelphia?

Sally: **Well** I grew up out in the suburbs. And then I lived for about seven years up in upstate New York. And then I came back here t’go to college. (Schiffrin 1987: 106)

“Well” can be deleted from Sally’s utterance without altering the grammaticality or truth conditions of what she says, but it is not conversationally superfluous. According to Schiffrin’s (1987) account, Sally’s use of “well” in (11) marks a response to a question that is not “fully consonant” with the expectations of the questioner, by rejecting the assumption that “Are you from Philadelphia?” has a binary answer (Schiffrin 1987: 106).

There is a substantial literature examining the discourse marking role of “you know” (Östman 1981; Schourup 1985; Schiffrin 1987; Erman 2001). Schourup (1985) gives an account of the “core meaning” of “you know” in its discourse marking role in terms of the speaker’s interest in ensuring that the audience grasps the speaker’s intended message:

[Uses of “you know” in its discourse marking role] represent discourse situations in which a speaker might wish to check up on the correspondence of his or her own communicative aims to what the addressee has been able to grasp from what has been said. (Schourup 1985: 128)

The great frequency of [“you know” in its discourse marking role] in [conversation is predictable from its core use. It appears in so many different places because this use is appropriate at any point at which the speaker is unsure of how well s/he is coming across. (Schourup 1985: 139)

Discourse marker uses of [know] are the most common occurrence of [know] in spoken English, and yet this type of use has hardly received any attention from philosophers (there are brief mentions of discourse markers in Camp and Hawthorne 2008: 1–2, and in Predelli 2013: 68 n. 13). Philosophers' near-exclusive attention to propositional content tends to make the variety of things we can do with [know] invisible, but a quick look at a corpus snaps those background features of language back into focus.

The prevalence of discourse marking uses of [know] might be taken to lend further support to an argument against the use of ordinary language in epistemology expounded by Hazlett (2010). Hazlett observes that non-factive uses of [know] can occur in non-philosophical conversations without any sense of unacceptability, as in the following example:

(12) Everyone knew that stress caused ulcers, before two Australian doctors in the early 80s proved that ulcers are actually caused by bacterial infection. (Hazlett 2010: 501)<sup>30</sup>

Hazlett argues that the apparent acceptability of non-factive uses of [know] in ordinary talk means that the ordinary concept of knowledge, which guides non-philosophers' use of [know], is not the same as the concept of knowledge that is of interest to epistemologists, which *is* factive. He concludes that "traditional epistemology shouldn't be especially interested in the concept of knowledge that serves as the meaning of 'knows' in ordinary talk" (Hazlett 2010: 499). (Of course, one might instead – like Baz – agree with Hazlett that there's a difference between ordinary uses of [know] and philosophers' uses of [know], and conclude instead that it is the concept associated with the philosophers' use that is of less interest.)

Given the frequency of discourse marking uses of [know], one could construct a very similar argument that ordinary language occurrences of [know] shouldn't be used in theorizing about knowledge, since the meaning of [know] includes uses that clearly are not "genuine references to knowing" (to use Nagel's phrase), but rather play a role in structuring and monitoring discourse. It would, therefore, be a mistake to look at ordinary use as a guide to what the philosophically interesting notion of knowledge is. We think, however, that such an argument isn't convincing.

As discussed above, in §3, there is cross-linguistic evidence that English [know] has different meanings when it refers to propositional knowledge and objectual knowledge. Other languages distinguish meanings of [know] that English lumps together. And there are similar cross-linguistic reasons in favor of thinking that discourse marking uses of [know] don't mean the same thing as propositional occurrences of [know]. While there is evidence that [know], with its propositional meaning, is a linguistic *universal* (Goddard and Wierzbicka 1994; Wierzbicka 1996), the discourse marking use of [know] is variable across languages, indicating that it has a different meaning in English than the propositional meaning of [know].<sup>31</sup> For example, Magnigová (2016: 60–1)

<sup>30</sup>Bach (2005a: 62) also observes that ordinary uses of "know" sometimes look non-factive: "For instance, we all know people who insist that they 'knew' things that they now acknowledge to be false. So does knowledge not even entail truth?"

<sup>31</sup>For a discussion of some evidence in favor of thinking that the propositional meaning of [know] is a linguistic universal, see Nagel (2017: 10): "The World Loanword Database, which covers 41 languages from all inhabited continents, lists words meaning *know* and *think* (both in the sense embedding a propositional complement) in every language (Haspelmath & Tadmor, 2009). Verbs meaning *know* and *think* not only seem to appear in all languages but are heavily used across languages, appearing among the thirty most common verbs in languages as diverse as Arabic (Buckwalter & Parkinson, 2014), Mandarin Chinese (Xiao, Rayson, & McEnery, 2009), and Russian (Sharoff, Umanskaya, & Wilson, 2013)". For an interesting

surveys different translations of the discourse marker “you know” into Czech, and finds that while a majority of discourse marking uses are translated with the Czech equivalent of “know”, there is also a great deal of variety that tracks the different types of discourse marking function that “you know” can play:

- (13) You know, it looked sorta funny.  
*Vypadalo to strašně srandovně, chápeš? (chápeš = you see)*
- (14) Still, you take what you can get, you know?  
*Člověk holt musí vzít zavděk tím, co je po ruce, že ano. (že ano = am I right?)*
- (15) “Guys like Barry, they have so much rage against women, you know.”  
*“Chlapi jako Barry v sobě dusí moře nenávisti k ženám, co?” (co? = is that right?)*
- (16) You know, this restaurant does have an indoor section.  
*Hele’, tahle restaurace má stolký i uvnitř. (heled = look)*

This aligns with the claim made in Chaume (2004: 843), that “In general, there is no one-to-one correspondence between two languages in the field of discourse markers”.

The fact that [know] appears to have different meanings in its objectual, propositional, and discourse marking forms means that it is a relatively straightforward exercise to exclude the non-propositional meanings of [know] when investigating whether the meaning associated with the propositional use of [know] corresponds with the concept of knowledge of interest to epistemologists. The prevalence of discourse marking uses of [know] in non-philosophical talk therefore doesn’t threaten ordinary language approaches in epistemology.

### 5.2 [Know] + prepositional phrase

The second interesting fact that jumps out of the sample of [know] is the existence of a reasonably common category of complements of [know] that, as far as we can tell, has never been discussed by philosophers, namely [know] + a prepositional phrase. The most common occurrences of [Know]+pp in the sample are the following:

1. known as 25 (48.1%).
2. know about 14 (26.9%).
3. known for 6 (11.5%).
4. known about 3 (5.7%).
5. known to 1 (1.9%).
6. knew of 1 (1.9%).
7. knows about 1 (1.9%).
8. knew about 1 (1.9%).

