Extra-compositional constraints and logical explanation

word count: 16,500

August 23, 2016

Abstract

The logic in natural language thesis, or simply the logicality thesis, forges an intimate connection between logic and language, trying to make respectable the idea that logical properties are instantiated in natural languages. Starting from the role of logical explanation in theoretical linguistics, I propose that logic and language are more intimately connected than some recent formulations of the logicality thesis suggest. In particular, the sceptical arguments due to Michael Glanzberg (2015) and Zoltan Szabó (2012) have no force against the logicality thesis understood as an explanatory hypothesis about language. Nor do they have force against a metaphysical claim about logico-linguistic properties. My discussion starts with three case studies of logical explanations targeting linguistic phenomena at the interface between syntax and semantics. These explanations are drawn from the recent linguistic literature and are due to Kai von Fintel (1994), Danny Fox (2000), and Tanya Reinhart (2006). The three logical explanations come out as explanatory on an epistemic view of explanation similar to the one proposed by Philip Kitcher (1989).

I argue that scepticism regarding the logicality thesis did not take logical explanations seriously as sources of evidence for the logicality thesis because it focused on compositional and truth-conditional phenomena rather than on extra-compositional linguistic constraints, which turn out to be crucial to logical explanation. I further argue that if we do not overemphasise the predictable differences between natural and artificial languages, we can formulate the logicality thesis on the basis of a robust point of similarity—viz. the modal character of inferences in both natural and artificial languages—that shows up in a unitary cohort of linguistic explanations and, moreover, is compatible with a pragmatic understanding of logic as a multi-faceted empirical discipline.

Keywords: *explanation, logic, unification, idealisation, extra-compositional constraints, truth conditions, logical form, semantics, the syntax/semantics interface, linguistics, general philosophy of science, philosophical logic.*

In thinking about the representational content of logic, some philosophers will be drawn, as I am, towards a simple yet strong metaphysical claim that there really are logical properties instantiated in key places of the world. One of the philosophers embracing this resolute view of logic, Gila Sher, puts it like this.[1]

> Logic, I believe, is grounded both in the mind and in the world, and its two grounds are interconnected. … Groundedness in the world is verdicality, i.e., compliance with strict standards of truth, evidence, and factual justification. Groundedness in the mind is conformity with pragmatic, conceptual, transcendental, linguistic, and possibly other extra-veridical norms. (Sher (2011, 354))

[1] Another realist about logical properties is McGinn (2000). Resnik (1999) identifies further realists and argues against their views.

Her claim, Sher adds, is "not deflationist [:] logic is both in the mind and in the world in a substantive sense, a sense that yields significant explanations, solves significant problems, and has significant consequences". (Sher (2011, 354))

I find this view congenial both in its metaphysical commitments (logic is part of the world as well as the mind) and its epistemic desiderata (that e.g. logic should provide significant explanations), but here I will be more concerned with the latter issue: the epistemic affordances of theories that appeal to logic.

Of all the theories that invoke logic, I will be looking at theories of natural language. My aim is to shed light on the ways logic may reside in natural language. I adopt as a working hypothesis Sher's unapologetic view of the metaphysical import of logic. I will not argue for this view here but for a natural extension of this view. My argument takes inspiration from an intriguing style of argument that I distil from scientific practice (and, in particular, theoretical linguistics) rather than from more fundamental considerations about the relation between logic and the world, as Sher's own argument does. I build a picture of how logic might plausibly be connected to language, and argue that this picture has a number of epistemic benefits as well as being metaphysically untroubled by a certain kind of sceptical argument.

Language can be seen as a complex cognitive ability, a competence in Noam Chomsky's sense.[2] This, as much, seems to be accepted by most philosophers of language and linguists. I think that logic provides an understanding of linguistic competence that goes beyond serving as a medium of representation for natural language sentences. Logic is explanatory, and the explanations that it provides target language rather than some mathematical subject matter. My plan here is to present a series

[2] See for instance Chomsky (2000) and Chomsky (1965).

of compelling case studies involving logical explanations of language. I put forward an epistemic view of *logical* explanation of grammaticality that emphasises one key explanatory value of logical properties. I then follow by arguing that scepticism regarding logical properties in natural language is inconclusive.

But who doubts that natural language has, among its other properties, logical properties? Here are two forceful exponents. One philosopher of language that is sceptical towards this thesis is Michael Glanzberg, who argues in a recent paper that

> [T]he logic in natural language thesis is false: we do not find logical consequence relations in our natural languages. (Glanzberg (2015, 115))

Glanzberg is not alone. After inspecting a couple of similar theses, Zoltan Szabo's conclusion ends up being as drastic as Glanzberg's.

> [A]ppeal to linguistic competence fails to save the intuition underlying logical hylomorphism. There are general reasons to doubt that logical competence (whether it is construed narrowly or broadly) can be part of linguistic competence. (Szabó (2012, 124))

Logical hylomorphism—the doctrine Szabo is alluding to—draws a sharp distinction between the logical *form* and extra-logical *matter* of natural language sentences and holds that the logical forms of sentences, in abstraction of their matter, explain the validity of arguments made out of such sentences. Zoltan Szabo thinks this is false.[3]

[3] More specifically,

> [According to logical hylomorphism] [t]he logical form of a sentence would be a characteristically logical arrangement of its parts; what gets arranged would be the sentences' extra-logical matter. The central thesis of this paper is that such a separation is impossible unless it is made actual by fiat. That is, except for artificial languages designed so as to have formulae factorable into logically significant form and logically insignificant

My task here is to argue for the opposite view that there is some significant logic to natural language and that logical competence is part of linguistic competence. I will call this thesis the logic in natural language thesis (borrowing Glanzberg's terminology), or, more succinctly, the logicality thesis.

Among the various sceptical stances toward the logicality thesis, I will focus on Glanzberg's and Szabo's criticism of the thesis because I share with them the assumption that speakers of natural languages are endowed with linguistic competence as well as their openness to a linguistic version of the logicality thesis—the (admittedly elusive) commitment that the logicality thesis should, to an important extent, stand or fall on empirical grounds.[4]

If the logicality thesis is to be a substantial empirical hypothesis, we should be able to determine its main theoretical virtues. One such virtue is its predictive power. Since an assessment of the predictive power of linguistic theories is a task better suited for linguists, I will instead emphasise another epistemic virtue of the logicality thesis, namely its explanatory power. I contend that one of the (if not *the*) most significant role of logic in linguistic theory is its *explanatory role*. Then the positive contribution of the paper is to show that logicality is best seen as explanatory on an epistemic view of explanation. In the end, what I hope to establish is that there are good epistemic reasons to endorse the logicality thesis, as well as no metaphysical obstacles to doing so.

matter, the separation cannot be made. (Szabó (2012, 107–8))

[4] Although many philosophers would agree in principle with this, in practice fewer are disposed to build on arguments provided by theoretical linguistics and, in general, they are less concerned with how to predict linguistic behaviour and what the resulting predictive hypotheses tell us about language.

1 Assumptions and aims

Theorists working at the interface between philosophy and linguistics will likely accept a distinction between logical forms as the semantically interpretable output of the natural language syntax and logical forms as more abstract properties of thoughts and sentences, which need not coincide with the former type of logical form. In effect this amounts to a methodological distinction between logical forms as resulting from the empirical study of language and logical forms as the result of (philosophical) logic study of inference. Rather than denying this methodological distinction, I will show that logical forms in the linguistic sense sometimes have extrinsic—relational or functional—properties which realise logical properties and forms in the philosophical logic sense.

In talking about the logicality of natural language, philosophers tend to overemphasise certain aspects of both language and logic, with the net effect of neglecting promising formulations of the logicality thesis. So I will take some time to emphasise the novel aspects of my logicality thesis, as opposed to the more familiar ones.

I will adopt a bare bones view of logicality. Thus I do not seek to spell out a complete set of properties that would make a piece of language logical. Nor do I aim to emphasise any particular logic over any other. To get some intuition about what logical properties are, it is useful to conceive of the logic of natural language as an extension of the first order logic (e.g. a higher order modal logic). This will do as long as we do not place high hopes on its perfect similarity to natural language. I'm not committed to any particular view of the 'right' logic, but just to the idea that some logic is necessary to explain certain facts of our language.

Logic, for present purposes, is a discipline that categorises inferences according to their modal status, that is, according to how *necessary* or *possible* they appear to be.[5]. We can think of the modal status of inferences in semantic, model-theoretic terms—although nothing here will hinge on that view. On the model-theoretic view, an inference or statement is necessary if it holds for all objects (properties, relations etc.) of an appropriate class. To illustrate, consider the statement *Every cat is a cat* or the corresponding inference: *if an object is a member of the set consisting of every cat, it is a cat*. Any of these is necessary if it holds no matter how we pick out the set of cats relevant to such sentences. Perhaps we mean cats in our neighbourhood (all of them), or fictional cats or any other possible cats. Since our statements will be true for any of these interpretations, they will come out as necessarily true. In other words, the different sets of cats that we might refer to are part of different models in which we evaluate our statements, and our statements come out as necessary if every model we can think of makes them true.[6]

The criterion of logicality implicit in our examples cannot single out logical properties (e.g. logical consequence). For there are necessities and possibilities that are not logical (e.g. physical possibilities). My intention is not to define logical properties, but to help identify them in the cases that we will survey. To single out logical properties, another step is considered standard: the identification of logical constants, things whose meaning is to be kept fixed across interpretations (e.g. *every* in our previous example). And this step is usually seen as up for debate. How to make sure

[5]This categorisation, in turn, is to be evaluated according to contextually given criteria of theory choice: fruitfulness, simplicity, and empirical adequacy

[6]It is, of course, important how we identify what counts as a relevant model for this evaluation. A necessary statement is true in every model selected with a given principle of selection (e.g. focusing on how matters stand in the actual world, as in the case where we pick out only actual cats).

that what we have picked are real logical constants, and how to distinguish them from other kinds of objects? I do not provide a principled answer to this question, but a pragmatic one. I claim that there is no disagreement about what we should keep fixed in certain scientific contexts, and such a distinction—the distinction between parts of the linguistic machinery that are kept fixed and those that are variable—is often made by the theorists themselves. In the end, whether logic is explanatory may be decided, in great part, by assessing the theoretical virtues of the hypotheses that appeal to such distinctions. (In brief, I am a holist about the validation of logically-driven theories.)

Logics will vary across many dimensions going much beyond the distinctions that I have just gestured at. However, what we need to retain from this is the requirement that a logic—a product of the discipline with the same name—needs to draw distinctions that can be modally categorised, as necessary, possible, or impossible. This is also the minimal sense of logic that allows us to talk about certain fragments of natural language that constitute good evidence for the logicality thesis.

This brings us to the logicality thesis and how we should conceive of it. For an opener, the logicality thesis need not be established quantitatively. It need not rely on counting the similarities and dissimilarities between some logical framework and some portion of natural language. Our task should be a qualitative one. Accordingly, I will not be seeking to establish the extent to which logical properties overlap in natural and in formal languages, but rather to check whether such an overlap is empirically justified and theoretically interesting. Thus even if the overlap is minimal, as long as it suggests new lines of research and involves properties that we independently expect to hold of human cognition, that overlap is significant enough to support the

logicality thesis.

