Where Do Cognitive Biases Fit Into Cognitive Linguistics? An Example from the ‘Curse of Knowledge’

VERA TOBIN

1 Introduction

It is a truth widely acknowledged (e.g. Piaget 1932; Fillmore 1975; Baron-Cohen 1995) that human beings have a remarkable ability to think about and adopt the perspectives of others. However, despite these spectacular performances of perspective taking, our abilities in this arena are both limited and riddled with biases. These biases are studied extensively in cognitive psychology, social psychology, economics, and cognitive approaches to decision making. Cognitive linguistics, however, despite its longstanding interest in viewpoint and perspective, has given them much less direct or explicit attention. This paper suggests that some models of meaning construction in cognitive linguistics are in fact very well suited to addressing the contributions of cognitive bias, and presents one illustrative integrated account.

The ‘curse of knowledge’ is a pervasive cognitive bias that makes it very difficult for us to imagine, once we know something, what it is like not to know it (Camerer, Loewenstein and Weber 1989; Birch and Bloom 2003). Here, I will argue that the curse of knowledge is an artifact of a more general cognitive shortcut that is implicated in features of ‘correct’ sentence interpretation such as presupposition projection, as well as in the phenomena that are traditionally described as curse-of-knowledge errors.

Linguistic forms that mark aspects of viewpoint may express aspects of a language user’s own immediate perceptions and conceptions, but can of
course also be deployed in service of much more complex representations. We can speak and think of other places and times; we can produce negation and counterfactual conditionals; we can represent the speech and thoughts of others (including imagined others); we can embed perspectives within other perspectives, layer them, or blend them in a host of fleeting or extended modes of discourse presentation.

Many models of meaning construction within cognitive linguistics (and elsewhere) describe that process in terms of mental representations. None of these models, to my knowledge, claims to model cognitive biases like the curse of knowledge, but mental spaces theory (Fauconnier 1985, 1997; Cutzer 1994) does have a theoretical apparatus that accounts for ways that structure tends to flow from one mental representation to another. This apparatus offers an elegant solution to a number of classic problems in semantics and pragmatics, none of which are normally understood to have anything to do with the egocentric biases implicated in drawing inferences or with our shortcomings in reasoning about other minds. But, it turns out, there are good reasons to think that at least some of them do, and that in particular the mental spaces account of presupposition captures something important about how that phenomenon reflects those biases.

2 Biases In and Of Perspective-taking

It is often the case that one person in an interaction knows more about some relevant subject than another person does. This state of affairs is known in strategic intelligence and economics as an ‘information asymmetry’. Asymmetric information can be of significant tactical value. Confidence artists, hustlers, and investment bankers all make their livings by cashing in on such advantages.

The career of the gambler Alvin Thomas, better known as ‘Titanic’ Thompson, provides a parade of illustrations of the advantages a carefully cultivated asymmetry can confer (compiled in loving detail in Cook 2011): A professional-caliber golf player and the model for the character Sky Masterson in *Guys and Dolls*, Titanic Thompson’s signature hustle was to beat a man playing right handed, then offer a double-or-nothing match in which he would play left handed—capitalizing on the fact that he was, unbeknownst to his opponent, actually a lefty. His greatest joy was rigging proposition wagers, bets on apparently random events about which he secretly had information his opponent did not: how many watermelons were stacked on a farm truck passing by (he had paid the farmer for a count that morning); whether he could hit a 500-yard golf drive (he could, when it was on the frozen Lake Michigan); whether he could cut a flower off its stem just by throwing playing cards (no trick, just practice).
More mundanely, workers generally know more about their own ability and motivation than prospective employers do. Most sellers are better informed about the true value of their products than buyers are, though a knowledgeable collector of Northern Soul records may well be better informed about the true value of a copy of Frank Wilson’s ‘Do I Love You’ than the seller who found it in a box in the basement—and if so, that buyer can make a tidy profit by capitalizing on this information gap. The way things seem to work, in other words, and our usual assumption, is that knowing more gives you the upper hand. But it turns out that this is not quite right.