Some of these phrases are clearly not what Nagel calls “genuine references to knowing”. “Known as”, for example, is used as a way of talking about what something is called, as in (17):

- (17) His body was thrown into Pamlico Sound, his head given as a trophy to Spotswood, who had it displayed on a tall pole in Hampton Roads, at a site now **known as** Blackbeard’s Point.<sup>32</sup>

discussion of the relative frequency of [know] and [think] across languages, see Wierzbicka (2006: 34–41). Thanks to Mark Dingemane for bringing this discussion to our attention.

<sup>32</sup>Woodard (2014).

Similarly, “known for” is used not to refer to knowledge, but to indicate what properties are commonly associated with the subject (see (18)):

(18) Dr. Allan Armitage is well **known for** his books, articles and lectures, all of which are delivered in his engaging and accessible style.<sup>33</sup>

In contrast, “[know] about” refers to some body of propositional knowledge. In (19), the speaker is saying that we know some fact or facts about the negotiations (*we know that the negotiations exist*, for example).

(19) But I just want to say what we **know about** the negotiations, Chris, at this point is that serious issues or yes, the centrifuges, the quality of the uranium and so forth, but the sanctions are – I’m told by people watching this closely, what is really holding this up at this point.<sup>34</sup>

“Known to” and “knew of” also both seem to be “genuine references to knowing”. “Known to” can take either a verb phrase or a noun phrase as its object. When something is “known to” do something, there is some action that is generally known that the subject performs – so “known to” plausibly refers to a body of propositional knowledge concerning the subject and that action. When “known to” is followed by a noun phrase, it refers to objectual knowledge (see (20)). “Knew of” also appears to refer to objectual knowledge (see (21)).

(20) The technique known as spoofing exploited a vulnerability **known to** the U.S. military.

(21) In the 1920s the great majority of American Protestants **knew of** Buddhism only vaguely.

## 6. Evaluating the Bach and Baz frequency claims

Now that we have a better map of how [know] is ordinarily used, and types of questions the corpus enables us to answer, we can return to the frequency claims made by Bach (2005a) and Baz (2012) that we quoted at the beginning of the paper, and determine whether we have the resources to verify whether or not they are true.

### 6.1 Bach

Let’s start by examining the passage from Bach (2005a):

Also, it is worth keeping in mind that most of the time, outside of epistemology, when we consider whether somebody knows something, we are mainly interested in whether the person has the information, not in whether the person’s belief rises to the level of knowledge. Ordinarily we do not already assume that they have a true belief and just focus on whether or not their epistemic position suffices for knowing. Similarly, when we say that someone does not know something, typically we mean that they don’t have the information. (Bach 2005a: 62–63)<sup>35</sup>

<sup>33</sup>Shinn (2014).

<sup>34</sup>2015, Interview with John Brennan. Source: SPOK: Fox.

<sup>35</sup>For some agreement with Bach about ordinary positive knowledge assessments, see Sripada and Stanley (2012: 7): “In many contexts, all we care about is that an agent had a true belief, or that the content

One immediate difficulty in verifying Bach's claims using linguistic corpora is the fact that he doesn't initially limit the scope of his claims to uses of language – his initial topic is “when we *consider* whether someone knows something”, which would include both talk and thought. Obviously linguistic corpora will not give us the resources to evaluate non-linguistic instances of considering whether someone knows something. But in the final sentence of the passage, he makes a claim specifically about language: “When we *say* that someone does not know something, typically we mean that they don't have the information”, so we will concentrate on whether the corpus can help us verify that claim.

First, we need to figure out what it means to “have the information”. In the broader context of Bach's article, he is discussing propositional knowledge, and in the quoted passage he talks about the propositional attitude of *belief*, so it is reasonable to assume that he is concerned with the state of having information that a proposition is true (rather than, for example, having information that constitutes a relation of acquaintance to an object). In a personal communication, Bach said “by ‘having the information’ I just meant (truly) believing that *P*”.<sup>36</sup> So Bach's claim is that when we say that someone doesn't know that *p*, we typically mean that they don't have a true belief that *p*.

Second, now that we have a better idea of what Bach's claim is, we need to collect the relevant occurrences of cases of “say[ing] that someone does not know something” (where “something” refers to a proposition) before we can verify whether or not it's correct. We can do that by searching the corpus for “[do] not know that”, which finds occurrences of “did not know that”, “do not know that”, and “does not know that”. There are 795 occurrences of those phrases in COCA as a whole, which break down as follows:

1. did not know that (536)
2. do not know that (176)
3. does not know that (83).

We can now formulate a question precise enough that the corpus will enable us to answer it. Within this set of knowledge denials, are knowledge denials in which a subject has a true belief that *p*, but is said not to know that *p* more or less common than knowledge denials in which a subject simply lacks a true belief that *p*?

We examined a random sample of 100 occurrences of “[do] not know that”.<sup>37</sup> The vast majority of sampled knowledge denials (97/100) do not conflict with Bach's claim that non-philosophical denials of knowledge are denials that the subject has a true belief that *p*. As an illustration, consider the following example, drawn from the random sample:

(22) KING: Help our memory. How did they enlist you to serve? Countess of Romanones: Well, I had been trying to get into things like Jacqueline Cochran's flying group – training program – but I was 20 years old and you had to be 25. I happened to mention this on a blind date in New York; **did not know that the man who was my**

---

of the knowledge is true. In such contexts, we don't care about the justificatory or basing requirements of knowledge – we care only about its factivity, i.e. that knowledge entails truth. So we are led to grant that someone knows something even though they clearly do not satisfy anything beyond the truth or true belief requirements of knowledge.”

<sup>36</sup>Bach's notion of “having the information” is similar to the notion of a “reality congruent” informational state, which includes both true beliefs and knowledge. See Song and Baillargeon (2008: 1789), for example. This notion is discussed in Nagel (2017).

<sup>37</sup>The sample is available to download here: [https://semanticsarchive.net/Archive/jhiMTE0Z/Corpus\\_samples\\_know\\_hand\\_tagged.xlsx](https://semanticsarchive.net/Archive/jhiMTE0Z/Corpus_samples_know_hand_tagged.xlsx)



**blind date was the head of secret intelligence for Portugal and Spain.** And they found that I was perhaps the kind of a girl that could go in there and infiltrate [sic] into Spanish society. KING: You had to kill a man once, didn't you? Countess of Romanones: Yes, I did. I did. KING: Our guest is Aline, the Countess of Romanones; her third book, *The Spy Wore Silk*, about a thwarted attempt of the assassination of the ...<sup>38</sup>

The knowledge denial in this passage is *not* a case of the Countess of Romanones possessing a true belief that the man who was her blind date was the head of secret intelligence for Portugal and Spain, while lacking sufficient evidence for that belief to count as knowledge – there's no indication that the countess had any suspicion that her blind date was head of secret intelligence for Portugal and Spain.

