Linguistic Intuition and Calibration

Jeffrey Maynes
St. Lawrence University

October 27th, 2012. Final Draft Copy.

The final publication is available at www.springerlink.com and in *Linguistics and Philosophy*, 35(5), 443-460.

1. Intuitions in Mentalistic Linguistics

A traditional source of evidence in linguistics is intuition, the judgments of competent speakers of a language.[1] Patterns in the intuitions motivate theoretical developments, and hypotheses are tested against those intuitions. In much of the literature on intuitions in linguistics, 'intuition' is used synonymously with 'grammaticality judgment,' which in turn is a (widely-recognized) misnomer for 'acceptability judgment.'[2] Acceptability judgments are a paradigmatic type of intuition, consider the following examples from discussions of binding theory:

(1) a. Jim_1 reasoned that he_1 got the promotion.

b. *He_1 reasoned that Jim_1 got the promotion.

In these sentences, the star (*) affixed to the front of (1b) indicates that the sentence is judged to be unacceptable (given that 'he' and 'Jim' co-refer, as indicated by the subscripts). On the basis of this judgment, the sentence is inferred to be ungrammatical (and indeed, the star is at times interpreted as if it represents this further claim). Whether

1 This paper has benefited enormously from conversation with and criticism from Steven Gross, as well as feedback from Jonathon Hricko, Nicholas Goldberg, Derek Leben, John Waterman, Karen Yan and two anonymous referees.

2 It is beyond the scope of this present essay to offer a full defense of a definition of `linguistic intuition'. Following Maynes and Gross (Manuscript), I will treat linguistic intuitions simply as judgments or reports which ascribe properties to linguistic, or language-like, items. This definition is intended to be minimally theoretic. Even if, for example, intuitions are *seemings* rather than judgments (Textor, 2009), a judgment or report of that seeming will still be required by actual practice.

linguists ought to be using intuitions as evidence for their views is a controversial issue in linguistic methodology. The practice is typically justified on the grounds that the object of study in linguistics is causal-functionally implicated in the production of the intuition (see section 2). That is, as a competent speaker, I use my competence with language to produce my intuitions, and it is this competence which is under investigation. It is therefore the etiology of the intuition which justifies using it as a piece of evidence for linguistic hypotheses.

In this essay, I consider three principal objections against the etiology argument. First, as Michael Devitt argues, it either assumes an implausible view of the mind or accepts that the processes by which intuitions are produced are mysterious. Either way, the etiology argument will not get much grip if it merely states that our linguistic knowledge or abilities are *somehow* implicated in the production of our intuitions. Second, intuition has come under methodological criticism both in philosophy and linguistics. Intuition, its critics argue, relies on the judgments of limited numbers of linguists actively engaged in the theoretical work that they are evaluating. These intuitions are subject to wide ranges of confounds and distorting influences which undermine their reliability. Third, if, as I accept here, we do not understand the processes involved in the production of intuition, we cannot unravel the causal chains in order to distinguish between those intuitions which reveal properties of the object of our investigation and those which do not.

I argue that these worries can be overcome, and the traditional practice vindicated. On a mentalistic interpretation of linguistics (see below), the etiology argument does

provide *prima facie* justification for the use of intuition. To overcome the problems enumerated above, one needs to show that linguistic intuition can be calibrated. A source of evidence is calibrated if errors and artifacts in the data can be satisfactorily resolved, such that its reliability can be assessed and its methods improved. I discuss calibration, and its relationship to linguistic intuition, in more detail in section 3. In section 4, I show that calibration is not merely *possible* for linguistic intuition, but actually empirically *plausible*. I discuss two cases drawn from literature in binding theory to demonstrate a case where linguistic intuition has been successfully calibrated. In showing that intuition can be calibrated, the three challenges can be met, and the traditional practice defended. The upshot for linguistics is the defense of a dominant method in the field. For the philosophy of linguistics, it is a clearer understanding of why linguists are justified in relying on these reports.

Mentalism, though controversial (see below), is the dominant approach in linguistics, and the account developed here takes it on board as an assumption. By a mentalistic interpretation, I mean the view that linguistic theory is (ultimately) a description or explanation of a psychological phenomena (e.g., some body of knowledge or information that speaker's possess, or some set of psychological processes that are common to all or most speakers).[3] Perhaps the most famous advocate of mentalism is Chomsky, who

3 The 'ultimately' caveat is required since a linguist may study inscriptions or utterances, and provide a theory of them, while still subscribing to mentalism. Consider Chomsky's distinction between descriptive and explanatory adequacy. A grammar is descriptively adequate if it successfuly predicts properties of sentences of that language. For a grammar to achieve explanatory adequacy, we must also be able to explain how a speaker could come to possess that grammar based only on the data available to the speaker. Chomsky, a proponent of mentalism in linguistics, holds that explanatory adequacy is the aim of linguistics. In this case, constructing a descriptively adequate grammar is a non-psychological account devised in service of an investigation which is (ultimately) of the minds of language users.

contends that the object of study is I-language, a generative procedure that produces certain types of linguistic representations, which, are used in conjunction with other psychological processes to produce inscriptions and utterances. One need not adopt the Chomskian position in order to subscribe to mentalism. Cognitive linguists, for example, reject the Chomskian claim that we possess an autonomous language faculty, but retain the mentalistic commitment (indeed arguing that this approach better reflects what we know about the mind than does the generative approach; see Evans, 2012 for a recent survey). Similarly, even if Devitt is right that there is a conceptual distinction between linguistics (the study of sentences) and psycholinguistics (the study of how the mind produces sentences), the account developed here would hold for work in psycholinguistics even if not linguistics proper.

The mentalism commitment is very general, and the precise nature of our linguistic abilities has been the subject of a great deal of work in linguistics and the philosophy of linguistics. A general account, however, suits a methodological analysis just fine. As I will argue later, the view developed here can be sustained even without a clear understanding of the nature of our linguistic knowledge or abilities and its place in the mind.