The evidential basis for the logicality thesis has been traditionally construed as pertaining to the compositional semantics of natural language—which builds the meaning of a complex expression from the way its component parts are arranged and their individual meanings.[7] This has been the main focus in both the Montagovian and Davidsonian semantic traditions.[8] For a correct assessment of the logicality thesis, we have to broaden our focus. It is still insufficiently acknowledged that compositional semantics is not—and has not been—the only target of semantic investigation.[9] As linguistic theory advances and ramifies into quasi-autonomous domains like syntax and semantics, we get new sources of evidence for the logicality thesis. The linguistic data considered here belong to mechanisms at the interface between syntax and semantics. These mechanisms are *extra-compositional*. Their principles are not principles of semantic composition but additional constraints on it. Often philosophers argue for (or against) the logicality thesis neglecting that linguistic competence includes extra-compositional mechanisms. The key purpose of §§2.1-2.3 is to showcase several extra-compositional mechanisms, which, I will argue in §3, endow logic with an explanatory role and constitute an unobjectionable source of evidence for the logicality thesis—as shown in the more polemical §4.

[7] For instance, Barbara Partee (2013) suggests that "the question of whether natural language is "logical" [...] may be viewed as a version of the question of whether natural language semantics is compositional". This is not to say logicality of language should be tied to compositionality or that we need to endorse a version of compositionality in order to speak of logicality, but only that the problem of logicality is often treated in the context of a compositional semantics.

[8] See e.g. Davidson (2003, 155–70), Montague (1974), and Partee (2004, 153–81). I will not rely, in my argument, on any specific formulation of compositionality, since the logicality thesis is independent of it.

[9] Arguably, the generative semanticists (e.g. Lakoff (1971)) posited the so-called trans-derivational constraints to deal in part with logical aspects of language. This is an interesting historical point that might bear on some present issues, but here I will strictly concentrate on more contemporary issues.

In line with Sher's resolute view, I am happy to assume that there indeed are various forms of logicality: logical properties of language, logical properties of human reasoning, and logical properties of the outside world. I adopt this view not only because I find it plausible but also in order to probe the metaphysical force of sceptical arguments to the contrary. Since I mainly draw on empirical data, I am more concerned though with the *epistemic* contribution of logic and its implications. In which ways—I would like to ask—does logic help us understand the world and in particular human cognition. Instead of approaching these questions head on, I start by explaining certain linguistic puzzles arising from language use and then talk at length about logical explanation, and the logic in natural language thesis.

2 Logical explanations

If grammar conceals logicality, we expect two things: (i) that grammatical forms may be partly determined by logical properties (e.g. consistency) and (ii) that ungrammatical forms may be determined by the lack of those properties (e.g. inconsistency). These expectations are borne out. The relevant range of cases supporting the logicality thesis is very rich indeed, and my selection will be somewhat arbitrary. I will focus on three kinds of linguistic constructions that conceal forms of logicality (or illogicality). These case studies are meant to bring to light an explanatory paradigm as well as its implications, rather than to express my judgement about the most promising explanations within that explanatory paradigm.

2.1 The contradiction-tracking 'but'

Inference and entailment are staples of logicality. The case of exceptive clauses is interesting because it exhibits a different paradigm of logicality: inconsistency. The relevant type of inconsistency is not open to view but shows itself indirectly in certain grammatical anomalies. It may show up in the following contrast, for instance.

(1) *But*-exceptives

   a. No one but John smokes.

   b. #Someone but John smokes.

von Fintel (1994) finds that the ungrammaticality of (1b) follows from two properties of *but*-exceptives. For an exceptive to be grammatical it should be that

i. the relevant property (e.g. that of not being a smoker) is true of all the individuals in question *apart from* the exceptional individual (e.g. John). Thus, property P must be true of the individuals in $D - e$, where D is the contextually relevant domain of individuals and e is an exceptional individual in the domain D that distinguishes himself by not having the relevant property.

ii. the relevant property is *false* about the individuals in D, in general. In our example, not all of the relevant people are non-smokers, because John smokes. So it is false that all the people in D are non-smokers.

Otherwise put, we can explain *but*-exceptives starting from statements about the domain D, namely the statements that D is such that (i) all individuals in $D - e$, for some choice of e, have property P and (ii) it is not the case that all individuals in D have property P. In interpreting *but*-exceptives, both statements are assumed to be true—and, according to von Fintel's hypothesis, this is the case in virtue of the

semantics of the relevant linguistic construction.

The two conditions are both true in (1a). Smoking is true solely of John (the exceptional individual, e). Everyone else is not smoking. So it is the case that non-smoking is a property of the individuals in $D-e$, in compliance with condition (i) and it is false, in compliance with condition (ii), that all the people in D have that property—John doesn't.

So our first example satisfies the two conditions, and is therefore acceptable. Are these conditions satisfied by the second example, (1b)? We can easily check whether they are. According to (i), smoking must be true of someone but not John; removing John from the relevant set of smokers will leave us with a set of smokers. This must be true, on the plausible assumption that there are smokers at all. Now, include John back into the relevant domain of people considered to be smokers and you must get, according to condition (ii), a falsehood. The trouble is that we cannot satisfy condition (ii) while at the same time preserving consistency. Adding John to the set of smokers would not make anyone less of a smoker, so it remains true that someone smokes. Recall that in the previous example, adding John back in the domain of individuals considered non-smokers made it false that *no one* smokes. In contrast, adding John back to the set of smokers will have no effect on whether *someone* smokes, so it can hardly make that false. Then condition (ii) cannot be fulfilled if we want to avoid contradiction.

Rephrasing all this, the explanation depends on the assumption that we do try to adopt both of these conditions in order to make grammatical sense of the exceptive *but*, thus taking the following two statements to be true.

(2) Contradiction in (1b)

i. $\exists x \in D - e : x$ is a smoker.

ii. $\neg \exists x \in D : x$ is a smoker.

Compare this with the two assumptions made in the grammatical case, (1a).

(3) No contradiction in (1a)

i. $\forall x \in D - e : x$ is a non-smoker.

ii. $\neg \forall x \in D : x$ is a non-smoker.

There is an important difference between the internal consistencies of the two pairs of assumptions. By taking both of the statements in (2) to be true we run into a contradiction, because (i) entails that someone is a smoker, whilst (ii) entails that no one is a smoker. This is not so in the case of (3). It is consistent to maintain that not everyone is a non-smoker (ii) but everyone except a given individual is a non-smoker (i). So the two pairs of assumptions differ with respect to internal consistency.

To leverage this difference, the final move in the explanation is to spell out the ungrammaticality of (1b) as a plain old logical contradiction. Grammaticality with *but*-exceptives requires consistency between the two assumptions, while inconsistency or contradiction generates ungrammaticality.

It looks like logic plays an important role in this explanation. The explanation appeals to a consistency test, whose possible results—consistency or inconsistency—determine possible linguistic properties—grammaticality or ungrammaticality. There are several ways of spelling out the required notion of inconsistency. A basic intuition is that a set of statements is inconsistent just in case it *cannot* be made true no matter how we imagine the world to be. A more precise definition would emerge from addi-

tional assumptions about logic. If we adopt a syntactic framework, a contradiction in a set of statements is what allows us to derive a sentence and its negation by way of a sound set of axioms and rules of inference. In a semantic framework, contradiction is falsehood in every model of that set of statements. But independently of these frameworks, we are facing a semantic explanation that appeals to an intuitive logical property. This is our first instance of a logical explanation of language.[10]

2.2 The equivalence avoidance of quantifier scope

Quantifiers and other operators often interact, giving rise to ambiguities. For instance, the sentences

(4) a. All passengers are not honest. $(\forall > \neg, \neg > \forall)$

b. A student listens to every professor. $(\exists > \forall, \forall > \exists)$

have two readings each. Example (4a) may mean that every passenger is dishonest or, more plausibly, that some passengers are dishonest. Example (4b) may mean that there is a student, say Tina, that listens to every professor, or that each professor is listened to by one student or another, not necessarily the same. We can cast these interpretations as ambiguities of scope. The different scopes of operators—their different ordering relative to each other, as indicated parenthetically in (4)— determine

[10] As we suggested earlier, a logical property can be intuitively characterised in modal terms. Contradictions are those statements that *cannot* be true, entailments are those statements that *must* be true, and so on. Different logics flesh out these notions in different ways, and it is their job to accurately do so, given the particular domain of reasoning they are trying to articulate. For my part, I am not concerned with spelling out logical properties using any particular logical framework (though see the appendix for a start on that front). It is enough for present purposes that those properties are recognisably logical, in the sense that we would attribute necessity, possibility or impossibility to stretches of reasoning that instantiate these properties. In this sense, the logical properties under study here are independent of the framework we may adopt to flesh them out.

different interpretations. The resulting interactions between operators in different positions relative to one another give rise to different meanings.

But these interactions disappear in certain contexts. While the first sentence in (5a) below, like the identical (4b) above, admits two readings, the very same sentence in (5b) does not. Compare:

(5) a. A student listens to every professor. And a teaching assistant does too. ($\exists > \forall$, $\forall > \exists$)

b. A student listens to every professor. And John does too.($\exists > \forall$, $\#\forall > \exists$)

The first sentence in each of these examples, *A student listens to every professor*, can be read in two ways in (5a) but in only one way in (5b)—as represented in the parenthetical annotations. The reading that for every professor there is a different student that likes that professor disappears in the latter. Why is that?

This is because, Danny Fox (2000) argues, the second sentence imposes its scopal structure on the first through a parallelism principle. The parallelism principle asks that the two sentences in the above configuration have parallel structures. If the second sentence comes to mean that there is a different person for each professor, then the first sentence needs to have the same interpretation. And if the second sentence has a distinct interpretation, the corresponding (parallel) interpretation should be instantiated by the first sentence. Generalising, if the second sentence is less flexible than the first, then both sentences will wind up having the interpretation of the less flexible sentence. Scopal rigidity extends by parallelism.

The explanation of the contrast in (5) comes down to identifying the difference between the second sentences in each of the examples. Let us pin down the intuitive

difference between the two. The second sentences say, respectively, that a teaching assistant listens to every professor and that John listens to every professor. These sentences differ in one respect. The former is more flexible than the latter, as it allows two interpretations. That a teaching assistant listens to every professor may mean two things, namely that different teachers listen to every professor and alternatively that a unique teacher listens to every professor. This is the very same feature noticed in the scopal ambiguities displayed by (4). In contrast, that John listens to every professor may mean only one thing as far as the scopal interaction between the quantifiers in subject and object position is concerned. Hardly anyone would read this sentence as saying that there is a different John that listens to different professors (except perhaps for Humpty Dumpty).

An additional assumption is needed to complete the explanation. Different (syntactic) orderings of the subject and object of a sentence are allowed only when these orderings produce different (semantic) interpretations. Then the orderings available for the second sentence are projected into the parallel ones available for the first sentence (if such orderings are possible there at all). What happens in our puzzling case (5b) is that even if the different orderings (and the corresponding interpretations) are possible in the first sentence taken independently (as in 4b), parallel ones are not possible in the second sentence, and this ends up constraining the orderings and interpretations available in the first sentence, via the parallelism principle.

According to this explanation, the semantics of scope has effect on its syntax. Crucially, the explanation relies on the identification of the different meanings or interpretations of a sentence. But whether a sentence produces one or more (propositional) interpretations is, semantically, a question of the *entailment relations* between these

interpretations. Two interpretations are identical if they mutually entail each other. They are different if they do not mutually entail each other. So we are facing another form of concealed logicality. The logical explanation that appeals to it is summarised below.

(6) a. A sentence of the form O_1O_2 (we represent the sentence as $S[O_1O_2]$), with two operators in subject and object position respectively admits an inverse syntactic ordering O_2O_1 iff the interpretation thus created is distinct and new, viz. iff $S[O_1O_2] \not\equiv S[O_2O_1]$.

b. For any two sentences that are part of a parallel-driven utterance (like 5a-b), their syntactic orderings must align in the sense that an ordering in the first sentence is allowed only if a parallel ordering is allowed in the second sentence.