Out of the random sample of 100 knowledge denials we examined, only *one* is a potential example of a subject being said not to know that *p* while she has a true belief that *p*. The remaining couple of examples fall into one of the following categories:

- anaphoric uses of “[do] not know that” that refer to earlier cases of know-wh:

(23) What this bill attempts to do – and it may not be perfect, and we know that – is to fix the broken borders, provide the interior enforcement, see that agriculture has a regular supply of labor, and provide a pathway to legalization so that the Homeland Security Department knows by photograph, by biometric identification who is in this country. **We do not know that now.**<sup>39</sup>

- uses of “[do] not know that” that are part of a reference to objectual knowledge:

(24) What could the cosmos need from or require of us? What sort of responsibilities flow from the idea that we all mirror the cosmos? What does it mean to say that the whole history of the cosmos is within a particular human being? I, for one, **do not know that whole history**, and I do not see how cosmic anchoring shows how my puny deeds will have cosmic significance.<sup>40</sup>

Why is it that the overwhelming majority of propositional knowledge denials are used to communicate that the target agent does not believe the proposition in question? (A referee observes that most epistemologists would probably find this fact “bizarre”.) That is, given that speakers could accurately describe such agents as not *believing* or *thinking* that *p*, why do speakers tend to describe such agents as not *knowing* that *p* instead? We think the explanation involves speakers simultaneously aiming to “maximize presupposition” (Heim 1991) on one hand, and aiming to avoid generating a false implicature on the other hand.<sup>41</sup> Consider the following two facts: (i) saying “S does not know that *p*” presupposes that *p*, so speakers can use such a sentence to add the proposition *p* to the conversational common ground; while (ii) a speaker who says “S does not believe/think that *p*” does not presuppose *p*, and tends to implicate that S believes/thinks *not-p*, as in (3).

(25) BUCKLEY: Well, this was – these were comments that came aboard Air Force One as the president was traveling up here. Ari Fleischer asked by reporters about it. Ari Fleischer saying that for the first time, that the president is saying that what Lott

<sup>38</sup>1991, Ed Bradley – The News Business. Source: SPOK: CNN\_King/.

<sup>39</sup>2007, Senate Votes for Cloture on Immigration Bill. Source: SPOK: PBS\_Newshour.

<sup>40</sup>Lawler (1997).

<sup>41</sup>Thanks to Emmanuel Chemla for discussion of this issue.

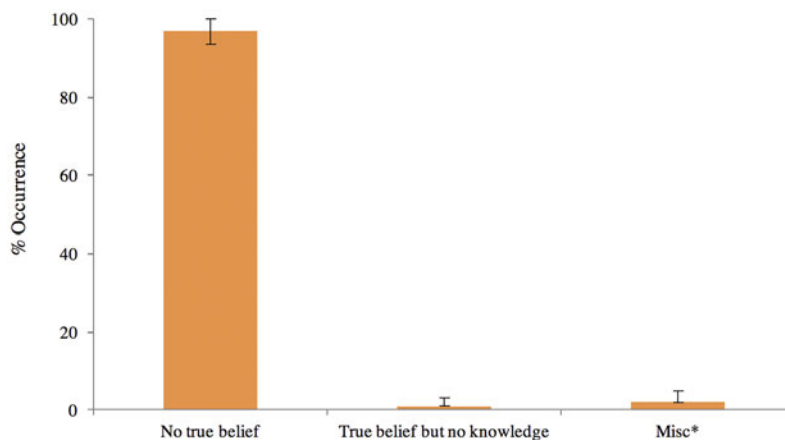


Fig. 3. Percentage occurrences of types of knowledge denials (“[do] not know that”), from a random sample of 100 occurrences in COCA. Error bars represent the 95% confidence interval of the proportion. \* “Miscellaneous” includes (i) anaphoric uses of “[do] not know that” that refer to earlier cases of know-wh; (ii) cases of “[do] not know that” that form part of denials of objectual knowledge.

said was wrong. It’s the first time that the administration is quoting the president as saying that what Trent Lott said was wrong. But they’ve also said that Lott has apologized, and the president **does not think that Lott should resign.**<sup>42</sup>

The use of the final sentence in (25) implicates that the president thinks that Lott should *not* resign. (This is an example of the phenomenon of “neg-raising” – see Horn (2001) for extensive discussion.)

Suppose one wants to communicate that *p* is the case and that *S* doesn’t have the belief that *p*. Saying “*S* doesn’t believe/think that *p*, but *p*” is not an effective way of doing that, because using such a sentence would typically implicate that *S* believes that not-*p*. In contrast, saying “*S* doesn’t know that *p*” is an efficient way of communicating that *p*, and it leaves open whether *S* doesn’t know that *p* because *S* doesn’t believe that *p* or because *S* believes that *p*, but lacks sufficient evidence to count as knowing that *p*. But it’s typically not hard to tell, in contexts like (22), which of those two more specific propositions the speaker intends to communicate by saying “*S* doesn’t know that *p*”. This combination of presupposition and implicature explains why speakers can use “*S* doesn’t know that *p*” to communicate that *S* doesn’t believe that *p*, but *p*.

The evidence from the corpus that we gathered therefore appears to support Bach’s claim about typical uses of knowledge denials outside of philosophy. But that appearance is misleading, for the following reason: Any piece of evidence that is compatible with Bach’s view that ordinary uses of “*S* does not know that *p*” are typically used to deny that *S* has a true belief that *p* is also compatible with the competing view that ordinary uses of “*S* does not know that *p*” are used to deny that *S* has a *justified* true belief that *p*. That’s because lacking a true belief that *p* is a way of lacking a justified true belief that *p*, so any evidence of the former will also be evidence of the latter. So what we have really found is evidence that *fails to disconfirm* both Bach’s view and the standard view that is his target. And after examining the corpus to try to find evidence in support of Bach’s frequency claim, it becomes clear that the only corpus-based evidence that would tip the balance between Bach’s claim and his opponent would be the typicality of

<sup>42</sup>2002, Iraq Terror Connection? Source: SPOK: CNN Iraq.

uses of “S does not know that p” that allow that the subject has a true belief that p, but denies knowledge on the grounds of lack of justification, which would show that Bach’s claim is mistaken. But we found that that type of use of “S doesn’t know that p” is rare.