There is one bit of Chomskian terminology, however, that I will taken on board in this essay. On most mentalistic views, linguistic performances (such as utterances or inscriptions) result from a wide range of cognitive processes interacting. The linguist, however, is usually interested in a narrow range of processes or information. For Chomsky, this will be the properties of the speaker's I-language. For cognitive linguists, this might be semantic structure. I will call these properties *competence-properties*, in

contrast with *performance-properties*, which are properties of the actual data produced by speakers. While this terminology is drawn from the generative program and Chomsky's use of the competence/performance distinction, my use can be extended to competing mentalistic views. What counts as competence-property rather than a performance-property will depend upon other theoretical commitments.

Devitt rejects the mentalistic interpretation of linguistics, and develops his account of linguistic intuition in order to block an argument in favor of the claim that linguistic rules are represented in the minds of competent speakers (Devitt, 2006a, Devitt, 2006b, Devitt, 2010). He identifies the traditional defense of intuition in linguistics as the *Voice of Competence* view. On the standard version of this view, linguistic rules are represented in the minds of competent speakers, and linguistic intuitions are computationally derived from those rules.[4] These rules thus provide the informational content to the intuitions, which are guaranteed to be (at least generally) reliable since this information issues from the competence under investigation. As linguistics is such a successful science, and has achieved this success on the basis of intuitions, we can abductively infer that the rules are, in fact, represented in the mind/brain. Devitt develops his model of linguistic intuition to block this abduction.

Devitt's view on the nature of intuition in linguistics is part and parcel of his views on intuition in general, which he defines as empirical, theory-laden, central-processor judgments which are relatively quick and unreflective. As theory-laden judgments, the epistemic force of an intuition is derived entirely from the quality of the theory

4 Devitt also identifies the *non-standard* Voice of Competence view, which is that linguistic rules are embodied but not represented.

underlying it. The subject's intuition is more akin to a theorist's judgment over data than a response originating out of a particular kind of competence. If a syntactician made a quick judgment about the syntactical structure of a sentence, we might evaluate its reliability based on her knowledge or theory of syntax. Intuitions are no different. In the case of acceptability judgments, this means that the quality of the theory of what constitutes an acceptable sentence is at issue. Linguists, who apply powerful theories of grammaticality (including an understanding of the distinction between acceptability and grammaticality), provide intuitions with the most epistemic weight. A consequence of Devitt's view, therefore, is that linguists ought to rely largely on their own intuitions (as a group).

While the theories of linguists are the best available, competent speakers of a language have reliable intuitions as well. It is not because they have direct access to the true rules, as the Voice of Competence view would suggest, but rather because they have ready access to a wealth of linguistic data. They regularly produce and consume sentences. It should come as no surprise that people with so much data at hand would have good intuitions about the properties of those sentences. When presented with a sentence, the competent speaker is able to ask herself if she would say it (or what she would think if someone else said it). This datum (among others) is supplied to the central-processor for reasoning about the sentence in light of the best theories the speaker has.

It is beyond the scope of this present project to defend the mentalist assumption and take up Devitt's arguments against it. Rather, my aim is to defend a place for linguistic intuition while accepting a mentalistic interpretation of linguistic intuition which does not

require that we adopt either the Voice of Competence view or Devitt's own proposal. Such an account is valuable for three principal reasons. First, it would fit with the prevalent understanding of linguistics by linguists, without taking on board Devitt's controversial arguments against mentalism or his account of intuition. Second, it is unclear that the Voice of Competence view accurately describes the views of mentalist linguists (Ludlow, 2011), and if so, an account which better accords with linguists' own views and practices is preferable. Third, the Voice of Competence view is inadequate on its own merits. The account I offer avoids the demand (which the Voice of Competence view cannot meet) that we give a satisfactory explanation of meta-linguistic judgment which shows how the informational content of an intuition is computationally derived from the language faculty.

2. The Etiology Argument

On the mentalistic approach, there is a natural justification for the use of linguistic intuition. The object of study is some sort of linguistic knowledge or ability (hereafter linguistic competence), which will be implicated in the production of our linguistic intuitions. Since this competence is part of the causal etiology of the intuition, the intuition provides a (highly mediated and fallible) window into that underlying competence (see Fitzgerald, 2009 for a recent defense of such a view). Vision research is an instructive analogy. My judgments about what I see are highly-mediated behavior. If my mind/brain forms a 2.5D representation before I see the lake before me, I am not introspectively aware of it. I'm aware of the lake, and if asked about what I see, I could report and describe it. This likely involves a host of complex cognitive capacities,

including the application of relevant concepts. This judgment is an important piece of evidence for theorizing about vision, as one attempts to reason back to the underlying causes of visual perception based on the reports of the end product. This analogy with vision was suggested by Chomsky in *Aspects*:

> A generative grammar attempts to specify what the speaker actually knows, not what he may report about his knowledge. Similarly, a theory of visual perception would attempt to account for what a person actually sees and the mechanisms that determine this rather than statements about what he sees and why, though these statements may provide useful, in fact, compelling evidence for such a theory (Chomsky, 1965, 8-9).

Work on visual illusions provides an illustrative example. In such cases, the perceptual report offered by the subject will be non-veridical. It is still, nevertheless, evidence precisely because it is the processes which produce the judgment which are under investigation, rather than the extra-mental properties of the visual stimulus. Consider Segall, Campbell and Herskovits important work on the Müller-Lyer illusion (Segall et al., 1966). When presented with this illusion, subjects tend to judge that the two lines of equal length are actually *unequal*. Segall et al., however, found that the degree to which people suffer from the illusion varies cross-culturally. They obtained this result by presenting subjects with the illusion, and asking which line was longer. The length of the lines was then adjusted until the subject reported that they were of equal length. These judgments were not used as evidence about the lengths of the lines, but were evidence for hypotheses about the visual processing which produced the judgments.[5]

5 Similarly, patients' judgments in blindsight cases are a crucial piece of evidence for understanding this surprising occurrence (Weiskrantz, 1990). In these cases, subjects report that they cannot see a stimulus.