The economists Colin Camerer, George Loewenstein, and Martin Weber (1989) tested this premise by putting people into a number of situations where they should, theoretically, have been particularly well positioned to predict and take advantage of the judgments of their less-informed counterparts: they had a financial incentive to do so (which ought to provide motivation to think carefully) and they received feedback about the accuracy of their predictions (which should help them to correct for errors in their assessments). Yet, despite these marks in their favor, the better-informed participants consistently let their private information affect their behavior even when it was to their advantage not to do so.

Camerer, Loewenstein, and Weber coined the term ‘the curse of knowledge’ to describe this tendency. More recently, some cognitive and developmental psychologists (e.g. Birch and Bloom 2003; Keysar, Lin, and Barr 2003; Birch 2005) have suggested that young children’s difficulty with false belief tasks, usually taken as evidence that they lack a ‘theory of mind’, are of a piece with curse-of-knowledge effects in adults. It may be that all of these difficulties reflect a single fundamental bias in social cognition that is more acute in younger children, who have less inhibitory control (Leslie and Polizzi 1998), but persists in some form for our entire lives.

In the realm of language, the curse of knowledge seems also to affect people’s judgments of the decodability of their own utterances, and especially of the degree to which meaning is ‘obviously’ conveyed by linguistic and paralinguistic structure. For instance, utterances are often potentially ambiguous, but speakers have many tools to constrain that ambiguity for their interlocutors’ benefit. However, while we do deploy those tools and believe that we are making ourselves clear by doing so, several studies (e.g. Newton 1994; Keysar and Bly 1995; Kelley and Jacoby 1996; Kraljic and Brennan 2005) suggest that our judgments in those moments are ‘cursed’. We are systematically biased to think that we have been understood. Even speakers who have been warned that what they are about to say is ambigu-
ous underestimate how ambiguous their own utterances will be and greatly overestimate the helpfulness of their attempts at disambiguation.

For instance, Keysar and Henley (2002) gave speakers syntactically ambiguous sentences and asked them to say them in such a way that another participant would understand them as unambiguous: e.g. ‘Angela killed the man with the gun’ said so as to convey the notion that Angela used the gun to kill the man, rather than that the man had the gun. Then the speakers were asked how they believed the listener would understand the sentence, and how confident they were in that assessment. The majority of speakers were both confident and wrong. Similarly, as illustrated in Figure 1, overhearsers who were given privileged information about another speaker’s intentions overestimated how transparent those intentions would be to a third party (Keysar 2000).

![Figure 1. Over-projection and assessing the transparency of intentions](image)

Results of this sort suggest that speakers don’t just rely overmuch on their own mental states to guide their performance, though they certainly do that; they also actively over-project information from their own perspective to their assessment of what is in the common ground. This brings us to the question of how best to incorporate these patterns of over-projection into our accounts of linguistic viewpoint phenomena that are frequently discussed in terms of what speakers and hearers are assumed to know, believe, bring readily to mind, or be able to infer about the matter under discussion. The classic example is the case of presupposition.
3 Presupposition

3.1 Overview

Presupposition has historically enjoyed considerable attention from linguists and philosophers of language, as a topic of interest to larger debates about the interaction and division of labor between semantics and pragmatics. The modern philosophical study of presupposition began with Gottlob Frege’s (1892) observations about some curious referential properties of proper names and temporal clauses. From there it served as a testing ground for truth-conditional accounts of the propositional status of definite descriptions (Russell 1905) and the relationships between sentences, statements, and speakers (e.g. Strawson 1952), and onward into a variety of theories that engage in one way or another with issues of contextual appropriateness. These latter approaches embrace positions from Karttunen’s (1973: 6) suggestion that presuppositions are a matter of ‘some … conversational principle’ regarding who ‘can be assumed to share the speaker’s beliefs’, to van der Sandt’s (1992) theory that presuppositions behave much like anaphora that need to be resolved with respect to their linguistic context, and Beaver’s (1999: 3) claim that presuppositional phenomena are ‘not naturally explained in terms of truth conditions but in terms of plausibility’.

Anyone sitting down to consider the phenomenon today can thus partake of well over a hundred years of prior theorizing and digging up of troublesome examples. Some of these investigations are of greater relevance to current trends in linguistic inquiry than others. As Levinson (1983: 167) remarked, ‘there is more literature on presupposition than on almost any other topic in pragmatics … and while much of this is of a technical and complex kind, a great deal is also obsolete’. In the thirty years since that writing, that body of scholarship has only grown larger, and its foundational works of course grow older every year.