To be clear, Bach’s claim is only about what is most typical, not that ordinary speakers wouldn’t be able to use knowledge denials in a way that acknowledges even a true belief might not count as knowledge. But the upshot of our failure to verify Bach’s frequency claim is that he is in no better an epistemic position than we are in respect to this claim – so it’s not something he is in a position to assert.

## 6.2 Baz

Now let’s consider Baz’s (2012) frequency claim about “I know that such and such”:

‘I know that such and such’ is far more commonly used in situations in which the obtaining of such and such is not in question and no one is in need of being assured of it. Think, for example, of the ‘I know’ of sharing a reaction to a piece of purported news, or the ‘I know’ of acknowledging a significant fact. (Baz 2012: 40)

The context in which Baz’s frequency claim appears is a discussion of Austin’s (1946) treatment of first-personal knowledge claims as performing a specific kind of speech act, namely the act of giving an *assurance* that something is the case. Austin compares the speech act of giving an assurance by making a knowledge claim with the speech act of *promising*. In both cases we do more than describe ourselves – we take “a new plunge” and give others a guarantee of something. On one recent account of assuring, the guarantee the speaker makes is that she has conclusive reasons in favor of the truth of the proposition assured (Lawlor 2013). Baz describes the conditions required for successfully performing an assurance as follows:

The use of ‘I know (that such and such)’ on which Austin focuses would be natural only in situations where the claim that such and such is grounded in some sort of expertise, as for example when the claimer is an expert in identifying birds, or perhaps in identifying some particular person’s moods. More generally, this use would be in place where the other is for some reason not in a position to assess one’s basis. (Baz 2012: 40)<sup>43</sup>

Baz gives the following two examples to illustrate “sharing a reaction to a piece of purported news” and “acknowledging a significant fact”, both of which are uses of “I know” that do not perform Austinian assurances (Baz 2012: 40 n. 38–39):

(26) ‘Jack and Jill are getting married!’ ‘I know!’ (with a tone of excitement, or, alternatively, with a sigh).

(27) ‘I know he is angry with me; I just haven’t had the time to speak with him about what happened’.

In order to verify Baz’s claim using COCA we need to look at a sample of propositional occurrences of “I know”, both occurrences that have explicit “that” clauses and

<sup>43</sup>Baz (2012: 40) claims that Austin, “taking his cue from the tradition’s obsession with knowledge as that which supposedly puts one in a position to give assurance ignores such situations [situations in which ‘I know’ is not used to make assurances]”. It is improbable that an acute observer of language (as Austin was) would be unaware that there are non-assurance uses of [know], since those uses are indeed quite common.

those that lack them. “I know” is a relatively frequent phrase in COCA, occurring 100,298 times. We collected a random sample of 100 propositional occurrences of “I know”, and hand coded them as either cases of *assurances* or *non-assurances*.<sup>44</sup> There is no algorithmic procedure for classifying the examples – it requires making judgments about the “total speech act in the total speech situation”.<sup>45</sup> Here are paradigmatic examples of assurances and non-assurances drawn from the sample:

### Assurances

(28) “I was having dinner with my own father a year after the war started in Iraq, and I said something that even raised a question about whether this was the right thing to do or not,” Morell recalls. “My father slammed his hand down on the table until the silverware jumped off, and he said, ‘But Iraq did 9/11!’ I said, ‘No, they didn’t, Dad. **I know**. Trust me. **I know**.”<sup>46</sup>

(29) Price is listed as the executive on a handful of inactive Louisiana businesses, including Street Life Entertainment, which Bryson Scott, who also is listed as an executive for the company, said was a record label. Scott identified himself as a rapper and said Price handled publicity. Scott also said he met Leonard Fournette during a barbecue at Price’s house when Fournette was playing football at St. Augustine High School in New Orleans. “**I know they were trying to run the BUGA Nation campaign**,” Scott said. Court records in Ascension Parish, where Price lives, show he has faced three civil lawsuits alleging unpaid credit card bills in recent years. Stumph said IWD Agency spoke to Lory Fournette only once, and Stumph said she met Leonard Fournette during a chance encounter in Baton Rouge. “He seemed very happy with the website and all the products,” Stumph said.<sup>47</sup>

### Non-assurances

(30) BEHAR: I see. Now, you’re going to – there’s a movie, I heard, that they might be making, a Lifetime movie about this story? C-SAVAGE: Yes. S-SAVAGE: Yes. BEHAR: Who’s playing you, Meryl Streep, I see her. C-SAVAGE: I’m going to play me and George Clooney is going to play Sean. BEHAR: Brilliant casting. C-SAVAGE: **I know**. I thought so.<sup>48</sup>

(31) ... the Republican National Committeeman there, a guy named Mickey Barnett, is calling for a hand recount. He said: “I just don’t believe 700 people went to the polls and did not vote for president.” That sounds a little familiar. From what you were just telling me about the 1 percent, 2 percent that don’t vote for president, perhaps you should call Mickey Barnett in New Mexico. I’m

<sup>44</sup>This sample is available to download here: [https://semanticsarchive.net/Archive/jhiMTE0Z/Corpus\\_samples\\_know\\_hand\\_tagged.xlsx](https://semanticsarchive.net/Archive/jhiMTE0Z/Corpus_samples_know_hand_tagged.xlsx).

<sup>45</sup>There is a possibility of bias given that the authors, rather than neutral coders, were coding the data. But in the cases where we approached the data with particular expectations, namely in the evaluation of the Bach and Baz claims, we found the opposite of our expectations, which should alleviate worries of the potential effects of bias somewhat. We have also provided links to the central samples to allow readers to evaluate our classifications.

<sup>46</sup>2015, ‘CIA vet: War far as I can see; U.S. has not done well in subduing the Islamic State.’ Susan Page, *USA Today*.

<sup>47</sup>2015, ‘Did Fournette’s family commit NCAA violation? BUGA Nation merchandise site raises questions.’ Josh Peter, *USA Today*.

<sup>48</sup>2011, ‘Did ‘Hanna Montana’ Destroy Cyrus Family? In Vitro Mix-up Nightmare; Interview With Julianne Moore.’ *Joy Behar Show*.