Illusions have also been fruitful in linguistics. Colin Phillips, Matt Wagers and Ellen Lau (Phillips, Wagers and Lau, 2011) catalog a series of linguistic illusions, including grammaticality illusions as in:

(2) More people have been to Russia than I have.

This sentence is ungrammatical because it does not include a comparison class for the expression 'more people' (it requires a class of individuals, and is not satisfied by the class denoted by 'I have'). Yet, people tend to find the sentence to be perfectly acceptable upon first glance. This data point can serve as useful evidence for a range of hypotheses about language processing. Just as in the vision cases, intuitions can serve as evidence for linguistic theses even if they do not indicate the truth of their contents. It is the fact that we judge it acceptable which is so useful in this case, not that this gives us good evidence that it is grammatical.

The structural similarities between intuitive judgments in vision research and in linguistics concern the relationship between performance-properties (e.g., what one sees, acceptability) and competence-properties (e.g., 2.5D sketch, grammaticality). While one cannot poll judgments about competence-properties directly, the competence-properties contribute to the etiology of the performance-properties. In both cases, there is a close (even if under-specified) partial causal-function relationship between the competence-properties and the performance-properties. Causal consequences are an obvious source of evidence for the study of their causes. The justification for the use of intuition, then, is

---

Nevertheless, their behavior indicates that they register it (i.e., behaving in some ways as if they were aware). The judgments of the subject would not constitute good evidence about the stimuli itself, but provide important evidence for hypotheses about visual processing, visual phenomenology and visual judgment. The subject is a reliable source of evidence solely on the grounds that his or her visual system (fitting a certain profile, e.g., suffering from a scotoma) is implicated in the judgment.

part and parcel of the general justification for using linguistic behaviors as evidence for linguistic hypotheses. On this proposal, there is no principled difference between the two. Linguistic intuitions simply are a subset of these behaviors which are readily testable and controllable.

There are, however, two principal deficiencies with the etiology argument. The illusion cases are once again helpful. When asking people to judge the length of the lines in the Müller-Lyer illusion, we *already know the lines are equal in length.* It is only given this knowledge that we are able to classify it as an illusion in the first place. The same holds in the case of a linguistic illusion. We need to already know that the sentence is ungrammatical in order to determine that subjects are under the spell of an illusion when evaluating it. Linguistic competence is only a partial cause of an intuition, and we need a means to determine when we are dealing with a case where intuition is indicative of performance properties (as in the illusion cases) rather than competence properties.

Nick Riemer presses this objection, arguing that the lack of a performance model is crippling to the reliance on intuition (Riemer, 2009, see also Bornkessel-Schlesewsky and Bornkessel-Schlesewsky, 2007). An idealized distinction between competence and performance allows linguists to delineate between the data points which the theory must accomodate and/or account for, and those which it need not. While this may be a fine methodological practice in some cases, eventually de-idealization is required (i.e., spelling out the distinction), otherwise, classification of performance errors will be *ad hoc*. To avoid this charge, linguists need to make predictions about acceptability to test their theories about grammaticality. The difficulty facing linguists, however, is that they lack a performance model. Without it, linguists cannot make any predictions about

acceptability judgments. Conflicts in the intuitive data, therefore, cannot be satisfactorily resolved. From the mere fact that our linguistic competence is part of the etiology of our judgment, it does not follow that we can distinguish its contributions from other causal influences.

The other problem noted at the outset is that intuition must also meet the methodological standards of successful science. Linguists often rely upon their own intuitions, or the intuitions of their colleagues, to motivate and test their theoretical claims. If this simply is the method of appealing to intuition, then even if intuition is causally related to the language faculty, it may provide a data source rife with so much noise that no meaningful conclusions can be drawn (see Labov, 1996, Featherston, 2007, Wasow and Arnold, 2005, Riemer, 2009, Ferreira, 2005 and Cowart, 1997). In the remainder of the paper, I supplement the causal etiology argument in order to overcome these two objections, and therefore, to justify the use of intuition in linguistics.

3. The Calibration Argument

3.1 Calibration

Intuitions have come under heavy scrutiny in philosophy, as well as in linguistics. Indeed, Robert Cummins argues that philosophical intuition finds itself in an untenable position, because it cannot be *calibrated* (Cummins, 1998). If presented with conflicting intuitions, he argues, it is impossible for us to determine which intuition (if any) is correct, and which is merely an error or artifact. The case is different for linguistic intuition, which is subject to calibration. Or so I will argue here. By showing that

linguistic intuition can be calibrated, the two objections to the etiology account will be answered.

Calibration is the process of assessing the reliability of a source of information and refining that method to render more accurate information (Weinberg et al. 2012). A natural way to do so is to find independent access to the target, and compare what we know about it to what our source of information tell us about it. Consider the case of an early astronomer, one using the first telescope capable of observing features of the Moon. The astronomer claims that the Moon has geographical features quite like the Earth, an observation which conflicts with extant theories about celestial bodies. Opponents contend that it is the telescope which is flawed, and that the observation ought to be dismissed. We could test this claim by aiming the telescope at nearby objects, ones to which we have independent access. We could then compare the reports from the telescope to other observations of the target, and verify that it produces accurate observations. If instead we found that the observations produced by using the telescope did differ from those of our other sources, the astronomer's observations could be dismissed as artifacts.

The astronomer's critic need not be satisfied by our terrestrial confirmation of the telescope's accuracy. Perhaps the atmosphere introduces anomalies into our observation that only effect celestial observations. If we have no independent access to the heavenly bodies, then we cannot calibrate the telescope using the method described above. Weinberg et al., point out that we have other means of calibration available, in particular, using our *theory of the instrument.* Our understanding of how the telescope works gives us reason to think it provides accurate reports, and we can identify artifacts by identifying

problems with the instrument itself. Calibrating a source of evidence typically involves bringing to bear both empirical calibration (through comparison with independent sources of evidence) and instrumental calibration (based upon our theory of the instrument itself).