2.3 Escaping equivalence with exceptional binding

Suppose Oscar loves himself but no one else loves him. We may try to express our thoughts about this state of affairs in two ways, with mixed success:

(7) a. #Oscar$_1$ likes him$_1$.

b. Only Oscar$_1$ likes him$_1$.

The indexes represent identity. That is, the noun phrases *Oscar* and *him* refer to the same individual under the intended interpretation. Yet the intended interpretation of (7a) is unacceptable when uttered out of the blue (with no associated ostension). *Him* may not refer to Oscar there. Only another pronoun, the reflexive *himself*, can do that job.

The situation is different with (7b). There, it *is* possible to read *him* as referring to Oscar. Why is that?

A suggestion that has recently gained traction in the linguistics literature is the one advanced by Tanya Reinhart (2006). Reinhart argues that in the context of (7a), the reflexive pronoun *himself* can express the intended interpretation of the plain pronoun *him* on which the pronoun refers to Oscar. Given that the interpretation of *him* as referring to Oscar is syntactically ill-construed, there is no reason to use the pronoun to express that meaning. There is a better pronominal form that conveys the same message, namely the reflexive. Now, turning to (7b), can we argue in the same way against the plain pronoun *him* in that context? Is the reflexive *himself* a more apt way of expressing what can be expressed using *him*? Some linguists argued that it is not. The reflexive *himself* in a context like (7b) would come to mean that only Oscar likes himself and no one else loves oneself. But this loveless state of affairs is not what we were supposed to be describing with (7b). We did not want to say that nobody else loves themselves. What we wanted to say is that nobody else loves *Oscar*. The plain pronoun *him* is better suited to express the latter state of affairs. So, the argument runs, even if the plain pronoun is syntactically not as good as the reflexive pronoun, it gets the better of the reflexive in the context of (7b), as it conveys a genuinely different message. No similar plea is available for the plain pronoun in (7a).

So (7) exhibits the contrast that it does because plain pronouns distinguish themselves from reflexives only in restricted contexts—contexts like (7b) rather than (7a). This distinction between pronouns conceals yet another form of logicality that can be uncovered by considering how the distinction unfolds at the propositional level.

The explanation just given appeals to a principle that may ring familiar:

(8) A plain pronoun x in object position (e.g. *him*) can replace a reflexive pronoun y (e.g. *himself*) in the same position for the purposes of referring to the individual denoted by the subject (e.g. Oscar) only if the plain pronoun can convey a different proposition. Thus x is allowed if there is a salient 'reflexive' proposition p of the form $\langle o\ \lambda y\ .\ Pyy\rangle$ expressed via the reflexive pronominal y, and this proposition is not in a relation of mutual entailment with the proposition q expressed via the plain pronominal x, a proposition of the form $\langle o\ o\ \lambda x\lambda z\ .\ Pzx\rangle$ (where o is the denotation of both the subject and object noun phrases and P is the property expressed by the verb phrase). In brief, x is acceptable only if $p \not\equiv q$.

The strategy underscoring this principle is the same as the one employed in explaining contrasts about scope. It is based on identifying distinct propositional meanings: propositions that are not mutually entailing. The strategy is applied to a new domain, to do with the binding of pronouns, but it continues to invoke a logical property—entailment. So this case too involves wielding logicality in order to explain grammatical unacceptability. On this logical explanation, example (7a) is ungrammatical because it does not bring out a new meaning relative to the reflexive-driven meaning.

To sum up, our three case studies represent different kinds of logical explanation. While the first case of *but*-exceptives involves ungrammaticality spelled out as inconsistency, the latter two cases involve ungrammaticality spelled out as equivalence. It is immediately clear that contradiction is bad, and for my purposes I will trade on this

intuition.[11] Equivalence, in contrast, does not arise suspicion at first. In what sense is equivalence bad, grammatically? The reason can be traced back to the relation between syntax and pragmatics. Given that certain linguistic forms are grammatically marked, and thus exhibit syntactic structures that should be avoided, the fact that they express equivalent meanings to those of syntactically better expressions will count against them. The syntactic costs incurred for adopting grammatically marked expressions are not feasible. Were it the case that these grammatically marked expressions told us something new (viz. non-equivalent), their novelty would outweigh their slight grammatical inadequacy, which will make them acceptable, on balance.[12] So the equivalence of meaning cannot rescue those expressions that stretch grammar.

3 Logical explanation as unification

What does all this tell us about the role of logic in explanation? The epistemic conception of explanation can help us pin that role down. We will make use of a particular epistemic view, the unificationist view (Kitcher (1989)).[13]

In adopting the epistemic view I do not deny any ontological claims that ontic theorists might want to make about the targets of linguistic explanation (mechanisms,

[11] Philosophically though, it is not as clear why we should avoid contradiction. Perhaps for its nefarious practical implications, as Turing once suggested (see also Sher (2011)). But clearly we have to say more in order to uncover the normative grounds against contradiction.

[12] It is important that the inadequacy should be minor, lest any ungrammaticality would come out to be worth its while. In the case of *him* vs *himself*, the inadequacy introduced by the former is minor, as it just lacks the reflexive feature of the latter, preserving all the others (it is still a singular, masculine noun phrase form).

[13] Kitcher himself advertises his view as capable of giving an account of grammatical explanation, but he does not offer us a proof of concept. At any rate, if what I'm saying here is on the right lines, Kitcher's contention is borne out.

causes, structures etc.).[14] In fact I agree that logical properties have a basis in the world. However, it is not easy to extract a view about what this basis amounts to from linguistic theory, which is mainly working at a level of abstraction distinct from the most plausible locus for cognitive capacities—the neurobiology of the brain. Nor does the field of neurobiology offer any ready resources to approach the problem of explanation in theoretical linguistics. Currently, the epistemic view of explanation is much better suited to illuminate logical explanation, since it can rely on the knowledge and analytical tools of the theorists rather than on more basic yet inaccessible facts. In giving an account of linguistic explanation, thus, we can do no better than appealing to the kind of information (theoretical assumptions, models, and observations) that the explainers themselves possess in that specific field of inquiry.

An explanation is, on this epistemic view, an argument or derivation whose result is not only to give a rationale for the phenomenon to be explained (the explanandum) but also to systematise the existing knowledge about that and other phenomena. How explanatory an argument is depends on how many other facts can be captured by the same kind of argument. If the argument reflects a pattern of reasoning that is useful for understanding different phenomena by way of similar arguments—arguments sharing the same pattern—then the argument is all the more explanatory.

Coupling this epistemic view of explanation with our case studies of logical explanation, we get a simple way of identifying the explanatory role of logic.

[14] For different version of ontic theories that imply ontic desiderata on explanation see e.g. Craver (2014) and French and Ladyman (2011).

3.1 Intra-linguistic unification

One sense in which our explanations provide unification is through the use of logical properties, which makes them share a common pattern of reasoning. The use of logical properties is just part of this common pattern. If we re-examine our case studies, we will quickly note that they all involve a comparison between logical forms, one of which is the result of the grammatical derivation which is supposed to go through for semantic interpretation and another logical form which is brought in only to test the former, via a comparison according to a certain criterion. This criterion, along with the logical forms which it governs, constitute an *extra-compositional* mechanism. This mechanism is extra-compositional because its purpose is not to construct linguistic structures (either syntactically or semantically) but rather to impose a constraint on the legitimate outputs of this construction.

Now, these criteria can in principle regulate comparison between logical forms and their alternatives in several ways. We may compare logical forms according to how syntactically anomalous they are (e.g. passive forms are less normal then active forms), or according to how syntactically and phonologically complex these forms are (e.g. a sentence with further embeddings is more difficult to build and pronounce), or according to some semantic criterion.

Our case studies rely on semantic criteria, and, in particular, on the logical relation between two logical forms. Whether or not a logical relation holds between two logical forms is the deciding factor for the grammaticality or ungrammaticality of one of these forms. Thus, whether the proposition expressed by a *but*-exceptive clause is consistent with its 'alternative' determines whether the exceptive clause is

grammatical. Whether a scope disambiguation or a exceptional binding pattern is logically distinct from (viz. non-equivalent to) an alternative logical form determines whether that scope or binding construction is grammatical. So comparison between logical forms according to logical criteria forms the core of the extra-compositional mechanisms invoked to explain our earlier data. And what is common between the arguments invoking such mechanisms is that the logical properties of the (linguistic) logical forms in question determine the grammatical status of these logical forms. (This is clear from the three principles invoked in our case studies, which I formalised in a bit more detail in the appendix.)

Here, we can plainly see that certain logical properties (individuated logically) combine with the logical forms of sentences (individuated linguistically) to determine the grammatical properties of the sentences in question. So we are not talking of either mere logical properties or mere lingusitic properties, but about a combination of the two. Both (linguistic) logical forms and their logical properties conspire to determine further grammatical properties, which we could not obtain without both of these components being in place.[15]

According to the unification picture, a key function of providing explanations is to systematise our knowledge by advancing arguments instantiating patterns of reasoning that are re-useable in arguing about various aspects of language compatible with our knowledge. It is in this power of systematisation that our logical explanations

[15] Earlier, I have called the logical properties of logical forms that make up our logical explanations *extrinsic* or *functional* properties of logical forms. By this I meant to hint at the comparison between two logical forms according to logical properties. Logical properties are extrinsic or functional in the sense that they come to be grammatically relevant only when a comparing (extra-compositional) mechanism is called for in order to test and attest the grammaticality status of a given logical form—which is not always the case, since some grammatical properties are determined by looking at at syntactic properties alone (e.g. case agreement).

excel. They bring well-known logical principles to bear on issues of grammaticality. The above logical explanations single out the relevant logical principles and connect them to judgements of grammaticality in an integrated argument.

To see that it is systematisation we are talking about, consider an alternative explanation of the *but*-exceptive cases. Recall that the challenge was to explain why the locution *all but one* is grammatical while *some but one* is not, despite their equivalent syntactic constituency. The logical explanation of the contrast appealed to the consistency of the former locution and the inconsistency of the latter. Suppose that we posit the alternative explanation that the ungrammaticality of (1b) is due to the fact that the two conditions cannot be fulfilled at the same time, rather than the fact that they are fulfilled in a way that leads to inconsistency. Thus the alternative explanation is not that we assume that the two conditions (i-ii, §2.1) are in place and derive a contradiction, but rather that one of the conditions fails (before a contradiction may arise), and because of this failure we are penalised with ungrammaticality. This may seem to explain equally well the data at hand, but would not generalise, making it a poor replacement for the corresponding logical explanation. Whatever its other merits or demerits[16], the argument thereby given would not share a pattern with the other two arguments and thus we would lose the logical explanation's force of unification.

[16] In §4.1 I argue that giving up logicality means giving up explanatoriness altogether, rather than just losing the unification power of the logical explanation (which amounts to losing only a part—a key part—of the explanatory value). In other words, without logicality we cannot even offer an argument that yields the explanandum as a conclusion, let alone an argument that shares a pattern with other explanatory arguments.

3.2 Extra-linguistic unification

We will have occasion to return to draw further consequences from our case studies, but for now what is important is that the appeal to logical properties is the common pattern of reasoning that makes our explanations more unified, and thus more explanatory. This is the *intra-linguistic* aspect of unification by logical properties, since, in putting forward logical explanations, we unify various empirical domains: at the very least, exceptives, scope, and binding. There is also an *extra-linguistic aspect* of unification, which is responsible for a good part of the surprise-value of logical explanations. That is, logical explanation provides a unification of the grammatical and reasoning realms. To see that, we need to flesh out what it means for logical principles to systematise existing knowledge.