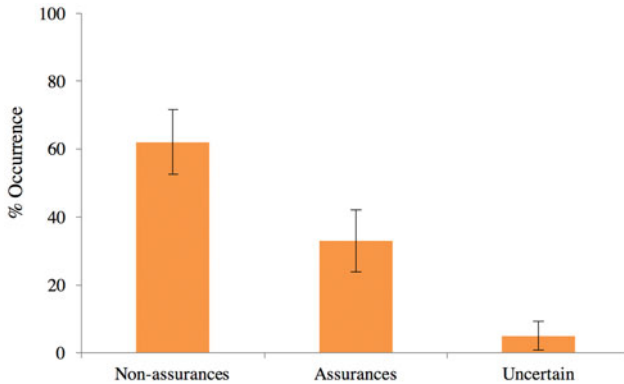


Fig. 4. Percentage of occurrences of “I [know]” that are assurances vs. non-assurances. From a random sample of 100 occurrences in COCA. Error bars represent the 95% confidence interval of the proportion.

not sure what the area code is there, but how do you reconcile those two completely contradictory points of view? **I know you don’t think it’s hypocrisy**, so please explain it to me. I’m really confused.<sup>49</sup>

All of the assurances shared with (28–29) the feature that the speaker had some expertise or special access to the fact being reported, in accordance with Baz’s characterization of Austin’s speech act of assurance. It also didn’t appear to be the case, in any of the examples classified as assurances, that the addressee of the speech act already believed the proposition being reported. The proportions of assurances, non-assurances, and cases that we were uncertain how to classify are as follows (see Figure 4):

- Non-assurances: 62/100.
- Assurances: 33/100.
- Uncertain: 5/100.

Baz’s claim that non-assurance uses of “I know” are “far more common” than assurance uses is backed up by this sample (assuming that having slightly less than twice as many in this sample is a way of satisfying “far more common”).

## 7. Conclusion: corpora, methodological hygiene, and ordinary language

We began this study with a suspicion, based on the unreliability of armchair frequency judgments, that Bach and Baz could be wrong about their characterizations of what the typical and more common uses of knowledge denials and “I know such and such” are. It turned out, however, that random samples drawn from COCA did not show that their armchair judgments were *false*. We found evidence that supports Baz’s claim that uses of “I know” are more frequently used to make non-assurances than assurances, and we failed to disconfirm Bach’s claim that ordinary uses of “I do not know that p” are simply denials that S has a true belief that p. Though we didn’t show that the content of either claim was false, we submit that neither Bach nor Baz had sufficient evidence to count as *knowing* that their frequency claims were

<sup>49</sup>2000, “Will Florida’s Courts Have the Final Say in the Presidential Election?” *CNN\_Crossfire*.

correct, and so neither was in a position (prior to this study) to assert the frequency claims that they made.<sup>50</sup> Moreover, since we also failed to disconfirm the hypothesis that Bach is arguing against – ordinary uses of the sentence “S does not know that p” are used to deny that S has a justified true belief that p – our evidence doesn’t put Bach in a position to know that his frequency claim is correct.

In addition to addressing general worries about the reliability and verifiability of armchair frequency judgments, looking at samples drawn from linguistic corpora can reveal new facts about the way linguistic expressions are used that have so far flown below philosophers’ radar. For example, while philosophers have devoted a great deal of attention to knowledge-that, knowledge-wh, and objectual knowledge, so far they have not discussed cases of knowledge + prepositional phrase complements. A cursory look at occurrences of [know] in COCA reveals that such occurrences are roughly as frequent as occurrences of objectual knowledge. And most striking of all, philosophers have so far not discussed discourse marking uses of [know], even though such uses are the most frequent occurrences of [know] in COCA.

When J.L. Austin begins his ordinary language investigation of the meaning of the expressions “intentionally” and “deliberately”, he says that the first step is to “consider some cases” of how those expressions are used. He mentions the possibility of looking at actual cases in which those expressions are used, before dismissing that option in favor of imagined cases:

First let us consider some cases. Actual cases would of course be excellent: we might observe what words have actually been used by commentators on real incidents, or by narrators of fictitious incidents. However, we do not have the time or space to do that here. We must instead imagine some cases. (Austin 1966: 429)


Austin was aware of, through contact with Arne Naess and other mid-century practitioners of “empirical semantics” and “occurrence analysis” (a forerunner of corpus linguistics) that extensive examination of actual cases was possible (Chapman 2011, 2014; Murphy 2015; Hansen 2017). But Austin’s consideration of actual cases is limited to his discussion of the case of *Regina vs. Finney*, in which the defendant, Finney, is convicted of manslaughter for scalding one of the mental patients under his care to death in a bath (Austin 1956–1957).<sup>51</sup> And there is reason to think that Austin was actually suspicious of the experimental and corpus-based methods of the empirical semantics movement (Chapman 2014; Murphy 2015; Longworth 2018). In contrast to Austin’s stance, we have aimed to show how a convincing and empirically grounded ordinary language approach to the investigation of lexical meaning should not ignore “what words have actually been used by commentators on real incidents” as a supplement to, and check on, philosophers’ judgments of what sorts of uses of language are the most common, ordinary, or typical.<sup>52</sup>

<sup>50</sup>For a similar worry about justification for armchair judgments about knowledge, see Hansen (2012, 2013), and a subsequent study (Hansen and Chemla 2013) that indicates that such armchair judgments are in fact in line with judgments collected in a formal experiment.

<sup>51</sup>There is a criticism of Austin’s methods, written around the same time as “Three Ways of Spilling Ink” was published (in which Austin’s remark about not having the time or space to consider “actual cases” appears), that argues Austin’s “intuitive” method of considering imaginary cases needs to be supplemented with a survey of “texts of actual language (spoken and written)” (New 1966: 374).

<sup>52</sup>Thanks to Zed Adams, Avner Baz, Emmanuel Chemla, Alex Davies, Mark Dingemans, Dan Harris, Shen-yi Liao, Eliot Michaelson, David Plunkett, Chris Potts, Sebastian Schuster, members of the Linguistics and Philosophy workshop at the University of Chicago, the Zürich Doctoral Workshop on ordinary language philosophy, and the symposium on metapragmatics at the 2019 Central APA for very helpful