Levinson was correct in his observation that ‘Much that [once] seemed confusing and mysterious has become clearer now that some basic distinctions and frameworks have been established’ (1983: 167–168). Nonetheless, there seems to be some room left at the table for new accounts of presupposition, particularly within the frameworks of dynamic semantics (e.g. Beaver 2001) and probabilistic processing models (e.g. Lassiter 2012)—and similarly, I would suggest, as an entryway for integrating research on reasoning under uncertainty with cognitive-linguistic models of meaning construction in discourse.

Broadly speaking, a presupposition is an inference or proposition whose truth is taken for granted in the utterance of a sentence. Presuppositions are usually tied to the use of particular lexical items or linguistic constructions,
called presupposition triggers. So, for instance, consider the word *regret* in Example (1), as discussed in Fillmore (1985: 249):

(1) John regretted signing the letter.

Under the standard interpretation of this example, the sentence *entails* that John felt regret, but *presupposes* that he signed the letter, and this is because, as a factive verb, *regret* presupposes its complement. Presuppositions typically survive negation, as seen in (2) and (3), both of which presuppose that Henrietta was late:

(2) Henrietta realized she was late.
(3) Henrietta didn’t realize she was late.

… unless the negation is taken to reject the entire framing of the presupposition trigger, in what is sometimes called ‘sentential’ or ‘metalinguistic’ negation.

(4) John didn’t *regret* signing the letter, because he didn’t sign it in the first place!
(5) I didn’t stop kicking my cat, because I never kicked her at all!
(6) The king of France isn’t bald, because there is no king of France!

The circumstances in which presuppositions survive within complex sentences, as they do in (3) and in a number of other contexts where entailments do not, and those in which they evaporate, as they do in (4–6), have been the source of much puzzlement for linguists and philosophers. This ‘projection problem’—the problem of how to describe and explain when presuppositions project from an embedded context and when they do not—has been a major focus of the research on presupposition since the 1970s, and it is at the heart of the mental spaces account of the mechanism Fauconnier (1985, 1997) calls ‘presupposition float’.

### 3.2 Mental Spaces: Inheritance and Presupposition Float

The behavior of presuppositions in embedded contexts is one of the foundational puzzles motivating the original formulation of mental spaces theory. The argument—which to readers of this volume may by now seem hardly necessary to make—is that aspects of meaning that are difficult or impossible to handle under a purely formal or logical analysis prove much more tractable when considered as the outcome of cognitive processes of meaning construction, particularly ‘the consequences of minds capable of setting up networks of mental spaces and optimizing the content within’ (Oakley 2009: 168).
There are two points here of relevance to the issue of cognitive bias and linguistic perspective: \textit{networks of mental spaces and optimizing the content within}. Unlike the possible worlds of truth-conditional semantics, mental spaces are partial and contingent. They are fundamentally cognitive rather than logical entities, constructed and revised in people’s minds in response to linguistic and other prompts. In other words, mental spaces are a model of the representations that constitute the meanings we construct as we think and talk. These are linked together in networks that are built up dynamically in working memory and can be stored in episodic memory. One mental space leads to another, and mental spaces can inherit structure from other spaces.

In mental spaces theory as originally formulated, the formal principles that govern this inheritance find their place in the theory not on the basis of their psychological or neurological plausibility (though the theory as a whole certainly aspires to be cognitively plausible), but because of their parsimony and utility within the logical structure of the mental spaces framework. In other words, their proposed dynamics are justified by their efficiency in accounting for existing puzzles in semantics and pragmatics. They are certainly not presented first and foremost as accurate descriptions of how or when humans make mistakes. On the contrary, the arguments for their validity are grounded in how well they explain subtle and complicated successful navigations of complex semantic and pragmatic structure. They are plausible insofar as they get meaning construction ‘right’. So how do they fit in with something like the curse of knowledge, which appears to be all about how people get things wrong?

As discussed in Section 3.1, a traditional defining feature of presuppositions has been that they can survive under negation or one of a variety of other embedding contexts, but it is tricky to predict when presuppositions of embedded clauses will give rise to presuppositions for an entire sentence, or of a multisentence piece of discourse.