Instrumental calibration for linguistic intuition depends upon our knowledge of how meta-linguistic judgment works. Not only have I made no assumptions about the specific cognitive underpinnings of linguistic competence here, but meta-linguistic judgment is itself not yet well understood. Devitt presses this challenge against the Voice of Competence view, a challenge one might also raise for the proposal I've offered here: 'if the explanation posits *x* as the cause of *y,* it must say enough about *the mechanism by which x* causes *y* to not leave this mysterious; a wave of the hand is not sufficient' (Devitt, 2006). I have left the mechanism by which the language faculty (or whatever it is that a linguist is studying) causes meta-linguistic judgment mysterious, and Devitt is right that merely noting that there is some story to tell is insufficient.

Instrumental calibration, however, is only one of our options for calibrating linguistic intuition. In the remainder of this essay, I argue that we can provide empirical calibration which is sufficient to justify the use of intuitions in linguistics. That is, other sources of evidence provide us with the tools to diagnose and correct errors in the dataset of linguistic intuition, and so, coupled with the *prima facie* justification offered by the cognitive etiology argument, linguists are justified in employing intuitions to test and develop their theories. The importance of calibration is a concession to Devitt's claim that we insufficiently understand the mechanisms by which intuitions are produced to hold on to the Voice of Competence view. Contra Devitt, this calibration argument

allows us to defend a traditional role for linguistic intuition without making intuitions out to be empirical judgments based upon our best linguistic theory.

3.2 Intuitings and Intuiteds

First, it is worth briefly pausing to elaborate on the epistemic structure of an appeal to intuition. It can be characterized in terms of a pair of inferences linking an *intuiting* to a linguistic thesis. The intuiting is the psychological event of having an intuition, in this case, the fact that one has made a meta-linguistic judgment that *p* (Lycan, 1988). The intuited is that *p*, it is the content of the intuiting. Consider a simple example:

(3) *Shirley book wrote.

The intuiting is the judgment that this sentence is unacceptable. The intuited is that (3) is unacceptable. A further claim is that (3) is ungrammatical. Making use of a distinction drawn by Bogen and Woodward, I argue first that the intuiting is the *data* for hypotheses about the *phenomenon* of grammaticality. In this section, I characterize the inferential structure of the connection between this data and the phenomena with a pair of inferences. In the remainder of the essay, I argue that by addressing each inference, I can resolve both problems with the etiology argument sketched above.

Bogen and Woodward's distinction between data and phenomena is designed to illuminate the space between theoretical claims about the object of investigation (the phenomena) and the observations (data) which are brought to bear as evidence.

> Data are, as we shall say, idiosyncratic to particular experimental contexts, and typically cannot occur outside of those contexts. Indeed, the factors involved in the production of data will often be so disparate and numerous,

> and the details of their interactions so complex, that it will not be possible to construct a theory that would allow us to predict their occurrence or trace in detail how they combine to produce particular items of data. Phenomena, by contrast, are not idiosyncratic to specific experimental contexts. We expect phenomena to have stable, repeatable characteristics which will be detectable by means of a variety of different procedures, which may yield quite different kinds of data (Bogen and Woodward, 1988, 317).

Consider their example of the 1973 discovery of weak neutral currents. One part of that discovery hinged on the behavior of neutrinos, which (for the most part) pass through the Earth without interacting with it. Producing usable data to record the behavior of neutrinos requires a great deal of technical sophistication. Scientists at CERN devised an experimental setup in which the neutrinos would first interact with matter to produce charged particles. These charged particles could then be detected with a bubble chamber, yielding the photographs which ultimately provided some of the key evidence for the existence of weak neutral currents. It is these photographs which constitute the data set, and it is the weak neutral currents which are the phenomena.

The crucial point for my purposes is that the aim of science is not to explain the data, but to explain the phenomena. The data are messy, and part of a complicated causal chain connecting them with the phenomena. Indeed, this chain is specific to the experimental context in which the data are elicited. In the weak neutral currents example, this chain stretches from the currents to the photographs, passing through a series of intermediate steps (interaction with matter and the bubble chamber). At issue is whether that connection suffices to provide an evidential relationship between the data and the phenomena.

This distinction maps quite readily onto the difference between the intuitings and the facts about grammaticality. An empirically adequate linguistic theory will predict that (3) is ungrammatical. What it will not do, however, is predict acceptability *judgments*. Judgments are the results of a complex, and (at present) mysterious process involving numerous performance systems in addition to linguistic competence. Not only would the predictions have to account for these performance systems, but any other factor that determined if the subject actually renders the judgment or not (including details about the context the subject finds herself in, such as, but not limited to, the experimental design). To make predictions about performance, additional assumptions are required.

The question, then, is whether this particular type of data (intuition) bears an evidential relationship to hypotheses about the phenomena. There are a pair of inferences linking intuitions to facts about grammaticality. The first is from the intuiting to the truth of the intuited content. That is, one must (nondemonstratively) infer from the fact that a subject judged a sentence to be acceptable, that the sentence is acceptable for that speaker. The second inference is from the intuited to the facts about grammaticality. Can one (again, nondemonstratively) infer that the sentence is grammatical from the fact that it is acceptable?

The two problems with the simple etiology argument can be understood in terms of this pair of inferences. Inferring from the intuiting to the intuited requires that we can determine (with sufficient confidence) that the intuiting was not the result of confounds in the experimental design, and that the intuitions of various speakers concern the same property, and are thus directly comparable. To infer from the intuiteds to linguistic theses, we need to be able to determine that the intuited is itself based on the competence

properties described and explained by linguistic theses. Therefore, if the two inferences are justifiable, then I will have successfully addressed the two principal problems with the causal account developed above.