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/epi.2019.15>

**Author ORCID.**  Nat Hansen, 0000-0001-5074-1075

## References

- Andow J.** (2015). 'How 'Intuition' Exploded.' *Metaphilosophy* **46**(2), 189–212.
- Austin J.** (1946). 'Other Minds.' *Aristotelian Society Supplementary Volume* **20**, 148–87.
- Austin J.** (1956–1957). 'A Plea for Excuses.' *Proceedings of the Aristotelian Society* **57**, 1–30.
- Austin J.** (1966). 'Three Ways of Spilling Ink.' *Philosophical Review* **75**(4), 427–40.
- Bach K.** (2005a). 'The Emperor's New 'Knows'.' In G. Preyer and G. Peter (eds), *Contextualism in Philosophy: Knowledge, Meaning and Truth*, pp. 51–89. Oxford: Oxford University Press.
- Bach K.** (2005b). 'Questions and Answers.' Comments on Jonathan Schaffer's "Knowing the Answer", Bellingham Summer Philosophy Conference.
- Barnes W.H.** (1953–1954). 'Talking about Sensations.' *Proceedings of the Aristotelian Society* **54**, 261–78.
- Baz A.** (2012). *When Words Are Called For: A Defense of Ordinary Language Philosophy*. Cambridge, MA: Harvard University Press.
- Bluhm R.** (2016). 'Corpus Analysis in Philosophy.' In M. Hinton (ed.), *Evidence, Experiment and Argument in Linguistics and the Philosophy of Language*, pp. 91–109. Berlin: Peter Lang.
- Braun D.** (2006). 'Now You Know Who Hong Oak Yun Is.' *Philosophical Issues* **16**, 24–42.
- Brogaard B.** (2009). 'What Mary Did Yesterday: Reflections on Knowledge-wh.' *Philosophy and Phenomenological Research* **78**(2), 439–67.
- Brownstein M. and Michaelson E.** (2016). 'Doing without Believing: Intellectualism, Knowledge-how, and Belief Attribution.' *Synthese* **193**(9), 2815–36.
- Buckwalter T. and Parkinson D.** (2014). *A Frequency Dictionary of Arabic: Core Vocabulary for Learners*. New York, NY: Routledge.
- Camp E. and Hawthorne J.** (2008). 'Sarcastic 'Like': A Case Study in the Interface of Syntax and Semantics.' *Philosophical Perspectives* **22**(1), 1–21.
- Chapman S.** (2011). 'Arne Naess and Empirical Semantics.' *Inquiry* **54**(1), 18–30.
- Chapman S.** (2014). 'Getting the Philosopher out of the Armchair: J.L. Austin's Response to Logical Positivism in Comparison to that of Arne Naess.' In B. Garvey (ed.), *J.L. Austin on Language*, pp. 108–23. Basingstoke: Palgrave Macmillan.
- Chaume F.** (2004). 'Discourse Markers in Audiovisual Translating.' *Meta* **49**(4), 843–55.
- Cortes I.** (2015). 'School Morning Madness.' *USA Today Magazine* **144**(2844), 68–9.
- Cumming G. and Maillardet R.** (2006). 'Confidence Intervals and Replication: Where Will the Next Mean Fall?' *Psychological Methods* **11**(3), 217–27.
- Daniulaityte R., Carlson R.G. and Kenne D.R.** (2006). 'Initiation to Pharmaceutical Opioids and Patterns of Misuse: Preliminary Qualitative Findings Obtained by the Ohio Substance Abuse Monitoring Network.' *Journal of Drug Issues* **36**(4), 787–808.
- Davies M.** (2002–). *Corpus del Español: 100 Million Words, 1200s–1900s*. <http://www.corpusdelespanol.org>.
- Davies M.** (2008–). *The Corpus of Contemporary American English (COCA): 520 Million Words, 1990–present*. <http://corpus.byu.edu/coca/>.
- Delinsky B.** (2015). *Blueprints*. New York, NY: St. Martin's Press.
- Engelen B.** (2010). *Schwierige Sprachliche Strukturen: Aufsätze zur Deutschen Grammatik*. Frankfurt am Main: Peter Lang.
- Erman B.** (2001). 'Pragmatic Markers Revisited with a Focus on *You Know* in Adult and Adolescent Talk.' *Journal of Pragmatics* **33**(9), 1337–59.
- Fischer E., Engelhardt P.E. and Herbelot A.** (2015). 'Intuitions and Illusions: From Explanation and Experiment to Assessment.' In E. Fischer and J. Collins (eds), *Experimental Philosophy, Rationalism and Naturalism*, pp. 259–92. London: Routledge.
- Fraña I.** (2017). *Concealed Questions*. Oxford: Oxford University Press.

comments. Kent Bach very generously discussed the content of the paper and his own view about "know" with us in detail. Anonymous referees provided helpful and constructive comments. This research was supported by Leverhulme Research Project Grant RPG-2016-193 and by an External Faculty Fellowship and a Geballe Fellowship at the Stanford Humanities Center.

- Fridland E.** (2015). 'Knowing-how: Problems and Considerations.' *European Journal of Philosophy* **23**(3), 703–27.
- Ginet C.** (2008). 'In Defense of a Non-causal Account of Reasons Explanations.' *Journal of Ethics* **12**(3/4), 229–37.
- Glick E.** (2013). 'Practical Modes of Presentation.' *Noûs* **49**(3), 538–59.
- Goddard C. and Wierzbicka A.** (1994). *Semantic and Lexical Universals: Theory and Empirical Findings*. Amsterdam: John Benjamins.
- Hansen N.** (2012). 'On an Alleged Truth/Falsity Asymmetry in Context Shifting Experiments.' *Philosophical Quarterly* **62**(248), 530–45.
- Hansen N.** (2013). 'A Slugfest of Intuitions: Contextualism and Experimental Design.' *Synthese* **190**(10), 1771–92.
- Hansen N.** (2017). 'Must we Measure what we Mean?' *Inquiry* **60**(8), 785–815.
- Hansen N. and Chemla E.** (2013). 'Experimenting on Contextualism.' *Mind & Language* **28**(3), 286–321.
- Hansson S.O.** (2006). 'Category-Specified Value Statements.' *Synthese* **148**(2), 425–32.
- Hasan A. and Fumerton R.** (2014). 'Knowledge by Acquaintance and Knowledge by Description.' In E.N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/spr2014/entries/knowledge-acquaintandescr/>.
- Haspelmath M. and Tadmor U.** (eds) (2009). *The World Loanword Database (WOLD)*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Hazlett A.** (2010). 'The Myth of Factive Verbs.' *Philosophy and Phenomenological Research* **80**(3), 497–522.
- Heim I.** (1979). *Concealed Questions*, pp. 51–60. Berlin: Springer-Verlag.
- Heim I.** (1991). 'Artikel und Definitheit [Articles and Definiteness].' In A. von Stechow and D. Wunderlich (eds), *Semantic: Ein internationales Handbuch der zeitgenössischen Forschung*, pp. 487–535. Berlin: de Gruyter.
- Higginbotham J.** (1996). 'The Semantics of Questions.' In S. Lappin (ed.), *The Handbook of Contemporary Semantic Theory*, pp. 361–83. Oxford: Blackwell.
- Horn L.R.** (2001). *A Natural History of Negation*. Reissue edition. Stanford, CA: CSLI Publications.
- Houser N.** (2015). 'History's Best Places to Kiss.' *Fantasy & Science Fiction* **128**(1/2), 129–49.
- Ichikawa J.J. and Steup M.** (2017). 'The Analysis of Knowledge.' In E.N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/spr2017/entries/knowledge-analysis/>.
- Jary M. and Stainton R.J.** (2017). 'Contextualism in Epistemology and Relevance Theory.' In J.J. Ichikawa (ed.), *The Routledge Handbook of Epistemic Contextualism*, pp. 480–92. New York, NY: Routledge.
- Kulynych J.** (2001). 'No Playing in the Public Sphere: Democratic Theory and the Exclusion of Children.' *Social Theory and Practice* **27**(2), 231–64.
- Lawler P.A.** (1997). 'Havel's Postmodern View of Man in the Cosmos.' *Perspectives on Political Science* **26**(1), 27.
- Lawlor K.** (2013). *Assurance: An Austinian View of Knowledge and Knowledge Claims*. Oxford: Oxford University Press.
- Liao S.-Y., McNally L. and Meskin A.** (2016). 'Aesthetic Adjectives Lack Uniform Behavior.' *Inquiry* **59**(6), 618–31.
- Longworth G.** (2018). 'The Ordinary and the Experimental: Cook Wilson and Austin on Method in Philosophy.' *British Journal for the History of Philosophy*, Online First. doi: 10.1080/09608788.2017.1413539.
- Ludlow P.** (2005). 'Contextualism and the New Linguistic Turn in Epistemology.' In G. Preyer and G. Peter (eds), *Contextualism in Philosophy: Knowledge, Meaning and Truth*, pp. 11–50. Oxford: Oxford University Press.
- Lupoff R.A.** (2015). '12:03 P.M.' *Fantasy & Science Fiction* **123**(3/4), 90–113.
- Magnigová J.** (2016). Subtitles and Their Czech Translation Counterparts. Master's thesis, Palacký University Olomouc, Olomouc, Czech Republic.
- Manning C.D., Surdeanu M., Bauer J., Finkel J., Bethard S.J. and McCloskey D.** (2014). 'The Stanford CoreNLP Natural Language Processing Toolkit.' In K. Bontcheva and J. Zhu (eds), *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: Systems Demonstrations*. Baltimore, MD: Association for Computational Linguistics, pp. 55–60.
- Mason L.** (2015). 'Teardrop.' *Fantasy & Science Fiction* **128**(5/6), 7–29.
- Meija-Ramos J.P., Alcock L., Lew K., Rago P., Sangwin C. and Inglis M.** (2019). 'Using Corpus Linguistics to Investigate Mathematical Explanation.' In E. Fischer and M. Curtis (eds), *Methodological Advances in Experimental Philosophy*. London: Bloomsbury.
- Mitias M.H.** (1982). 'What Makes an Experience Aesthetic?' *Journal of Aesthetics and Art Criticism* **41**(2), 157–69.