For instance, the cleft construction in a sentence like \textit{It was meeting his niece that changed Jack’s mind} triggers the presupposition that something changed Jack’s mind. This presupposition can project from a variety of embedded contexts, as seen in (7a–be).

\begin{enumerate}
\item [7] a. It wasn’t meeting his niece that changed Jack’s mind. (negation)
\item b. If it was meeting his niece that changed Jack’s mind, he won’t mind babysitting her. (antecedent of a conditional)
\item c. Was it meeting his niece that changed Jack’s mind? (question)
\end{enumerate}
d. Maybe/it’s possible that it was meeting his niece that changed Jack’s mind. (possibility modal)

e. Presumably/probably it was meeting his niece that changed Jack’s mind. (evidential modal, probability adverb)

f. Maria thinks that it was meeting his niece that changed Jack’s mind. (belief operator)

Such embedded presuppositions can be cancelled or blocked under the right circumstances, and it has proved difficult on a sentence-level analysis to provide categorical accounts of what those circumstances might be. For instance, consider a pair of examples like (8), discussed in Fauconnier (1997):

(8)  a. If John has children, then John’s children are bald.
     b. If John is here, then John’s children are in New York.

Despite their structural similarities, (8a) and (8b) do not behave the same way with respect to presupposition. The sentence as a whole in (8a) does not presuppose that John has children, while the sentence as a whole in (8b) certainly does. Why?

The mental spaces account of this phenomenon is as follows: presupposed information propagates through a mental space network according to a general principle of *presupposition float*: Information grammatically introduced as presupposed relative to one mental space will ‘float’ to the next space and continue to propagate to the next higher space, unless structure already in that space either entails the presupposition or is incompatible with it. So (8a) sets up a space (‘if…’) that is hypothetical with respect to the base (i.e. the sentence as a whole), as well as an expansion space...
(‘then…’) that is subordinate to that hypothetical space. In the hypothetical space, the proposition that John has children is set up non-presuppositionally. In the expansion it is set up presuppositionally. And so when it ‘floats’ from the expansion up to the next higher space, it ‘meets itself’ and stops—and as a result it is not projected to the base and does not hold for the sentence as a whole. Because the antecedent in (8b) does not set up structure that entails that John has children, the presupposition in that case floats freely all the way up through the network, and the sentence as a whole does in fact presuppose that he does.

The focus in the Fauconnier account is on the circumstances under which presuppositions are blocked from propagating all the way to the base, because this is the crucial question for the projection problem. But the presupposition float principle also means that whatever we know, we by default ascribe to the speaker’s reality and the conversational ground. This tendency obtains especially for that which we know implicitly; that’s the special thing about presupposition, as we’ll see in more detail in the next section. This projection of information within the mental space network begins to resemble the apparently erroneous projections associated with the curse of knowledge. In other words, the kinds of mistakes catalogued in the literature on egocentric biases predict that we will systematically project information from certain less general contexts to more general ones, and so—in the Mental Spaces account—do the patterns of sentence interpretation seen with embedded presupposition.

These examples demonstrate that there is at the very least a suggestive family resemblance here, but we have not yet shown that the patterns of projection from less general to more general in these two sets of phenomena are more than broadly congruent. To see most clearly how the habits of thought captured by the curse of knowledge are part and parcel of ordinary language processing at the semantics/pragmatics interface, we will want to consider the phenomenon of so-called ‘presupposition accommodation’ (Lewis 1979).¹

### 3.3 Presupposition Accommodation

Most presupposing constructions can be used informatively; that is, they can carry presuppositions that are not already part of the common ground and yet still be felicitous. For example, if I say, ‘I have to pick up a package

¹ Like almost everything to do with presupposition, the notion of ‘accommodation’ is not without its critics (see, for instance, Gauker 2008). However, the linguistic behavior itself—the fact that things that look like presuppositions can be used appropriately in contexts where the presupposed propositions are in fact discourse-new information—is generally accepted as a phenomenon in need of explanation.
for my sister’ to someone who has no idea whether or not I had any siblings, my interlocutor can easily accommodate the presupposition triggered by my sister (namely, that there exists a person who fills this role), taking that information to be part of a newly expanded but still backgrounded shared set of contextual assumptions for the discourse.