Let us assume that the broadest category of knowledge that linguistic explanations tap into is knowledge of cognition. We explain linguistic phenomena by working with cognitively-significant knowledge that includes grammatical principles, judgements of acceptability, and other theoretical assumptions about cognition. The unification picture, we have said, casts explanatory arguments as those arguments that share a significant pattern with other arguments. The extra-linguistic aspect of unification and thus of explanation is that our explanations reveal patterns of reasoning that are common not only at the level of elaborate human reasoning but also, unexpectedly, at the grammatical level. Logical contradiction and equivalence do not appear only when thinking (incorrectly) that, e.g., it is raining and it is not raining, or when thinking (correctly) that, say, John is a man because it is not the case that he is not a man.

Logical principles also constrain, as our cases illustrate, the grammar of certain constructions, and ultimately how we use words: conjunctions like *but*, quantifying determiners like *all*, and pronouns like *him*. Thus logical contradiction and equivalence are not only vindicating or vetoing patterns of reasoning but also grammatical patterns. Hence logical principles provide an intriguing local unification of our logical and grammatical capacities. (The unification is local because it concerns the explanation of a limited class of linguistic constructions and it is not meant as a wholesale unification, which would lead us to the strong but unsupported view that grammar is logic.)

This explanatory role of logic is surprising, yet not so surprising as to deprive our explanation of cohesiveness or credibility. We have independent reason to accept logical principles as a component of our cognition. And we accept—at least if we work within the paradigm proposed by Chomsky—that language will be responsible for at least part of the traits of human cognition. We should then accept that the logical aspects of cognition may be built into our linguistic capacity. Our logical explanations realise this possibility in a concrete fashion.[17]

Given this explanatory success, we can say even more. It is appealing to posit that all the inferential abilities required for competence with the sentences surveyed earlier

[17] The question of what Kitcher (1989) calls stringency arises in the case of logical explanation as well. Stringency is the degree of similarity between arguments, or, otherwise put, the degree to which they share a common pattern. So how stringent is a logical explanation? One way to view the commonalities between logical explanations is to see them as invariance properties. On this view, contradictions are those statements which cannot be true in any model. And equivalences are among those statements that are true in every model. Then the hypothesis emerging from the first case study is formulated in terms of invariance of falsehood (viz. contradiction), and the hypothesis in the second and third case studies are formulated in terms of invariance of truth, or truth preservation (namely, equivalence). Indeed, these invariance properties are a good ground for the logicality of certain linguistic computations, and thus for the stringency of the logical explanation, which in appealing to these linguistic computations are also invoking a common logical pattern of explanation.

are part of a more complex logical ability. Recognising contradictions and recognising logical equivalence relations should be part of a unique ability of testing the truth values of statements e.g. by traversing a space of models (roughly, situations in which these statements can be evaluated). One way we can see this is as a semantic extension of Evans's (1982) generality constraint: if one can form the thought that *a* is *F* then one may also form the thought that *a* is *G*, where *a* is a statement, *F* stands for *true in every model* and *G* stands for *false in every model*. Clearly, the task is more complex than when the predicates are basic—as Evans envisaged it—and requires the evaluation of thoughts rather than merely their construction. But this is also a natural hypothesis; the simplest and the strongest we can get. Thus, until serious reasons for doubt show up, the logical ability needed to figure out the grammaticality status of *but*-exceptives, non-reflexive binding and scopal constructions can safely be considered one and the same ability. This view of a unique logical capacity at the service of grammar complements well the epistemic view of logical explanation as logically-driven unification. And, together, these views line up with Sher's contention that logic resides in language and mind (inter alia), and that it provides a better understanding of the world.

4 Defending logicality against scepticism

In arguing against the logicality thesis, Michael Glanzberg and Zoltan Szabo put forward views of natural language semantics that leave no room for logical properties. In assessing their arguments, I will focus on two issues. First, what kinds of linguistic phenomena are considered as evidence for reaching sceptical conclusions about the

logicality thesis? Can the logic in natural language thesis be proven true, given an enlarged evidential basis including the earlier data? My response to the latter question will be positive. Basically, sceptical accounts overlook evidence by their very theoretical set up. I will argue for this in the next section and then, after rehearsing the evidential and epistemic advantages of the logicality thesis, I turn to the second question of whether, once the new evidence is counted in, the sceptical objections do not still pinpoint to more foundational drawbacks of the logic in natural language thesis (§4.2).

4.1 Evidential question

So what is the evidence for languages' lack of logical properties? Both philosophers are very clear about the source of evidence. Szabo focuses on compositional truth-conditions and Glanzberg stipulates that semantics is just about truth conditions. To begin with, Szabo announces early on in his paper what view he wants to rule out.

> What my view is really incompatible with is the broadly Davidsonean idea that we can squeeze genuine logical forms out of a compositional semantics for natural language. That, I think, is a hopeless project. (Szabó (2012, 107–8))

If this broadly Davidsonian idea is all there is to the logic in natural language, I cannot disagree. The logical explanation surveyed above are not even fit to defend the broadly Davidsonian thesis. But, as we will see, Szabo seems to be committed to a stronger claim than this particular passage suggests, and, be that as it may, Szabo

does not so much as argue (or even purport to argue) that there is a principled reason to tie the logicality thesis to what we can squeeze from the compositional and truth conditional view of semantics, Davidson-style. Why should we content ourselves with this potentially limiting view of semantics as the gateway to logicality?

Glanzberg, on the other hand, does aim to give us some reason to restrict the logicality thesis to truth conditional semantics, whether this is to be conducted in the (truth-theoretic) manner of Davidson or in the (model-theoretic) manner of Montague. He contends that the semantics of natural language is absolute, rather than relative to a space of models. Even in the model-theoretic tradition initiated by Montague, what semanticists are concerned with is not how linguistic constructions behave in alternative models—consisting, roughly, of the objects, properties and situations that would satisfy those linguistic constructions. Instead, linguistic analysis aims to state the right truth conditions. These are part of the absolute model (the real world, if you wish) relative to which linguistic expressions are to be evaluated. In one fell swoop, Glanzberg confines semantics to the job of stating truth conditions *and* excludes logical properties from the domain of semantics. Since logical properties are defined in terms of a space of models—and thus are global properties grounded in this multiplicity of models—they cannot be within the reach of semantics given that the empirical domain of semantics consists solely of an absolute model—a local property grounded in such a model.

One apparent problem for his view, Glanzberg notes, is that certain properties of quantifier phrases are stated in terms of global properties, as invariance of a property over a class of models. Glanzberg argues that this is not a real problem for the view that semantics deals in truth conditions (a local matter) rather than in a space of

models (a global matter). He says:

> Attending to the local versus global distinction, we can reconcile two facts that might have seemed in tension. First, familiar determiners in natural language have more or less the semantics that logical theory says they should. [...] But the reason is simply that semantics of natural language uses local properties of quantifiers in spelling out the semantics of determiners. These are already available to absolute semantics. Now, this does not mean we can never look at the global notion of quantifier in thinking about natural language. The basic idea for giving absolute truth-conditions is the local one, and in fact, sometimes *we can get interesting further results* out of local definitions. But on occasion, *we learn something by abstracting away from absolute truth conditions*, by looking at global generalized quantifiers [:] [...] natural language determiners express restricted quantification. In looking at this sort of global property, we are not simply spelling out the semantics of a language. Rather, we are abstracting away from the semantics proper – a specification of contributions to truth-conditions – *to look at a more abstract property of an expression*. It turns out, in this case, abstracting away from the universe of discourse is the right thing to do. Particularly when asking about *logical or more generally mathematical properties of expressions*, this sort of abstraction can be of great interest [...] even if it goes beyond the semantics of any language per se. This sort of possibility shows how we might take the step from semantics proper to logic [...] (Glanzberg (2015, 99–100), my emphases)

One thing is clear at the outset: there is a distinction between the global (logical) properties and the local (truth conditional) ones. We can blackbox the meaning of global properties for the moment and focus on how the argument is framed. The question is what we state when we state global properties. In Glanzberg's words—I italicised the relevant phrases above—we get interesting further results, thus learning something by abstracting away from absolute truth conditions. What are these results about? What it is that we learn when we move from local to global properties? The answer is, as the above paragraph explicitly states, that we learn something about the more abstract properties of expressions. A natural way to read this answer is that these abstract properties, for all their abstractness, are *linguistic* properties as well, since they belong to linguistic expressions. Yet Glanzberg hastens to add that the properties in question are not part of semantics.[18]

If we endorse the view of semantics according to which semantics is exclusively in the service of stating truth conditions, we ban logical properties defined model-theoretically from semantics. On the resulting view, the 'abstract properties' of expressions are not semantic. There are two limiting points here. First, this stipulation in itself does not vindicate scepticism about the logic in natural language thesis. At best, it just vindicates scepticism about a related—yet narrower—thesis. We can

[18] Why not? I am not entirely sure. Glanzberg suggests, following Lepore's (1983) argument, that semantics must stay absolute, on pain of loosing a feature of our meaning, namely that it is stable. Admitting multiple models in our semantic theory is presumed to not preserve this stability or realism about meaning. Glanzberg goes further than Lepore in drawing the conclusion that the exclusion of logical (global) properties is the price to pay for realism—and the avoidance of relativism—about meaning. But he does not say why we cannot use an absolute model to satisfy or realist commitments about truth-conditions and a space of model to satisfy our other linguistic commitments, including logico-linguistic properties. Surely, this theoretical apparatus consisting of models is meant to do the representational work we want to put it to. So why not use both local and global properties for different representational purposes in semantics? As I will register later (in §3.2), I cannot find any satisfactory answer on behalf of scepticism.

call it the logic in natural language *semantics* thesis. Recall though that Glanzberg's anti-logicality thesis was a bit more ambitious. As he puts it, in full generality, "we do not find logical consequence in our natural languages" Glanzberg (2015, 115). So perhaps we should limit this claim to semantics. That remains a strong claim and a pain for the advocate of the logicality thesis. But—and this is the second and more important point—we do not have to stomach the stipulation about the domain of semantics either.

Instead, we should keep to a *semantic* version of the logic in natural language thesis, albeit not the familiar one. On the proposed version of the thesis, the aforementioned abstract properties of linguistic expressions are not only linguistic properties but semantic ones. It is only that they are *not* part of the truth conditional content of expressions or the compositional mechanism that composes these expressions' meanings. Bluntly put: scepticism about the logicality thesis trades on misconstruing what a semantic property can be. What a semantic property can be is nicely illustrated by the non-compositional mechanisms described in our case studies.

The running theme of the previous sections has been that logical properties—(in)consistency and (non)equivalence—determine judgements of grammaticality about the distribution of *but*-exceptives, non-reflexive pronouns as well as the scope of quantifiers. The corresponding explanations target the interface between syntax and semantics, which means that they involve both syntactic and semantic properties. The syntactic part is the grammaticality status of the logical forms involving *but*-exceptives, non-reflexive pronouns and quantifier scope. The semantic part is the presence (or lack) of consistency or equivalence between these logical forms and their alternatives—the alternative logical forms they are to be compared against.

Our logical explanations connect these syntactic and semantic parts.