- Moretti F.** (2013). ‘Operationalizing’: Or, the Function of Measurement in Modern Literary Theory.’ *Stanford Literary Lab Pamphlet* 6, 1–13.
- Murphy T.S.** (2015). ‘Experimental Philosophy: 1935–1965.’ In T. Lombrozo, J. Knobe and S. Nichols (eds), *Oxford Studies in Experimental Philosophy*, vol. 1, pp. 325–68. Oxford: Oxford University Press.
- Nagel J.** (2014). *Knowledge: A Very Short Introduction*. Oxford: Oxford University Press.
- Nagel J.** (2017). ‘Factive and Non-factive Mental State Attribution.’ *Mind & Language* 32(5), 525–44.
- New C.** (1966). ‘A Plea for Linguistics.’ *Mind* 75(299), 368–84.
- Newton N.** (1989). ‘On Viewing Pain as a Secondary Quality.’ *Noûs* 23(5), 569–98.
- Oderberg D.S.** (1989). ‘Johnston on Human Beings.’ *Journal of Philosophy* 86(3), 137–41.
- Östman J.-O.** (1981). *You Know: A Discourse Functional Approach*. Amsterdam: John Benjamins.
- Pinillos A.** (2012). ‘Knowledge, Experiments, and Practical Interests.’ In J. Brown and M. Gerken (eds), *New Essays on Knowledge Ascriptions*, pp. 192–219. Oxford: Oxford University Press.
- Pinillos A. and Nichols S.** (2018). ‘Skepticism and the Acquisition of Knowledge.’ *Mind & Language* 33(4), 397–414.
- Predelli S.** (2013). *Meaning without Truth*. Oxford: Oxford University Press.
- Price H.** (1965). ‘Belief ‘in’ and belief ‘that.’’ *Religious Studies* 1(1), 5–27.
- Roque L.S., Kendrick K.H., Norcliffe E., Brown P., Defina R., Dingemans M., Dirksmeyer T., Enfield N., Floyd S., Hammond J., Rossi G., Tufvesson S., van Putten S. and Majid A.** (2015). ‘Vision Verbs Dominate in Conversation Across Cultures, but the Ranking of Non-Visual Verbs Varies.’ *Cognitive Linguistics* 26(1), 31–60.
- Rumfitt I.** (2003). ‘Savoir Faire.’ *Journal of Philosophy* 100(3), 158–66.
- Saebø K.J.** (2004). ‘Natural Language Corpus Semantics: The Free Choice Controversy.’ *Nordic Journal of Linguistics* 27(2), 197–218.
- Saunders J.T.** (1962). ‘Fatalism and Linguistic Reform.’ *Analysis* 23(2), 30–1.
- Schaffer J.** (2007). ‘Knowing the Answer.’ *Philosophy and Phenomenological Research* 75(2), 383–403.
- Schiffrin D.** (1987). *Discourse Markers*. Cambridge: Cambridge University Press.
- Schiffrin D.** (2001). ‘Discourse Markers: Language, Meaning, and Context.’ In D. Schiffrin, D. Tannen and H.E. Hamilton (eds), *The Handbook of Discourse Analysis*, pp. 54–75. Oxford: Blackwell.
- Schourup L.C.** (1985). *Common Discourse Particles in English Conversation*. New York, NY: Garland Publishing, Inc.
- Schourup L.** (1999). ‘Discourse Markers.’ *Lingua* 107(3–4), 227–65.
- Sharoff S., Umanskaya E. and Wilson J.** (2013). *A Frequency Dictionary of Russian: Core Vocabulary for Learners*. New York, NY: Routledge.
- Shinn M.** (2014). ‘The Digital Tool Shed.’ *Horticulture* 111(4), 58–61.
- Song J.H. and Baillargeon R.** (2008). ‘Infants’ Reasoning About Others’ False Perceptions.’ *Developmental Psychology* 44(6), 1789–95.
- Sripada C.S. and Stanley J.** (2012). ‘Empirical Tests of Interest-Relative Invariantism.’ *Episteme* 9(1), 3–26.
- Stanley J.** (2011). *Know How*. Oxford: Oxford University Press.
- Stanley J. and Williamson T.** (2001). ‘Knowing How.’ *Journal of Philosophy* 97(8), 411–44.
- Stone J.** (2001). ‘What is it Like to Have an Unconscious Mental State?’ *Philosophical Studies* 104(2), 179–202.
- Sytsma J., Bluhm R., Willemsen P. and Reuter K.** (2019). ‘Causation Attributions and Corpus Analysis.’ In E. Fischer and M. Curtis (eds), *Methodological Advances in Experimental Philosophy*. London: Bloomsbury.
- Taylor P.W.** (1954). ‘Four Types of Ethical Relativism.’ *Philosophical Review* 63(4), 500–16.
- Tversky A. and Kahneman D.** (1973). ‘Availability: A Heuristic for Judging Frequency and Probability.’ *Cognitive Psychology* 5(2), 207–32.
- Vesey G.** (1954). ‘Unthinking Assumptions and their Justification.’ *Mind* 63(250), 226–33.
- Vetter B.** (2014). ‘Dispositions Without Conditionals.’ *Mind* 123(489), 129–56.
- Vihvelin K.** (1996). ‘What Time Travelers Cannot Do.’ *Philosophical Studies* 81(2/3), 315–30.
- Wierzbicka A.** (1996). *Semantics: Primes and Universals*. Oxford: Oxford University Press.
- Wierzbicka A.** (2006). *English: Meaning and Culture*. Oxford: Oxford University Press.
- Wolgast E.H.** (1960). ‘The Experience in Perception.’ *Philosophical Review* 69(2), 165–82.
- Woodard C.** (2014). ‘The Last Days of Blackbeard.’ *Smithsonian* 44(10), 32–41.
- Xiao R., Rayson P. and McEnery T.** (2009). *A Frequency Dictionary of Mandarin Chinese: Core Vocabulary for Learners*. New York, NY: Routledge.