Under the common ground account (as seen in Heim 1983, 1992, following Karttunen 1974 and Stalnaker 1970), presupposition is a precondition for felicitous utterances, which imposes requirements on the context that must be satisfied. That ‘context’ of utterances consists of the common ground, defined as mutual knowledge or common beliefs among participants of a conversation. Under this account, (9) has as a requirement that the common ground includes the proposition that someone broke the vase; in other words, a speaker who asserts (9) does so assuming that this requirement has been satisfied.

(9) It was Margaret who broke the vase.

The utterance thus requires that it is taken for granted and not subject to (further) discussion that someone broke the vase. But, as Lewis (1979: 340) observes:

It’s not as easy as you might think to say something that will be unacceptable for lack of required presuppositions. Say something that requires a missing presupposition, and straightaway that presupposition springs into existence, making what you said acceptable after all. (Or at least, this is what happens if your conversational partners tacitly acquiesce.)

Indeed, it is not only possible but very common for presuppositions to be informative.

The particular subtlety of the way that an informative presupposition introduces new information into the discourse is that, by definition, it is not asserting the presupposed content but treating it as both already known and already part of the shared, implicit context. Thus presupposition accommodation also involves presupposition float, even when the presupposition is not syntactically embedded. While an explicit assertion obtains and can be disputed at the level of the speaker’s reality (SR in Figure 3), presuppositions seem to do something more: ‘floating’ by default, unless they meet themselves or are contradicted, up to the common ground.
So here we have the first ingredient of projection from embedded contexts into more general ones. The examples from the cognitive bias literature, however, don’t stop there—they show that people habitually ascribe information from one limited or embedded perspective in reasoning about a separate limited perspective. In addition to presupposition float, mental spaces theory holds that structure is sometimes also automatically projected in the opposite direction, via a principle called *space optimization*: ‘relevant structure not explicitly contradicted is inherited within the child-space’ (Fauconnier 1997: 112). The utterance ‘I wish that whiny brat Edward were your brother instead of mine’, for instance, sets up a base space of the speaker’s reality, and a subordinate ‘wish’ space that is counterfactual with respect to that base. In the base space, Edward is the speaker’s brother. In the wish space, he is not. Many facts and inferences about Edward that obtain at the base level are blocked from being projected to the wish space. For example, in the base space, Edward and the speaker have the same parents, share half of their DNA, and so on. Those properties do not project to the wish space because they are incompatible with the counterfactual. But other features of Edward that are facts in the speaker’s reality space are projected by default to the wish space, for example his (alleged) quality of being a whiny brat.

When space optimization works together with presupposition float, then, embedded or informative presuppositions can float up to the common ground and from there be projected into new child spaces via optimization, as illustrated in Figure 4. This result is exactly parallel to the ‘cursed’ information propagation pattern discussed in Section 2 and illustrated in Figure 1. In both cases, privileged information—information salient to the speaker that was not originally shared by her interlocutors—propagates through the network and is ascribed by default to all participants.
This kind of accommodation projection is connected to the curse of knowledge in two ways: First, because presupposition accommodation always involves presupposition float, it makes embedded information available for profligate projection into our representations of other embedded perspectives in the network, just as we see with analogous privileged information in the wide array of studies on the curse of knowledge and the illusory transparency of intentions. Second, it is also necessary *because* of people’s biased inclination to overestimate what of their knowledge is already in the common ground.

Recall from Section 2 that curse-of-knowledge effects are particularly well documented in the context of linguistic communication. Indeed, recent studies (Bromme, Jucks, and Runde 2005; Wu and Keysar 2007) suggest that the more information speakers share in a conversation overall, the more they overestimate their interlocutors’ knowledge about new topics. This means that even if we started out with a state of affairs in which the rule for these expressions was that one was licensed only ever to use presupposing constructions to reference information that was already in the common ground, we would very quickly run into a great many cases where a speaker had mistakenly assumed that some referent was known and salient to her audience, or that her audience shared her background knowledge and beliefs about the topic at hand. Thus, even if no one were deliberately using presuppositions to introduce new information, the occasion would certainly, indeed frequently, arise.