4. The Empirical Calibration of Linguistic Intuition

4.1 Between Intuitions

One way to calibrate linguistic intuitions is to compare one subject's intuitions to the intuitions of another. Broad trends might suggest which intuitions are idiosyncratic and which ought to be pursued, while more sophisticated analysis might reveal the confounds which prompted certain judgments. This is particularly useful with the first inference, that connecting the judgment of a meta-linguistic property to the truth of that judgment. For example, can we conclude from a speaker's judgment that a sentence has a certain interpretation, that the sentence has that interpretation for the speaker? At first blush, this question seems almost tautological. If a sentence is judged acceptable, then, by virtue of judging it acceptable (accepting it), *the sentence is acceptable.* This is tempting, but not quite right.

'Acceptability' is a context-sensitive term, and if linguists rely on the subject's sense of what is acceptable, then the meaning of this predicate is liable to change across-subjects (and perhaps within-subjects as well). One might judge a sentence acceptable because it is well-formed, while another judges it acceptable because it fits with stylistic and prescriptive grammar considerations. Indeed, even linguists may operate with subtly different notions of 'acceptability' (Schütze, 1996, 44-48). Linguists have traditionally

relied upon their own intuitions, a practice Chomsky defends on the grounds that linguists are more sensitive to the sense of 'acceptability' at play in the investigation (Chomsky, 1986, 36).[6] There are obvious methodological problems, however, with relying on the intuitions of only linguists. First and foremost, it is the linguists who construct theories on these issues, and these theories are liable to infect their intuitions.[7] Second, it is a relatively small sample size, one which is not representative of the population at large. Further, these appeals typically lack experimental control on possible confounds, such as lexical variants and priming by other sentences (see Schütze, 1996 for a survey of confounds on linguistic intuitions).

These worries, a sampling of the methodological objections against linguistic intuition, can be addressed by the use of improved methods for gathering and analyzing intuition. These methods allow linguists to calibrate linguistic intuition through comparison with other intuitions. Consider, for example, work on intuitions relevant to binding theory by Frank Keller and Ash Asudeh using the technique of Magnitude Estimation (ME) (Keller and Asudeh, 2001 and Asudeh and Keller, 2002). ME is a technique classically used to measure sensations, where subjects are asked to assign values based on a comparison between a stimuli and a baseline (usually measuring intensity). What the ME method offers, Keller and Asudeh argue, is a greater degree of

6 This practice is not merely relying on the intuitions of a lone linguist. Colin Phillips points out that in practice, each linguist's intuitions are tested against those of her colleagues in conversation, presentation and publication (Phillips, 2009).

7 This claim has been explored empirically. Culbertson and Gross (2009) found that linguists and subjects with varying levels of experience with linguistics and cognitive science (including subjects who have taken courses in cognitive science, but not in linguistics) perform consistently on acceptability tasks. Interestingly, Dabrowka (2010) reports disagreement amongst linguists' intuitions based upon the theories they accept, but that the linguists' were more likely to have intuitions which run *counter* to their theories! This work is just a further example of the use of between intuition calibration.

gradience in grammaticality judgments, rather than either the standard binary methods traditionally employed (e.g., acceptable or non-acceptable, grammatical or non-grammatical) or the use of point scales.[8] Further, ME studies admit of more sophisticated statistical analyses (parametric, rather than non-parametric analyses, though see Sprouse, 2007, 2011 and Weskott and Fanselow, 2011 for a critical discussion of the virtues of ME).

Keller and Asudeh look at a particular kind of phenomenon in binding theory research, Picture NPs (PNPs). A pronoun or anaphor is in a PNP position if it is inside an NP. This commonly occurs with picture nouns, such as in the following sentences:[9]

(4) a. Hanna$_1$ found a picture of herself$_1$/her$_1$.

b. Hanna$_1$ took a picture of herself$_1$/*her$_1$.

c. Hanna$_2$ found Peter$_1$'s picture of *him$_1$/himself$_1$.

These sentences introduce complications for attempts to specify the locality constraint in Principles A and B. These principles are complementary. Anaphors must be inside the local domain (Principle A), and pronouns must be outside it (Principle B). If either of these conditions is violated, then the sentence (with the co-referential interpretation between referring expression and the pronominal expression) is ungrammatical. Yet, in sentence (4a), this complementary distribution breaks down, and both the anaphor and pronoun are acceptably interpreted as referring to Hanna. This complementary distribution resurfaces in sentences (4b) and (4c). The relevant addition seems either to be a possessor (4c), or a certain type of verb (4b).

8 The use of graded acceptability judgments does not prejudge how to interpret the results. For example, supposing that acceptability is a graded property leaves open whether grammaticality is a binary or graded property. See Featherston, 2007.

9 Note that these are the 'standard intuitions' about these cases (not Keller and Asudeh's results).

The sentences reported in (4) can be taken as the intuitions in need of calibration. While various solutions have been presented to account for these intuitions, a persistent obstacle is the lack of clarity about the data. As Daniel Büring notes in his text on binding theory, 'it is a well-acknowledged fact that the data in this area of [binding theory] are complex and hard to judge' (Büring, 2005, 51). Keller and Asudeh argue that the subtlety and sensitivity of ME studies offer a way to resolve this complex region of the data.

Keller and Asudeh tested 52 volunteer subjects, all native speakers of English and all naive of syntactic theory. The subjects were first provided with instructions on assigning magnitudes. They were given a line, and asked to assign any number to it. They were then instructed to provide numbers to new lines which were proportionate to the difference in length between the lines, and told that they could estimate acceptability in the same way. Acceptability was defined in terms of the possibility of coreference between two terms in the sentence. Subjects were provided with 48 sentences, 24 of which were four lexical variants of each type of sentence tested in the experiment, and 24 were filler sentences. Subjects were then asked to rate the acceptability of the sentences.

This work raises doubts about the intuition reported in (4a). While we began with the intuition that both the anaphor and pronoun were equally acceptable, Keller and Asudeh found that anaphors were judged more acceptable than pronouns. It was not the case, however, that pronouns were judged fully *unacceptable* either indicating that the data is more complex than our initial intuitions suggested. Perhaps their most striking finding was that anaphors can be bound from outside a PNP, even if the PNP contains a possessor. Compare (4c) to the intuitions reported (5) below:

(5) Hanna$_1$ found Peter$_2$'s picture of her$_1$/*herself$_1$.