## Appendix: Examples of philosophers making armchair frequency claims about language (drawn from JSTOR)

### “most commonly we use”/“most commonly used”

I think myself that we very rarely, if ever, use *observe* in ordinary language in connection with the senses except in connection with the sense of sight, to mean either *see* or *look at*. Also I think that *perceive* is **most commonly used** as a rather formal synonym for *notice*. (Barnes 1953–1954: 263)

The same kind of analysis can be applied to the rather difficult example of ‘believing in oneself’. **Most commonly we use** this phrase in a negative form. The trouble with Tom is that he does not believe in himself: he will never make a success in his life. (Price 1965: 19)

### “we typically mean”/“typically we mean”

We do find it natural to utter such a sentence as ‘Sarah’s reason for deciding to enter the room was that she wanted to get her glasses’, but **typically we mean** by this just that Sarah decided to enter the room in order to get her glasses, i.e., that what she decided was: to enter the room in order to get her glasses, i.e., that her reason for entering the room was part of the content of the intention formed by her decision. What **we typically mean** by ‘the reason for which S decided to do A’ is the reason for which, according to the intention formed in the decision, S will do A when and if she carries out the decision. (Ginet 2008: 233)

When we say that a decision is good, **we typically mean** that it is good *qua* decision. (Hansson 2006: 426)

Consider what **we typically mean** when we say someone is acting in a childish way or when oppressed groups claim they are being treated like children. We are asserting that childishness is the opposite of the desired state of adult responsibility. (Kulynych 2001: 257)

What **we typically mean** by ‘self-consciousness’ or ‘self-awareness’ is more complex: it involves an awareness of oneself as a subject of experience, with unified mental states, who persists through various changes. (Stone 2001: 198 n. 11)

### “we most commonly mean”

One thing making this example difficult, however, is what **we most commonly mean** when we speak of wondering if there is a pain: when we wonder this we *do try to detect or become aware of* a pain by means of a sensation, or else, as in the first example, to analyze the sensations we are already aware of having. (Newton 1989: 586)

### “we ordinarily mean”/“ordinarily we mean”

Now, when we ask, ‘What makes a pumpkin yellow?’ do we really have in mind a causal kind of question? Are we in other words, asking: what causes the pumpkin to be yellow? What can the answer to this question be? **Ordinarily, we mean**, I should say: what *makes* the pumpkin yellow is simply the fact that it possesses the property, color, yellow. (Mittias 1982: 159)

He certainly cannot mean what **we ordinarily mean** by ‘human being’, since what **we ordinarily mean**, if indeed there is any definite meaning at all, is ‘human organism’, ‘member of the species *homo sapiens*’. (Oderberg 1989: 138)

In saying that it is within one’s power to bring about a situation **we ordinarily mean** that he has the requisite skills and resources, that no one has bound him hand and foot to prevent his so doing, and so on. (Saunders 1962: 30)

We begin by seeing whether there is a *rational* procedure for settling moral conflicts, or whether *good reasons* can be given for justifying moral beliefs, or whether there is a *reasonable* way of arriving at a moral decision. What does this mean? It means what **we ordinarily mean** by using the terms ‘rational’, ‘good reasons’, and ‘reasonable’. (Taylor 1954: 516)

If the justification the philosopher looks for is not what **we ordinarily mean** by ‘justification’ then the products of his search cannot have the sort of significance they would have if it were. (Vesey 1954: 226)

I think they both mean by ‘can’ what **we ordinarily mean**. And what **we ordinarily mean** when we say that someone can do something is that she has both the ability and the opportunity to do it. (Vihvelin 1996: 318)

This experience is not the *whole* of what we mean when we say we see something, because **ordinarily we mean** also to imply that that thing is before us. (Wolgast 1960: 166)

**Nat Hansen** is Associate Professor of Philosophy and a member of the Centre for Cognition Research at the University of Reading. He has written papers on contextualism, experimental semantics and pragmatics, the meaning of color terms, and ordinary language philosophy.

**J.D. Porter** is the Associate Director of the Literary Lab at Stanford University. Working primarily in the broad field of digital humanities, he specializes in race and ethnicity theory, literary modernism, and poetics. He has work forthcoming in the journal *Cultural Analytics* and through Cambridge University Press (in the collection *Ralph Ellison in Context*).

**Kathryn Francis** is a Lecturer in Psychology at the University of Bradford, UK. Previously, she worked as a Postdoctoral Research Fellow at the University of Reading in the Department of Philosophy and the School of Psychology and Clinical Language Sciences. Kathryn’s research interests lie at the intersection of experimental philosophy and psychology (epistemology, moral psychology) and she has co-authored papers in the sciences (*British Journal of Psychology*, *Scientific Reports*, *PLoS ONE*) and in philosophy (*Ergo*).