Yet, crucially, the semantic part is not constitutive of the truth conditional content of sentences like e.g. *No one but John smokes* or *#Someone but John smokes*. These sentences, if we are to believe the earlier explanations, are indeed engendering consistency and inconsistency (respectively), but not as part of their truth conditions. For, surely, the consistency and inconsistency in question cannot spell out the truth conditions of those sentences. Instead, these logical properties have a more roundabout tie with the worldly conditions under which such sentences are true. They are spelling out the semantic conditions under which certain sentences are grammatical. Hence, according to our logical explanations, logical properties are conditions that should be met *before* questions about the sentences' truth conditions can even arise.

In more linguistic terms, an otherwise interpretable syntactic tree (or derivation or logical form) will be grammatically licensed only if its semantic interpretation compares favourably with the semantic interpretation of a another ('alternative') syntactic tree, which is constructed in order to test the former tree—viz. the one that is to be grammatically licensed and thus deemed well formed. In testing our tree, we compare it to its 'alternative' in logical terms. The type of question we ask in testing this tree may include: is its semantically interpreted logical form logically equivalent with this other logical form? Or: is it logically consistent with this other logical form? By noting that these questions concern the relations between two logical forms, we can ascertain that they are not about rules of composition and of actual mapping between syntax and semantics. We cannot even ask the earlier questions if we limit ourselves to the study of the (purported) compositional process that builds the truth

condition of a sentence. Yet this is precisely the kind of questions that we ask in order to explain the grammatical features of contrastive *but*, plain pronouns and quantifier scope in the three cases and others.

The main problem with scepticism about the logicality thesis is evidential, as it overlooks the empirical significance of non-compositional, non-truth conditional semantic properties that conceal logicality. The problem stems from the way Glanzberg and Szabo set up the kind of properties that should be scanned and exploited in order to establish the logicality thesis. For both philosophers these properties can only be truth conditional, compositional properties. Given these strictures, the task of arguing against the logicality thesis appears to be overambitious.

To be fair to Szabo, he does note, at some point, the "apparent relevance of logical features of linguistic expressions to their grammaticality" and points us to the kind of data we have seen in the previous sections (Szabó (2012, 121)). This is just one small step away from granting the logicality thesis. Yet, strikingly, Szabo also thinks that there are conclusive reasons for doubting that logicality is part of linguistic competence. Isn't there a tension between Szabo's two commitments? For if logical properties determine (or are relevant for) grammaticality, this is reason enough to believe that they do so in virtue of the logical forms of linguistic constructions, and, sidelining the notion of logical form altogether, it is surely reason enough to believe in an intimate connection between logic and language, and thus in the logicality thesis.

There is, I think, a straightforward explanation for why Szabo's commitments are not in fact in tension, but this does no good to scepticism about logicality in natural language. Szabo's contention, like Glanzberg's, is due to a restrictive view of

semantic properties and thus of semantics. He would not see the logical properties involved in the linguistic hypotheses explored here as validities in virtue of logical forms of sentences (in the logical sense). Nor would he see the mechanisms posited by those hypotheses as belonging to the compositional mechanism. This much is clear from the passage quoted at the beginning of this section. While I think this is about right[19], the restrictive view of semantics does not vindicate his argument against a linguistic version of the logicality thesis. One key target of his criticism is the thesis that there are "inferences whose validity can be recognized on the basis of linguistic competence alone, and among these inferences whose validity can be recognized on the basis of grammatical competence alone" (Szabó (2012) [120; cf. 122]). But the grammaticality of e.g. *but*-constructions and exceptional binding constructions *is* supposed to be recognisable in virtue of our recognising the validity (and other logical properties) of the derivations that license the grammaticality of those constructions. And our recognition of these logical properties *is* supposed to be part of our grammatical competence.[20][21]

[19] This is right in the sense that to explain our data we need more than the logical forms of *sentences*—we also need the logical forms of their competitors as well. And, as we have suggested, the relation between the logical forms of sentences and the logical forms of their competitors is not part of the compositional mechanism.

[20] This is not a wholesale rejection of Szabo's position, which, I take it, is mainly directed against logical hylomorphism—the thesis that sentences have both forms and matters, where the latter are responsible for logical properties, e.g. validity. But this does not touch on the logicality thesis. Even if it were true that no such ('purely logical') forms can be identified, this would not entail the stronger claim about linguistic competence, namely that linguistic competence does not involve logical properties. There is much more to linguistic competence than 'purely' logical forms. (Strictly speaking, linguistic competence is not tied to 'linguistic' logical forms either, as there are versions of syntactic minimalism that do away with that level of representation altogether; see e.g. Collins (2007, 811).) My objection at this juncture is that Szabo slips too easily from claims about logical forms of sentences to claims about the relation between logical and linguistic competence, as can be seen by inspecting e.g. Szabó (2012, 124) (the paragraph cited at the beginning of the paper, on p. 3) as well as the passage just cited.

[21] Related to Szabo's critique of the logicality thesis is an argument due to Timothy Williamson. Williamson calls into question the following principle that connects logical and linguistic competence:

(9) Necessarily, whoever understands the sentence 'Every vixen is a vixen' has a disposition to assent to it. Williamson (2007, 100)

So, the question arises, why can we not endorse the logic in natural language by looking beyond the confines of compositional, truth-conditional contents? What is wrong with the notion of logicality invoked in our case studies?

Before trying to answer what's wrong with the logicality invoked in our case studies (see §4.2), I would like to elaborate on what is good about the notion of logicality emerging from these case studies.

4.1.1 Beyond truth conditions

Our case studies involve genuine logical properties, properties that are not merely based on regimentation in a formal language. Logicality plays the key role in explanation. Had we invoked only non-logical properties instead, we could not have got

Williamson then discusses the case of negative polarity items (e.g. *any*) which seems to support the thesis that grammatical properties are determined by the downward entailing property. This is a logical property. Here is Williamson's explication of the property as well as his conclusion regarding its relevance to the connection between logical and linguistic competence.

> Suppose, nevertheless, that our classification of [negative polarity] strings as well- or ill-formed does depend on some prior classification of contexts as downward entailing or not. The question remains: *is that classification available for unconscious reasoning that would issue in conscious assent to supposedly analytic sentences?* To identify a context as downward entailing involves a more sophisticated logical insight than identifying a particular argument as valid, since it requires the validation of an abstract pattern of argument. For example, identifying negation as a downward entailing context requires checking this schema, for arbitrary sentences "A" and "B": If "A" entails "B" then "It is not the case that B" entails "It is not the case that A." That is just the kind of abstract formal reasoning task on which humans perform worst. Contrast that with our high level of reliability in determining whether strings with negative polarity items are well-formed. Thus the evidence suggests that the unconscious logic in question is not at the service of the cognitive processes that normally produce conscious assent to sentences like (1) ["Every vixen is a vixen"]. Such cases therefore fail to support a modification of the conclusions reached so far [namely, the sceptical conclusions about thesis (9) and the like]. (Williamson (2007, 108–9), my emphasis)

This may be right, in general. Logical properties that are determinant of grammatical at the sub-personal level need not—and usually do not—manifest themselves in assent to sentences overtly expressing those logical properties. But note that the explanatory role of logical properties does in no way depend on conscious assent to sentences having the corresponding logical forms.

our explanations off the ground.

Some structural properties—and logical consequence is one of them—should make their way into our linguistic explanations on pain of not being able to separate between argument patterns that are true accidentally and those that are true necessarily. This would be bad because the minimal properties of being true accidentally or necessarily are empirically relevant in our cases. Recall the hypothesis that the sentence *A student likes every professor* admits an inverse scope reading in virtue of its having a non-equivalent reading, a novel reading that differs from the surface scope readings. The assumption is needed in order to explain the scopal phenomena in our second case study. Can we distinguish between scopal readings locally, invoking only the absolute truth-conditions favoured by Michael Glanzberg and other philosophers of language?

Let us consider what our sentence's absolute truth conditions might be in this case. Suppose that the domain of people to which one refers (or quantifies over) contained only one professor. In that case, the two scopal readings would be truth conditionally equivalent, as can be easily checked (it can only be the case that a student likes the given professor, thus excluding the cases where there are different professors who may be the subject of students' liking in different ways). But, plainly, this would have no bearing at all on the grammatical (scopal) properties of the sentences in question. The contextually shifty truth conditions of quantified sentences cannot capture the steadfast availability of scopal readings. How the world actually is, and what the contextual truth conditions are, cannot make any difference to the grammatical status of our quantified sentence. If all that mattered for grammaticality were the truth conditions of the sentences given the actual universe of quantification (according

to how the world and utterance context actually are), then the contrast emphasised eariler in (4) would remain unexplained.

In fact, what the grammaticality of our earlier scope construction covaries with is a logical property. This is much more stable than truth conditional properties, matching the stability of the grammatical judgements about scope. Accordingly, we need to distinguish between necessary and accidental patterns of inference (and between the corresponding logical properties) in order to make the right predictions about scope in our second case study. We need to know, in the particular case of (4), if the two scopal readings are logically equivalent—not whether these readings happen to be true in the actual context. (The argument and its conclusion carry over to the other case studies as well.) To escape this conclusion, a sceptic needs to convince us that the logical properties covary with more fundamental non-logical, truth-conditional properties which can do the same explanatory job. Until then, the best hypothesis we have about these cases is that logical properties are necessary parts of the explanation. Happily, there is no supplementary cost for our linguistic theory. All that we need to require is a minimal constraint on the logico-linguistic structure: some distinction should be made between necessary and contingent properties of inferences of a certain limited pattern. We can then account for the data adopting a permissive criterion of logicality compatible with any number of logics.

4.1.2 Logical constants as evidence?

Logicality affords us many other interesting explanations of apparently disparate phenomena. What is attractive to the logicality thesis is not the wholesale logicality of language, but the potential to explain these local phenomena under the unifying

banner of logicality. To bring this out more vividly, I discuss several empirical points against the logicality thesis and argue that they are either orthogonal to that thesis or in fact reinforce it.

One source of dissatisfaction with the logicality thesis—expressed by both Glanzberg and Szabo—is that it is very hard to individuate logical constants in natural language and then collect them in a clean class, along the model given by first order logic. I cannot make good on the problem of logical constants, but I have a reason not to worry about it too much. After all, where does the expectation that the logic of natural language will be as clean and supple as first order logic? Surely, not from the recent history of linguistics and philosophy.[22]

For instance, the syntactic differences between the functional items (e.g. conjunctions, quantifiers, determiners, pronouns) in natural language and their counterparts in first order languages have been clear for quite some time now. Just by following such clear differences we get plenty of differences at the level of semantics, assuming that the logical consequence relations are individuated in part syntactically. Glanzberg and Szabo seem to count these differences as supporting their point against logicality in natural language.[23] But these obvious differences count in their favour only if we had highly unrealistic expectations about the kind of logicality to be found in natural languages.

[22] The mismatches between classical logic and natural language have been clear since, at least, Montague (1974) and Barwise and Cooper (1981). Consider for instance the type of expressions that can be conjoined with & and *and* (respectively) in first order logic and natural language. This is enough to generate different classes of logical consequence relations, as I point out below.

[23] For instance, Glanzberg (2015) speaks of a more *permissive view* of logic that would make the logicality thesis more plausible, but also less interesting. But the permissive view of logic should be our norm insofar as we treat logic as a theory grounded in the world and subject to critical scrutiny. Cf. §4.3.

While it is controversial that we have fixed criteria for separating logical constants from non-logical terms, we can still discriminate between them in practice (cf. Chierchia (2013, 52)). In our case studies, the relevant logical constants are those that produce consistency and equivalence tests over pairs of logical forms, namely conjunction and equivalence. So we seem to have no practical problem in identifying the logical constants needed to formulate the linguistic explanations that concern us.