These two pairs are also complementary. This motivates the claim that an anaphor in a PNP with a possessor must be bound by that possessor. (5) is unacceptable with 'herself,' since it cannot be bound by 'Hanna' (a Principle A violation) and must be bound by the PNP's possessor ('Peter,' with which is not coindexed). In (4c), 'himself' is bound by 'Peter,' the possessor of the PNP, consistent with the claim. These intuitions, however, are not upheld by Keller and Asudeh's results. They found that subjects judged both variations of (5) to be equally acceptable (and highly acceptable at that). Judgments on (4c), however, were similar to those found for (4a), where an anaphor was found fully acceptable, and a pronoun was only moderately acceptable. These results suggest that our initial intuitions (or those of linguists) did not accurately represent the intuitions of English speakers about these sentences, and the revised data challenges existing theoretical views about PNPs.

Crucially for my purposes, Keller and Asudeh's work is a case study in the value of calibrating intuition through comparison with other intuitions. The intuitions reported here in (4) and (5) may have been standard, but the sentences are also difficult to judge. While far from the final word, Keller and Asudeh's work raises doubts about the accuracy of these intuitions, and suggests that more subtle and fine-grained data better captures intuitions about these sentences. Experimental methods have also been used to confirm key intuitions used by linguists. Sprouse and Almeida checked every example from a linguistics textbook (*Adger's Core Syntax*) and found a high degree of consistency between the intuitions reported in the text, and those of naive respondents (Sprouse and Almeida, forthcoming; for other examples of linguists' intuitions being corroborated by

experimental work, see Cowart 1997, Phillips, 2009, den Dikken et al., 2007).[10] In each of these cases, between intuition calibration has been useful in sorting out the reliable intuitions from the less reliable.

4.2 Independent Access

Comparing intuitions is useful, but not sufficient. Uniformity is of little significance if we are uniformly mistaken in our judgments. In a study of English speakers in Philadelphia, Labov asked his subjects about the use of the positive 'anymore,' as in the following sentences:

(6) a. John is smoking a lot anymore.

b. Harry likes rock music anymore.

c. Anymore, I hate to go in town anymore.

A number of speakers reported that (6a) and (6b) are unacceptable. Yet, some of these very same speakers used sentences with the positive 'anymore' construction. Indeed, Labov cites an example of one woman, who rejected (6a) and (6b), but proceeded to use (6c) during the interview! The 12 speakers who used the positive 'anymore' despite having intuitions that it was not acceptable 'said that they had never heard it before, did not recognize it as English, thought it might mean [the opposite of its meaning], and showed the same signs of bewilderment that we get from Northern speakers outside the dialect area' (Labov, 1975, 106). If 'recognized it as English' is interpreted to mean 'recognize it as part of the dialect of English they speak,' then the judgment is clearly

10 This use of between-intuition calibration bears directly on the debate over the claim that linguists are justified in using their own intuitions in place of the intuitions of naïve speakers, whether because the linguists' intuitions are similar to those of naïve speakers, or because the linguists' intuitions are superior (as on Devitt's view, see Section 1). See also note 7 above.

false. Similarly, their judgments about the meaning are false.

This case is a challenge to the inference from the intuiting to the truth of the intuited. If the 'anymore' case generalizes, then there are a number of cases in which the intuited is false, and our having that intuition is not a reliable indicator of the truth of its content (e.g., that the sentence is acceptable in the speaker's own idiolect). Without reason to think that the result is an artifact of Labov's experimental design, we have no reason to think that improvements in design will solve this problem. Instead, the solution is one which Labov uses, calibrating intuition by checking it against another data source. Labov identifies the error by comparing this intuitive data with behavioral data (through spoken corpus and elicited speech tasks) he and his collaborators obtained which clearly indicated these subjects use the positive 'anymore.'

The use of independent sources of corroboration to calibrate linguistic intuition is also crucial to the second inferential step connecting the data (intuitions) to the phenomena under investigation (e.g., grammaticality). If presented with inconsistent intuitions, the linguist may wish to reject one as an artifact or an error. Riemer worries that, unless the linguist has a performance model, this cannot be done in a manner which is not question-begging. The linguist might simply classify intuitions as performance errors because the very theory being evaluated predicts that they could not reflect competence-properties. A full performance model is unnecessary to avoid this charge. Independent sources of evidence can provide reason to conclude that an intuition is indicating performance-properties rather than competence-properties, and coincidence of intuition and independent evidence might suggest that the competence-properties are indicated. In this section, I once again turn to a case study in binding theory to demonstrate the plausibility

of the empirical calibration of linguistic intuition.

In the preceding discussion of work in experimental syntax, I briefly presented the work of Keller and Asudeh on Picture NPs (PNPs). They found that subjects judged sentences with an anaphor in a possessed PNP position and sentences with a pronoun in that position to be equally acceptable (see sentence (5) above), a finding that runs contrary to many views in binding theory. In addition to their work (among others) clarifying the data, these predictions have also been put to the test using other methods, including behavioral tasks and eye tracking data.

Subjects in a study by Runner, Sussman and Tanenhaus were presented with three dolls, and a set of pictures on a wall (Runner et al. 2006). The subjects were also outfitted with eye-tracking equipment. The speakers were then given a three step instruction, such as:

(7) Look at Ken. Pick up Joe. Have Joe touch Harry's picture of him/himself.