If (as they do) hearers will accommodate such uses, it is a short step to conventionally using presuppositions informatively—if it is clear that hearers *can* accommodate these uses of presupposing constructions, why not for efficiency’s sake bring uncontroversial information into the common ground?

---

**FIGURE 4. Projection through the network**
ground as a presupposition (cf. Givón 1995)? From there, of course, a variety of rhetorically motivated, and even underhanded, possibilities arise for using presupposing constructions as a way of treating potentially controversial information as uncontroversial, less readily available for refutation, and more readily available for default projection across a network of embedded perspectives.

4 ‘Cursed’ Projections

Studies of the curse of knowledge and related biases often treat them as a sort of lapse from ideal approaches to communication and understanding. The analysis in Keysar (2007), for instance, dwells on how these effects ‘introduce a systematic reason for miscommunication’ and suggests that our reliance on our own perspective in discourse means that we should consider misunderstandings not as ‘a product of noise and random error’ but an endemic product of our ‘disregard for the perspective of the other’. With presupposition, however, the structures of default projection from the origo to other perspectives constitute a completely ordinary and even necessary component of meaning construction.

Cognitive linguists are used to thinking about the quirks of other general human cognitive abilities as being reflected in and constitutive of the organization of language. As, for instance, psychological models of memory and categorization have been central to models of both grammatical categories and lexical semantics in cognitive linguistics, so too can psychological models of heuristics and biases in reasoning under uncertainty help to explicate many of the processes that structure pragmatic inferencing and information packaging (Lambrecht 1994) for referring expressions and clausal structure.

This approach fits in well not only with the linguistic-theoretical structures of mental spaces theory, but also with two related bodies of recent research on how people think and behave when they interact with one another: the role of motor and perceptual simulation in both linguistic and non-linguistic understanding, and the pervasiveness of immediate, unconscious behavioral alignment in face-to-face interaction.2

Together, the research on simulation and the research on interactive alignment suggest that we go through the world continually, automatically, and mutually synching ourselves up with the people we encounter: We echo their actions and utterances directly (Garrod and Pickering 2009) and perceptually (Pecher and Zwaan 2005); we borrow their syntax (Du Bois

---

2 For a fine overview of the former and its potential role in linguistic performance, see Bergen (2012); for the latter, see Louwerse et al. (2012).
— and they do the same with us. Even people with amnesia converge on shared terms for novel objects with their interlocutors over the course of a conversation (Duff et al. 2005). At the same time, we need to fill out these echoes rapidly and behind the scenes. Mapping information from our own perspective is a quick and efficient way of generating good approximations. The resulting frequent and persistent projection of perspectives is in keeping with both the kind of default interpretation we see across both standard curse of knowledge effects and the structural features of presupposition projection.

This phenomenon, incidentally, is one of the ways that stories can exploit the curse of knowledge to create clever misdirections and surprises. Many twist endings hinge on a hidden coreferentiality—Jekyll and Hyde are the same man, the Planet of the Apes was Earth all along, Harry Lime was the ‘third man’ at his own funeral, Darth Vader was once Anakin Skywalker. To have a character say ‘Mr. Hyde and Dr. Jekyll are two different people’ may well serve only to alert the reader that things are far more likely to be the other way around. But a sentence like ‘in case of the decease of Henry Jekyll… all his possessions were to pass into the hands of his friend and benefactor Edward Hyde’, presupposes, rather than asserts, its deceptions. When a crucial bit of misinformation is delivered via something other than a discourse-foregrounded entailment, the mechanisms of ‘project what you know’ that govern presupposition are particularly likely to produce the desired result: a misleading, but ultimately aesthetically appealing, systematic overpopulation from embedded perspectives.

Research on egocentric biases and reasoning under uncertainty thus turns out to provide psychologically and neurologically grounded support for some of the technical details of mental spaces theory that were previously motivated primarily for their explanatory and descriptive utility in handling linguistic data. The case of presupposition provides a promising model for integrating biases into our understanding of ‘correct’ interpretive phenomena, and also for the converse: we can and should use our finely observed understanding of linguistic phenomena to tease apart the uses and limits of egocentrism and automatic alignment in social cognition.

---

References