The problem of logical constants is a problem in the philosophy of logic, which is relatively independent from the issue of whether logical properties are present in natural languages. It is still unclear, insofar as logic—rather than language—is concerned, to what extent we can demarcate between logical and non-logical terms or, in effect, whether we need to do so in order to speak of logical properties. Cf. Sagi (2014), van Benthem (1989).[24] Since the distinction between logical and non-logical terms is still controversial, it is doubtful that such a distinction can advance our understanding of the relationship between logic and language. For now we would be better off using an intuitive idea of logicality, seen as a way of classifying patterns of inference. This is what I did in discussing the logical properties behind our linguistic

[24] Here is the spirit in which van Benthem approaches logical constants.

> The professed purpose of this paper has been to analyze various strands in the intuitive notion of logicality, and then to show these at work in the widest possible setting.
> Perhaps it is only fair to add explicitly that this is an expression of a view opposed to the traditional idea of regarding logic as being primarily concerned with the study of 'logical constants' (whatever these may be). Logic, in our view, is concerned with the study of logical phenomena: and these occur all across language, not just with any distinguished group of actors.
> This view is more in line with that of Bernard Bolzano, who saw the task of logic as providing a liberal study of various mechanisms of consequence. [reference suppressed]. (van Benthem (1989, 336–37))

constructions (but see the appendix for a more concrete proposal).[25]

There are further reasons to think that the argument from logical constanthood is weak. Let us look at some purported logical constants in natural language, such as quantifiers and negation. Glanzberg (2015, 113) takes the properties of quantifiers and negation to be a problem for the advocate of the logicality thesis. He draws our attention to the contrasting behaviour of quantifiers and negation in such sentences as the following.

(10) *Distributional differences between 'every' and 'a'*

a. #Max is every friend of mine.

b. #There is every book on the table.

c. There is a book on the table.

(11) *Scopal differences between 'every' and 'a'*

a. John did not read every book. non-ambiguous

b. John did not read a book. ambiguous

The point of these examples is the mismatch between the syntactic and semantic differences between natural language quantification and quantification in first order logic, and thus a mismatch between the logical terms in natural language and logic. There is nothing that prevents us from saying, in a first order language, that e.g. every friend of mine is a person identical with Max: $\forall x(Fx \rightarrow x = m)$. Yet we cannot put this thought in words in a natural way, as (10a) testifies. Surely, Glanzberg argues,

[25] Assuming some criterion of logical constanthood (whether categorical or degree-based), our understanding of the extent to which natural language terms are logically constant will advance with out understanding of language, as the result of a wholistic comparison of different theories with respect to their epistemic virtues. There is no reason to draw a distinction between logical and non-logical terms beforehand. But once more the logicality thesis and its theoretical significance can be assessed independently of one's favourite criterion for logical constanthood.

the explanation of such distributional differences has nothing to do with logic. Nor is there any logical rationale for the contrast between how the quantifiers *every book* and *a book* interact with the negation *not*. As we can see in (11), when one reports, using normal intonation, that John did not read every book, she cannot be saying that John read no book at all (i.e., for every book, it is not the case that John read it). Her utterance of (11a) then can be read in one way and one way only. On the other hand, when one says that John did not read a book (11b), she may mean two things: that there is a book John did not read, or that in fact John read no book. So, the argument concludes, a linguistic analysis of quantification goes well beyond the bounds of logic.

What should we say on encountering such knotty behaviour coming from, of all things, quantifiers and negation? Before closing the case, I would like to suggest, we may consider (i) the aims of the logicality thesis as well as (ii) the potential linguistic explanations of the irregular behaviour of our logical constants. The first point is relatively straightforward. The existence of syntactic quirks in natural language is perfectly compatible to its having logical properties. (That the difference is syntactic in nature is argued in Beghelli and Stowell (1997).) All that these examples show is that the class of consequences in natural language will differ from the class of consequences in first order logic. But, as noted before, this was to be expected and it does not show that there are no interesting overlaps between the two classes. What's more, such overlaps may be empirically relevant, which brings us to another line of response on behalf of the advocate of the logic in natural language. In fact—and this is my second point—there really are logical explanations of the contrasts noted by Glanzberg.

Consider (11) first. What is characteristic of these sentences is that they consist of two operators, namely a quantifier and a negation. The two operators may, in principle, engage in scope interactions. Mayr and Spector (2010) argue that a double-operator sentence admits an inverse scope interpretation only when the inverse scope interpretation is asymmetrically entailed by (and thus is weaker than) the surface scope interpretation. This is enough to account for the contrast between (11a) and (11b). Note that (11a), which says that John did not read every book, produces a stronger inverse scope reading: for every book it is true that John did not read it. But the present hypothesis does not allow stronger inverse scope readings, and rightly so, as (11a) does not have any such reading. In contrast, (11b), which says that John did not read a book, does have its inverse scope reading predicted by our hypothesis and indeed this (weaker) reading is available. It says that there is a book that John did not read. This is weaker than the (surface scope) reading that John read no book.

To be sure, the hypothesis that inverse scope readings must be weaker has not been definitively established, and committing to it is a risky affair. Yet it gives us a clue as to the possibilities of a certain kind of linguistic explanation that has recently registered a number of successes. It is intriguing only just for the fact that it has the same flavour as the explanation of our second case study.[26] On the epistemic account, the present explanation counts as a good explanation because it increases the unity of a theory of linguistic scope.

Next up, we have the contrast in (10b-c). The first sentence (b) says that there is every book on the table, which is an egregious thing to say. The second sentence (c) replaces the universal quantifier with an existential quantifier, improving the resulting

[26] In fact it is meant as a strengthening of the so-called economy principle underlying our second case study. We provided a version of this principle in (6).

reading considerably to the point of normality: it now says that there is a book on the table. Barwise and Cooper (1981) point out that the contrast between the universal and existential quantifiers can be traced back to the distinction between strong determiners and weak ones. A determiner (e.g. *every*) is strong iff the generalised quantifier that it determines—which can be represented as a set of sets of the objects in the universe—is such that it includes its restrictor (e.g. the set corresponding to *man* in *every man*) in every model—viz. no matter what the universe happens to be. It is tautological that *every man* includes all the men no matter in which imaginary or actual context we happen to truthfully utter such a quantified phrase. But it is not tautological that *some men* includes all the men no matter what the universe looks like. In fact, it will seldom do. Thus, *some* is a weak determiner. Barwise and Cooper (1981) submit that only weak determiners can appear in *there*-constructions. Their hypothesis thus provides an explanation for the contrast in (10b-c).

Both asymmetrical entailment and the distinction between weak and strong determiners are notions that, on the standard model theoretic view, can be defined only in terms of global properties—in terms of how they behave across a space of models. As such both explanations of Glanzberg's data are logical, in his favourite sense. This is not to exclude other (non-model-theoretic) views of logicality that can reflect the very same properties of quantifiers. Various views of logic may flesh out this notion of logicality in different ways but to the same effect of explaining the linguistic contrasts in question. Equally important is that the logically driven accounts are not only compatible with the familiar mismatches regarding logical constants (e.g. quantifiers and other operators) exhibited by natural languages and formal ones, but in fact, the logically driven accounts *explain* these mismatches. The cases we

have just looked at are interesting because they show how logical explanation itself draws the boundaries of the logic in natural language thesis, in a theoretically fruitful way.

Before moving on, two of this section's conclusions stand out. First, the logicality thesis is not a quantitative but a qualitative thesis. It does not matter, for establishing this thesis, how much natural languages resemble a given standard of logic but rather whether there are illuminating accounts of linguistic phenomena revealing properties that are deemed logical, in a fashion that is agnostic with respect to which logic provides the right standard. Second, to find logicality in natural language we may have to dig deeper, since surface differences between logic and natural language may well reveal logicality (rather than deny it) once we take advantage of the logical explanations' unifying force. We have seen how certain differences in the behaviour of quantifier phrases and negation in natural language and their logical counterparts in first order logic are explained by appeal to the structural logical properties that they are presumed to induce. Crucially, the fact that these logical properties are logical—in the minimal sense of endowing inferences with a certain modal status, e.g. with necessity or impossibility—goes well beyond collecting data points for the logicality thesis, showing that there is some unity to the data as well as supporting a plausible metaphysical interpretation of the data: as noted, our case studies point to a distinct logical capacity that services grammar, locally.

4.2 Idealisation and stipulation

There is a more foundational argument against the logicality of language that is particularly relevant to logical explanation. Both Glanzberg and Szabo seem to imply

that this argument is problematic for the logic in natural language thesis. So let us see what the worries are, and then ask whether they detract from the explanatory value of logic, as illustrated by our case studies. Here is Glanzberg:

> [T]he logic in natural language thesis is false: we do not find logical consequence relations in our natural languages. [...][There is] a three-fold process that allows us to get from a natural language to a logical consequence relation. The process involves identifying logical constants, *abstracting away* from other features of meaning, and *idealizing away* from quirks in the structure of human languages. The relation between logic and natural language is thus less close than advocates of the logic in natural language thesis would have it, while the three-fold process allows that there is some connection. (Glanzberg (2015, 114–5), my emphases)

Szabo, on his part, voices a similar concern regarding abstraction. He argues for the limitations of abstraction by focusing on the following constructions. These constructions, as Szabo points out, are exhibiting higher and higher levels of abstraction.

(12) Alex is a father; therefore Alex is a father or Alex is a mother

(13) a is a father; therefore a is a father or a is a mother.

(14) a is F; therefore a is F or a is M

(15) p; therefore p or q

He then says:

> When it comes to explanation by abstraction, the more abstract the

> better. The fact that [(12)] is an instance of [(13)] goes some distance towards explaining its validity, the fact that it is an instance of [(14)] goes further, and the fact that it is an instance of [(15)] is as good as it gets. This is as it should be – eliminating more and more clutter we get a clearer and clearer view of how the truth of the premises guarantees the truth of the conclusion. [But …]
>
> Explanation by abstraction is just a step away from no explanation at all. It is roughly akin to saying "The validity of this inference is self-explanatory – you will see it for yourself as soon as I remove the irrelevant details that obscure your insight." It is an attractive view that the limits of logic are set by the scope of adequate explanation by abstraction. Logical validity is epistemically fundamental – to explain it all we can do is remove the dust and hope that validity will shine through. Other validities are not self-evident – to explain them we have to appeal to necessary truths. (Szabó (2012, 125))

As I will note below, there are differences of commitment and emphasis between the two sets of remarks, but, first, there are some interesting commonalities that should not escape our notice.

It is important to see Glanzberg's and Szabo's claims in the context of an argument against an empirical version of the logic in natural language thesis. Otherwise these worries may seem to miss their target. Glanzberg construes the task of getting from language to a logical consequence relation as a task in which (philosophical) logicians should engage in order to define their language-independent subject matter—a subject matter that has no explanatory connection to linguistic constructions and

thus no empirical relevance. Szabo's argument too seems to concentrate on the problem of how to demarcate logicality from non-logical things, which he presumes to be an explanatory task for the philosophical logician independently of any other empirical concern. He construes the task as aimed at explaining the validity of an argument rather than the grammaticality or felicity of a linguistic construction. None of these is what the advocate of an empirically relevant logicality thesis is aiming at, since s/he will want logic to do some explanatory work regarding language.[27] Both arguments, as exhibited by the previous passages, would grossly misconstrue the logic in natural language thesis insofar as they were aimed at it at all. But connecting them with other assumptions the two authors make, we can understand these arguments as targeting empirical versions of the logicality thesis.