The action instruction (the third sentence) contains a possessed PNP, and subjects were given either the variant with the pronoun or the variant with the anaphor. First, the behavior of the subject is observed when following the instruction (i.e., which picture they touch with the doll). The behavior of the subject provides a clear indication of how the subject has understood the PNP and the reference of the pronoun or anaphor without relying upon the processes of meta-linguistic judgment (thus providing independent access to the subject's reference assignments). Following the intuitions reported in (4c) and (5), as well as the dominant views of linguists working on binding theory, we should expect that subjects would not use Joe to touch the picture of Harry if 'him' is used, but to touch the picture of Harry if 'himself' is used. While subjects did perform as expected

with the pronoun, they violated the prediction with an anaphor just over thirty percent of the time. That is, subjects treated 'himself' as bound by a term outside the PNP, such as 'Joe.' This finding lends support to Keller and Asudeh's contention that anaphors can be bound from outside the possessed PNP, and similarly raises doubts about the complementary domains of anaphors and pronouns in these cases.[11]

The complementarity of the principles might be defended on the grounds that reference resolution proceeds in two-stages, and that the complementary distribution is present at the first stage, but not the second (and thus it does not show up in the resulting action). To test this hypothesis, Runner et al. looked at the eye movements of the subjects. Eye-movements are 'time-locked' to hearing referential expressions (Tanenhaus et al., 1995), that is, when hearing a referential expression, a subject's eyes move to observe the possible referents for the expression. In other words, the use of eye-tracking techniques allows for the experimenter to collect evidence about comprehension in real-time, and without requiring the considered judgment or behavior of the participant. By looking at the subject's eye-movements, and then inferring back to the processing of the subject at that time, Runner et al. argue that the binding constraints are not operative even at an earlier stage.[12]

This case shows the fruitfulness of behavioral and eye-tracking data in calibrating intuition. The complementarity assumption was backed by intuitions like those reported

11 This is not to say that Runner et al.'s findings align perfectly with those of Keller and Asudeh, or that corroboration between different sources of evidence is a straightforward affair. While this case usefully demonstrates the plausibility of empirical calibration through independent access to the target, many cases in actual work will be subtle, complex and require careful analysis of the differences between the causal chains connecting the data and phenomena in each case.

12 Runner et al.'s work is the final word on this issue. For other results (both challenging and backing) Runner et al.'s, see Nicol and Swinney, 1989; Badecker and Straub, 2002; and Xiang, Dillon and Phillips, 2009. I am obliged to an anonymous reviewer on this point.

in (4) and (5). Keller and Asudeh report intuitions which challenge those initial judgments. Runner et al. are able to find support for Keller and Asudeh's findings *without relying on intuition*. In these cases, linguists are interested in the interpretation the speaker assigns to the sentence, and which interpretations are and are not available to that speaker in order to reason back to the operative binding constraints. The behavioral data obtained in the first experiment provides independent access to the interpretation which the speaker settles upon, while the eye-tracking data provides independent access to the interpretations which the speaker considers at various stages in processing. Coincidence between these data sources confirms which interpretations are available, and which are not available (or at least not considered).[13] If the data sources did not coincide, then it would be an open empirical question how to resolve the conflict, in the same way that conflicting results from different experimental designs force scientists in all fields to determine whether the discrepancy falsifies the hypothesis under investigation or an auxiliary hypothesis or assumption (such as those concerning the design, or as in the eye-tracking example, differences between stages of processing). In this case, Runner et al.'s results do coincide with the intuitions reported by Keller and Asudeh, strengthening the case built upon those intuitions. That this type of calibration is not only in principle possible, but empirically plausible, shows that linguistic intuition can indeed be calibrated, and so meet the challenges mounted against the etiology argument.

The example also shows the importance and relevance of the mentalistic assumption to the calibration argument. Attempts to salvage complementarity trade on the

13 One of the advantages of intuition is that it is readily extended to cases the subject might not have considered otherwise, such as to interpretations the speaker would not naturally have considered. As noted below, this advantage of intuition provides good reason to use these other methods in conjunction with intuition, rather than as replacements for it.

assumption that these rules describe psychological processes, which are themselves subject to verification from psycholinguistic techniques. In turn, the mentalistic assumption justifies the use of psycholinguistic techniques and the application of the results to linguistic hypotheses. An opponent of the mentalistic interpretation might argue that how we *process* sentences involving pronouns and anaphors in possessed PNPs is irrelevant to the facts about the sentences themselves. While a calibration argument can, in principle, be mounted with or without the mentalistic assumption, the case for plausibility I have offered here depends upon it.[14]

The availability of other sources of evidence does not mean that intuition can be surrendered in favor of them, or that it has no contribution to make of its own. It has a low resource cost (large numbers of intuitions can be collected quickly) and can be extended to an indefinite range of sentences. Perhaps the most important use is in terms of *negative behavior.* A sentence may not be used because it *cannot* be used correctly (that is, it is unacceptable), or because it simply does not arise (perhaps the situation warranting it does not occur). Intuitions are easily manipulated in experimental contexts, providing the linguist with a means to tease apart these two options. The flip side to this point is that intuition raises questions of ecological validity which can be avoided through the use of corpus data. This, however, merely reinforces the claim that intuition should not be used exclusively, and that intuition is best used in conjunction with other sources of evidence.

---

14 This case represents just a few of the other sources of evidence brought to bear on linguistic hypotheses. Working linguists make use of a wide array of sources, including corpus analysis, developmental data, reaction time studies, neuroimaging, clinical analysis of patients with an aphasia and a wide array of behavioral tasks. Calibration can be pursued with a range of tools, some more appropriate to each context than others.

5. Conclusion

The account developed here charts a middle route between Devitt's two proposed justifications for the use of linguistic intuition. Whether the Voice of Competence view is plausible on its own merits, or the actual view of any linguists, it is taken to be justified by appeal to our theory of the instrument (meta-linguistic judgment). Since we lack an adequate theory, this approach is insufficient to defend present linguistic practice. Devitt takes the view that we should instead treat our intuitions as mere cases of theoretical judgments. The judgment that a sentence is acceptable is no different, in principle, from the judgment that one expression c-commands another. While I have not challenged this view here, it is inconsistent with the traditional interpretation of intuition within linguistics. The view I have defended here is a form of this traditional view, without requiring instrumental calibration.