The story seems to be like this. Natural language semantics has as its main job to state the truth conditions of natural language sentences and the way the truth conditions are compositionally derived. But revealing how the truth conditions take shape is all there is to the semanticists' empirical contribution. Everything else may be of theoretical interest but it is bound to go beyond the empirical domain of semantics. We can get to the extra-empirical logical stuff by abstracting and idealising away from the daily bread of the semanticist—the truth conditions. Logic is what comes out of abstraction and idealisation. What we then come to theorise about is not logical properties of linguistic expressions but rather some other kind of content (pure logical content?), following from our logical commitments. The only sense in which logical properties belong to natural language, the argument goes, is determined by termino-

[27] To claim that logical properties are explanatory—which is what the logicality thesis is based on—is not so much to assert that logical properties can be explained by something more fundamental, but that they can figure in explanations of facts that are not purely logical, namely linguistic facts.

logical fiat rather than in virtue of the logical properties' explanatory role.

This story, while no doubt coherent, fails to establish scepticism towards the logicality thesis. For it to have any bite at all, it needs to show two things. First, it needs to show that it is ad hoc to stipulate logical properties (and forms) and thus that there is no motivation for positing logicality. Second, it needs to show that the abstraction and idealisation involved in positing logicality distort reality. For if we want abstraction and idealisation to play a negative role in an argument against logicality in natural language, we have to understand them in the usual way as processes that distort reality. Abstraction, on this understanding, is an act that neglects details (about e.g. the shape of physical objects), assuming they had no impact on the subject matter of our theory, while idealisation is the act of making falsifying assumptions about our subject matter (e.g. assuming that there is no friction between two bodies).

But none of these points is borne out, first, because stipulating logical properties in natural language is often empirically driven, and, second, because if we were to believe that there are logical properties in natural language, simply saying that they are obtained by abstraction or idealisation would not impinge on these metaphysical beliefs. All scientific modelling involves abstraction and idealisation, and yet this does not automatically make it false.[28] So why suppose that we distort reality in thinking that language has logical properties?

[28] A mainstream view in the philosophy of science, put forward in different forms by Woodward (2005) and Strevens (2004), is that scientific models, even if obtained by abstraction and idealization, target *difference makers*, and are true in virtue of their capturing these difference makers. Moreover, virtually all the discussions of abstraction and idealisation, see e.g. Strevens (2004), Strevens (2008), and Elliott-Graves and Weisberg (2014), are put forward under the assumption that the abstraction and idealisation processes are compatible with the truth of theories and models that are abstracting and idealising away from the world. So, in our linguistic case, scepticism should establish that logical properties are not difference makers or, equivalently, that the kind of abstraction and idealisation involved in deploying logical properties in logical explanation distorts language (the target system) to the point of falsifying these explanations.

Logicality-sceptics give us no answer to this question. Still, scepticism about logic in natural language might seem appealing to a deflationist about logic—one that wants to deny that there are genuine logical properties, thinking (perhaps) that they are reducible to more basic non-logical properties. Certainly, if there are no logical properties, we would not expect to find them in natural language. But then, if this is all the argument from abstraction and idealisation comes down to, why argue for deflationism in such a roundabout way, focusing on natural language when (apparently) nothing about natural language can settle the matter? I do not think that scepticism regarding logicality is motivated by deflationism, but even if it were, it would go through too much linguistic trouble without achieving a better grounding, and, at any rate, it would be no better than the more committed metaphysical position assumed here.

To conclude this section, the metaphysical inconclusiveness of the sceptical argument based on abstraction and idealisation adds up to its more general empirical and epistemic deficiency, emphasised in the previous sections. There is no epistemic advantage to be had for separating logic and language in a way that undercuts substantial interactions between the two. On the contrary, there are constructive reasons to endorse the logicality thesis. Logic provides a way of systematising our knowledge of human cognition. Logical explanations endow certain logical properties with a distinctive epistemic role in understanding the grammaticality status of linguistic constructions involving, among other things, doubly quantified sentences, the binding behaviour of pronouns, and determiner phrases modified by exceptives. Since logical properties occupy a key place in linguistic explanations, they are both empirically fruitful in making sense of certain linguistic judgements and epistemically

fruitful in making sense of the data in a way that systematises our knowledge of cognition. They also have a plausible metaphysical interpretation in terms of a complex logical capacity closely interacting with grammar. While, I grant, this metaphysical claim is not evident, considerations of abstraction and idealisation from linguistic facts do not make it any more doubtful.

4.3 Diagnosing scepticism

Why do philosophers of language such as Glanzberg and Szabo come to so forcefully oppose the logicality of language despite there being good reasons—as I claimed—to be sympathetic to the logicality thesis?

Here is what I think is at the core of the disagreement about logicality in natural language. First comes the already familiar issue that Glanzberg and Szabo focus on a traditional view of semantics as concerned with the compositionally derived truth conditions. The second issue, also intimated earlier, is that our sceptics take classical logic as a touchstone of logicality and then argue that natural language falls far short of that standard. These two ideas—truth-conditional semantics and classical logic—are powerful ideas that have caught the philosophical imagination, so there may be some reason to focus on them. And if we do, we may reach sceptical conclusions about the logic in natural language.[29]

[29] I have already quoted passages that make clear that both Glanzberg and Szabo focus on compositional truth conditions as their main source of evidence (see §4.1). Regarding the nature of logic, Szabo points out that "a separation [between the form and matter of linguistic constructions] is impossible unless it is made actual by fiat. That is, except for artificial languages designed so as to have formulae factorable into logically significant form and logically insignificant matter, the separation cannot be made" (Szabó (2012, 107–8)). Glanzberg too finds it telling that "natural language does not really come pre-equipped with a distinguished class of logical constants" Glanzberg (2015, 98). But we should be impressed with the problem of making a separation between logical form and matter or between logical constants and non-logical terms only if we assume that some such distinction is already given in logic

Nevertheless, there is much more than these two ideas to the study of logic and language. Both logic and language can be approached from different directions—using different tools and with different interests in mind. So by putting logic and language together and trying to evaluate their relationship, it is only fair to take into account these different perspectives. The trouble is that scepticism about natural language does not acknowledge that logic can be treated as an empirical discipline and that linguistic explanation may be non-compositional, while these two assumptions would be good grounds for the logic in natural language thesis.

Certain contrasts between logic and language are pertinent only if we see logic as a given, say, as this or that system of rules that we learn in school. Both Glanzberg and Szabo draw some rhetorical effect out of taking something like first order logic as a model of logicality. Perhaps the exercise is worth engaging in for propaedeutic reasons. Of course there are real differences between natural and artificial languages (especially if we look at first order logic), concerning logical constants, syntactic forms, the class of logical consequences, and the presence (or conspicuous absence) of lexical entailments.[30]

Discovering these differences is useful in learning how to approach both languages, whether formal or natural. But this is telling us only a partial story about logic in natural language, one that we sort of know. If, instead, we see logic as preoccupied with properties of inference, which can in turn be approached from many perspectives, then of course logic will not present itself as pre-packaged with constants, forms and rules. This places logic together with the other sciences.[31] And if we see logic as

for good, and we have to match that achievement in our study of language.

[30] For an overview of (and a theory about) these somewhat unruly lexical entailments see Pustejovsky (1991).

[31] It is up to the logician to posit the best constants, forms and rules that account for some class of

scientific modelling of inference, the question of where logic resides in (rather than where it doesn't) becomes more tractable.

5 Conclusion

Logicality appears at the confluence of two epistemic virtues, empirical adequacy and generality (unification). In virtue of meeting these epistemic virtues, logicality is explanatory. Then, especially if we have independent grounds for believing in the existence of logical properties, it is reasonable to posit that grammar, broadly conceived as the capacity partly responsible for judgements of acceptability, is endowed with an extra-compositional mechanism that checks a limited number of logical properties. Indeed, this is precisely what a number of theorists do.[32] Although it is still premature to decide whether these hypotheses are true, they are epistemically attractive in view of their unification of novel pieces of evidence.

So what is the significance of logical explanations for the logic in natural language thesis? There are two separate messages that my discussion makes possible.

First, we may come to think that there are some serendipitous points of incidence between logical properties and linguistic properties, which amounts to admitting that there are logico-linguistic properties and that logic is part of natural language,

inferences, as it is up to the physicist to posit the relevant physical primitives and laws to account for some class of physical phenomena (cf. Williamson (2013)). So even if linguists and philosophers of language may have to take a shot at choosing logical constants, this does not distinguish them from logicians, biologists, physicists and other scientists. Not all logical constants are equal, since not all the logical constants can help us picture the world correctly. Clearly, Prior's (1960) tonk operator—which lets us tonk-join any proposition p to a necessarily true one t, and also lets us infer p from a tonk-joined pair of p and t—will not figure among the logical constants of any realistic logic or linguistic semantics.

[32] For instance, Fox (2000), Chierchia (2013), Gajewski (2009)

to a certain extent. The next questions are (i) to what extent is logic part of language (see Chierchia (2013) for an exploration of this idea) as well as (ii) what sort of logical framework is adequate for backing the ensuing empirical claims. (In fact, question (ii) has independent significance but it becomes more urgent as we discover that logic has an extensive explanatory power with respect to natural language.)

Second, even if we don't come to commit to mixed, logico-linguistic properties, we can conclude—more safely—that logic is a good way of organising a number of apparently unrelated empirical findings.

Both messages support the logic in natural language thesis as a metaphysical and epistemic thesis respectively. But in the light of the present argument the latter is more minimal and less controversial.

Even supposing that scepticism got lucky and the world is, in fact, as it prescribes (viz. devoid of logico-linguistic properties), what scepticism still misses is that were logic to enter linguistic explanation, it could do so not only via regimentation—by putting e.g. j for John and S for $swims$, thus giving us a representation of the logical form of *John swims*—but also by giving us access to deeper structural properties that are expressive enough to represent grammatical properties, and make predictions about them. It is significant that such structural logical properties play a role in explanation. In effect, acknowledging the existence of logical explanations demands some additional scrutiny. We may ask a question that the logicality sceptic cannot ask, namely, Why are logical properties so good at representing grammatical properties?

References

Alkistis Elliott-Graves, and Michael Weisberg. 2014. "Idealization." *Philosophy Compass* 9 (3): 176–85. doi:10.1111/phc3.12109.

Chierchia, Gennaro. 2013. *Logic in Grammar. Polarity, Free Choice, and Intervention*. Oxford: Oxford University Press.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge Mass.: MIT Press.

———. 2000. *New Horisons in the Study of Language and Mind*. Cambridge, UK: Cambridge University Press.

Clemens Mayr, and Benjamin Spector. 2010. "Not Too Strong! Generalizing the Scope Economy Condition." In *Proceedings of Sinn Und Bedeutung*, edited by Martin Prinzhorn Viola Schmitt and Sarah Zobel. Vol. 14.

Collins, John. 2007. "Syntax, More or Less." *Mind* 116 (464): 805–50. doi:10.1093/mind/fzm805.

Craver, Carl. 2014. "The Ontic Account of Scientific Explanation." In *Explanation in Special Sciences: The Case of Biology and History*, edited by M.I. Kaiser et al., 27–52. Springer Netherlands. doi:10.1007/978-94-007-7563-3_2.

Davidson, Donald. 2003. *The Essential Davidson*. Edited by Ernie Lepore and Kirk Ludwig. Oxford: Oxford University Press.

Evans, Gareth. 1982. *The Varieties of Reference*. Edited by John McDowell. New York, Oxford: Oxford University Press.

Filippo Beghelli, and Tim Stowell. 1997. "Distributivity and Negation: The Syntax

of Each and Every." In *Ways of Scope Taking*, edited by Anna Szabolcsi, 65:71–107. Springer Netherlands.

Fox, Danny. 2000. *Economy and Semantic Interpretation*. Cambridge Mass.: MIT Press.

Gajewski, Jon. 2009. "L-Triviality and Grammar." `http://www.gajewski.uconn.edu/papers/Logic.pdf`.

Glanzberg, Michael. 2015. "Logical Consequence and Natural Language." In *Foundations of Logical Consequence*, edited by Colin R. Caret and Ole T. Hjortland. Oxford University Press. doi:10.1093/acprof:oso/9780198715696.001.0001.

Jon Barwise, and Robin Cooper. 1981. "Generalized Quantifiers and Natural Language." *Linguistics and Philosophy* 4: 159–219.

Kitcher, Philip. 1989. "Explanatory Unification and the Causal Structure of the World." *Scientific Explanation* 13. University of Minnesota Press Minneapolis: 410–505.

Lakoff, George. 1971. "On Generative Semantics." In *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*, edited by Danny D. Steinberg and Leon A. Jakobovits, 232–96. Cambridge University Press.

Lepore, Ernest. 1983. "What Model Theoretic Semantics Cannot Do?" *Synthese* 54: 167–87.

McGinn, Colin. 2000. *Logical Properties*. Oxford: Oxford University Press.

Montague, Richard. 1974. *Formal Philosophy*. Edited by Richmond H Thomason.

New Haven Conn.: Yale University Press.

Partee, Barbara H. 2013. "On the History of the Question of Whether Natural Language Is 'Illogical'." *History and Philosophy of the Language Sciences*. https://hiphilangsci.net/2013/05/01/on-the-history-of-the-question-of-whether-natural-language-is-illogical/.

Partee, Barbara H. 2004. *Compositionality in Formal Semantics: Selected Papers by Barbara H. Partee*. Malden, MA: Blackwell Publishing Ltd.

Prior, Arthur. 1960. "The Runabout Inference-Ticket." *Analysis* 21 (2): 38–39. doi:10.1093/analys/21.2.38.

Pustejovsky, James. 1991. "The Generative Lexicon." *Computational Linguistics* 17 (4): 409–41. http://dl.acm.org/citation.cfm?id=176321.176324.

Reinhart, Tanya. 2006. *Interface Strategies: Optimal and Costly Computations*. Cambridge Mass.: MIT Press.

Resnik, Michael D. 1999. "Against Logical Realism." *History and Philosophy of Logic* 20 (3-4): 181–94. doi:10.1080/01445349950044134.

Sagi, Gil. 2014. "Formality in Logic: From Logical Terms to Semantic Constraints." *Logique & Analyse*, no. 227: 259–76. doi:10.2143/LEA.227.0.3053506.

Sher, Gila. 2011. "Is Logic in the Mind or in the World?" *Synthese* 181 (2): 353–65. doi:10.1007/s11229-010-9796-1.

Steven French, and James Ladyman. 2011. "The Ontic Account of Scientific Explanation." In *Scientific Structuralism*, edited by P. Bokulich and A. Bokulich, 25–42. 10.1007/978-94-007-7563-3_2: Springer Netherlands.

Strevens, Michael. 2004. "The Causal and Unification Approaches to Explanation

Unified—causally." *Noûs* 38 (1): 154–76. doi:10.1111/j.1468-0068.2004.00466.x.

———. 2008. *Depth: An Account of Scientific Explanation*. Harvard University Press.

Szabó, Zoltan G. 2012. "Against Logical Form." In *Donald Davidson on Truth, Meaning, and the Mental*, edited by G. Preyer. Oxford: Oxford University Press.

van Benthem, Johan. 1989. "Logical Constants Across Varying Types." *Notre Dame Journal of Formal Logic* 30 (3).

von Fintel, Kai. 1994. "Restrictions on Quantifier Domains." PhD thesis, University of Massachusetts Amherst. `http://www.semanticsarchive.net/Archive/jA3N2IwN/`.

Williamson, Timothy. 2007. *The Philosophy of Philosophy*. Oxford: Blackwell Publishing.

———. 2013. "Logic, Metalogic and Neutrality." *Erkenntnis*, 1–21. doi:10.1007/s10670-013-9474-z.

Woodward, James. 2005. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.

Appendix

Let us restate in a more precise fashion the abstract logical properties implicit in our case studies with an eye to imposing constraints on a more comprehensive logic for extra-compositional linguistic mechanisms. These logical properties are meant to

represent the functional properties that linguistic logical forms need to have in order to be part of logical explanations.

(16) For a *but*-exceptive clause to be grammatical, it needs to be *pivot-sensitive* in the sense of instantiating a property P (the property of not being a smoker in (1a), for instance) and a quantifier Q (e.g. *someone* in (1b)) with the following higher order property

$$\lambda\langle P, Q\rangle \,.\, \exists e : Q(x \in D - e) : Px \wedge \neg Q(x \in D) : Px.$$

The latter conjunct of this property is the 'alternative' that is instrumental in checking consistency at some extra-compositional stage in the derivation. (We call a property pivot-sensitive if there is a pivotal element e—the exception—which does not have P, and thus makes the second conjunct true, preserving consistency with the first conjunct.)

(17) A scopal construction is grammatical iff its operators O_1 and O_2 (any of the quantifiers or the negation) are *non-commutative* in the sense of satisfying the higher order property:

$$\lambda\langle O_1, O_2\rangle \,.\, [[O_2 S][O_1 P]] \not\equiv [[O_1 P][O_2 S]]$$

where S and P are the contextually salient restrictors of the operators, if any. The second member of the inequality/non-equivalence ($\not\equiv$) is the surface scope logical form that is used to test the first member for equivalence at some extra-compositional stage of the derivation. The equivalence test needs to fail in

order for the inverse scope logical form represented by the first member of the inequality to be licensed.

(18) A binding reading is allowed in a clause iff it instantiates a non-myopic relation relative to a linguistic context. A relation R is *non-myopic* in context C iff it satisfies this higher order property:

$$\lambda\langle R, C\rangle \,.\, \forall x \forall y CRxy \not\equiv CRxx.$$

Note that R should satisfy non-myopia in a certain linguistic context, namely in the scope of C. C stands for an operator that is part of the logical form of the sentence in which R appears, e.g. *only* in our previous example, (7b). A relation R is myopic when it does not discriminate its second argument from the first, treating it as equal to the first argument. The second member of the inequality ($\not\equiv$) is the test case against which we check for equivalence at some extra-compositional stage in the derivation. The first member of the inequality (the logical from whose binding structure is being probed) is licensed if the equivalence fails.

We can further relativize all the semantic values to a possible world parameter—the world of the context of utterance. Then $\equiv$ will be treated as mutual inclusion between sets of worlds and $\wedge$ as non-empty intersection of such sets. The ensuing (possible worlds) semantic values will form the models at which the previous properties are to be evaluated. I will leave the semantic definitions implicit and insist instead on the role of the implicit models for our properties as well as the resulting extra-compositional character of the semantic computation.

A common feature of pivot-sensitivity, commutativity, and non-myopia is that they rely on two logical forms which are part of a test. The first logical form is to be the result of the compositional mechanism, while the second logical form is extra-compositional, being brought in only for the purpose of testing the first. It is worth stressing that the second logical forms need to be extra-compositional just in their *grammatical role*; they may otherwise derived compositionally as any other logical form, but the computation they are part of is not meant to promote them as the interpretation of a given linguistic construction. What gets promoted is only the first kind of logical forms. We can consider the second logical forms as part of a reference-set, to borrow Reinhart's (2006) terminology—a set of alternatives to the (first) genuine logical forms being tested. The tests themselves are plain consistency and equivalence tests, using plain conjunction and equivalence, as we can see in the lambda-expressions expressing the three properties. Basically, the higher order properties defined in (16)-(18) are necessary conditions for the success of the said tests.

Note that the conditions that each of $\langle P, Q\rangle$, $\langle O_1, O_2\rangle$, and $\langle C, R\rangle$ need to satisfy according to the higher order properties stated in (16)-(18) become truth-evaluable statements once we supply particular values for the members of each of these pairs. The logicality involved in our fragments of natural language can be stated, equivalently, either in terms of the three higher order properties or in terms of the truth-evaluable statements constituting the defining conditions for those higher order properties. The soon-to-become truth-evaluable statements are these:

(19) $\exists e : Q(x \in D - e)Px \wedge \neg Q(x \in D) : Px$ cf. (16)

(20) $[[O_1S][O_2P] \not\equiv [[O_2P][O_1S]]$ cf. (17)

(21) $\forall x \forall y\, CRxy \not\equiv CRxx$ cf. (18)

To discuss logical explanation we can look at the properties of these three conditions, and all we say here about the three conditions can be reformulated in terms of the earlier higher order properties from which they originated.

Now for each particular choice of pairs of properties filling in the pairs of variables $\langle P, Q \rangle$, it should be the case that the resulting truth-evaluable statement is either consistent or inconsistent. For instance, as argued in the text, if we substitute $\langle [\![\text{non-smoking}]\!], [\![\text{all people}]\!] \rangle$ in the right places in (19), it will result in a consistent statement, since it is consistent to hold both (i) that all the people except a certain e are non-smokers and (ii) that it is false that all the people (including e) are non-smokers. In contrast, for the pair $\langle [\![\text{smoking}]\!], [\![\text{someone}]\!] \rangle$, (19) grammatically determines a contradiction.

The other two cases will be treated in a similar fashion. For each particular choice of object filling in the pairs of variables $\langle O_1, O_2 \rangle$ and $\langle C, R \rangle$, each of the above statements (20) and (21) will be either consistent or inconsistent. Otherwise put, the corresponding equivalences (formed by using $\equiv$ instead of $\not\equiv$ in (20) and (21)) will be either true under all interpretations (viz. no matter what the world of evaluation is) or not true under all such interpretations. To give just one example, (20) will be inconsistent for the pair $\langle [\![\text{John}]\!], [\![\text{every professor}]\!] \rangle$, which means that the corresponding equivalence between the two logical forms is logically true (viz. true at all the possible worlds) for the same pair. As we have seen, this means ungrammaticality. For grammaticality, we need non-equivalent logical forms, such as the ones determined by $\langle [\![\text{some student}]\!], [\![\text{every professor}]\!] \rangle$.

We have thus described a mechanism that checks the (in)consistency of conjunctions or equivalences formed from pairs of logical forms. (In)consistency is determined by evaluating statements across possible worlds, which are part of the models for such statements. These are the essential ingredients of extra-compositional logical explanations.

One may still quibble with the claim that models themselves are explanatory (for we can define the previous logical properties without them). Nonetheless, it is less doubtful that they are an essential part of an explanatory mechanism in the sense of helping us draw empirically and explanatory useful distinctions.

Models can be separated in two empirically relevant sets: one set that satisfies pivot-sensitivity, non-commutativity, and non-myopia, and one set that does not. The models that do satisfy these properties give us consistency of conjunctions and non-equivalences (respectively) and thus grammaticality. The others give us inconsistency and thus ungrammaticality. Models would, therefore, play a substantial explanatory role in the context of a model-theoretic account of our target phenomena. This makes it clear, I think, that models, although not part of the truth-conditional contents of *but*-exceptives, scope constructions and binding exceptions can be bona fide parts of the theory that accounts for these linguistic phenomena. These models also help us state empirically relevant logical properties.