The use of linguistic intuition is justified on the grounds of two related arguments. First, on mentalistic approaches to linguistics, intuitions are taken to be causally related to the knowledge or abilities which are under investigation. This gives us *prima facie* reason to use them to get at those objects of investigation in the same way that perceptual reports help us get at the cognitive underpinnings of perception. This account is insufficient, however, as it does not enable us to disentangle the complex causal chains that resulted in the intuition. These webs can be disentangled, however, by calibrating intuition through comparison with other data sources. On the basis of these other sources of evidence, the reliability of linguistic intuition can be empirically established and errors in our intuitions can be reliably diagnosed. In addition to providing *prima facie*

justification for the use of intuition, the mentalist assumption justifies the use of psycholinguistic data to corroborate and calibrate linguistic intuition. Within these mentalistic approaches, the use of linguistic intuition to evaluate linguistic hypotheses can proceed on solid ground. On the basis of intuitive data, linguists can make a series of *ceteris paribus* predictions, where the *ceteris* clause covers all of the contributions from performance factors independent of the competence under investigation (however that is understood in that tradition). Through calibration with other sources of evidence, linguists have a leg to stand on when investigating whether all things really are equal.

References

Asudeh, A. & Keller, F. (2002). Experimental Evidence for a Prediction-based Binding Theory. (In M. Andronis, C. Ball, H. Elston, & S. Neuvel (Eds.), *Papers from the 37th Annual Meeting of the Chicago Linguistic Society*, volume 1. Chicago Linguistic Society.)

Badecker, W. & Straub, K. (2002). The Processing Role of Structural Constraints on the Interpretation of Pronouns and Anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 748-769

Bogen, J. & Woodward, J. (1988). Saving the Phenomena. *Philosophical Review*, 97(3), 303–352

Bornkessel-Schlesewsky, I. & Bornkessel-Schlesewsky, M. (2007). The wolf in sheep's clothing: Against a new judgment-driven imperialism. *Theoretical Linguistics*, 33(3), 319-333

Büring, D. (2005). *Binding Theory*. (Cambridge University Press.)

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. (The MIT Press.)

Chomsky, N. (1986). *Knowledge of Language: It's Nature, Origin and Use*. (Praeger.)

Conroy, A., Takahashi, E., Lidz, J. & Phillips, C. (2009). Equal Treatment for All Antecedents: How Children Succeed With Principle B. *Linguistic Inquiry*, 40, 446–486

Cowart, W. (1997). *Experimental Syntax*. (Sage.)

Culbertson, J. & Gross, S. (2009). Are Linguists Better Subjects? *British Journal for the Philosophy of Science*, 60, 721–36

Cummins, R. (1998). Reflection on Reflective Equilibrium. (In M.R. DePaul & W. Ramsey (Eds.), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. Rowan & Littlefield Publishers.)

den Dikken, M., Bernstein, J.B., Tortora, C. & Zanuttini, R. (2007). Data and Grammar: Means and Individuals. *Theoretical Linguistics*, 33, 335–352

Devitt, M. (2006a). *Ignorance of Language*. (Oxford University Press.)

Devitt, M. (2006b). Intuitions in Linguistics. *British Journal for the Philosophy of Science*, 57, 481–513

Devitt, M. (2010). Linguistic Intuitions Revisited. *British Journal for the Philosophy of Science*, 61(4), 833–865

Evans, V. (2012). Cognitive Linguistics. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(2), 129-141

Featherston, S. (2007). Data in Generative Grammar: The stick and the carrot. *Theoretical Linguistics*, 33, 269–318

Ferreira, F. (2005). Psycholinguistics, Formal Grammars and Cognitive Science. *The Linguistic Review*, 22, 365–380

Fitzgerald, G. (2009). Linguistic Intuitions. *The British Journal for the Philosophy of Science*, 61(1), 123-160

Keller, F. & Asudeh, A. (2001). Constraints on Linguistic Coreference: Structural vs Pragmatic Factors. (In J.D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum.)

Labov, W. (1975). *What is a Linguistic Fact?* (Peter de Ridder Press.)

Labov, W. (1996). When Intuitions Fail. (In L. McNair, K. Singer, L. Dolbrin, & M. Aucon (Eds.), *Papers from the Parasession on Theory and Data in Linguistics*. Chicago Linguistic Society.)

Ludlow, P. (2011). *The Philosophy of Generative Linguistics*. (Oxford University Press.)

Lycan, W. (1988). *Judgment and Justification*. (Cambridge University Press.)

Maynes, J. & Gross, S. (Manuscript) Linguistic Intuitions.

Nicol, J. & Swinney, D. (1989). The Role of Structure in Coreference Assignment During Sentence Comprehension. *Journal of Psycholinguistic Research*, 18(1), 5-19

Phillips, C. (2009). Should We Impeach Armchair Linguists? (In S. Iwasaki (Ed.), *Japanese/Korean Linguistics*. CLSI Publications.)

Phillips, C., Wagers, M. & Lau, E. (2011). Grammatical illusions and selective fallibility in real time language comprehension. (In J. Runner (Ed.), *Experiments at the Interfaces*, volume 37 of *Syntax & Semantics*. Emerald Publications.)

Riemer, N. (2009). Grammaticality as Evidence and as Prediction in a Galilean Linguistics. *Language Sciences*, 31, 612–633

Runner, J.T., Sussman, R.S., & Tanenhaus, M.K. (2006). Processing Reflexives and Pronouns in Picture Noun Phrases. *Cognitive Science,* 30, 193-241

Schütze, C.T. (1996). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. (The University of Chicago Press.)

Segall, M., Campbell, D. & Herskovits, M.J. (1966). *The Influence of Culture on Visual Perception*. (The Bobbs-Merrill Company.)

Sprouse, J. (2007) *A Program for Experimental Syntax: Finding the Relationship Between Acceptability and Grammatical Knowledge*. Dissertation, University of Maryland.

Sprouse, J. (2011). A test of the cognitive assumptions of magnitude estimation: Commutatitivty does not hold for acceptability judgments. *Language*, 87(2), 401-407

Sprouse, J. & Almeida, D. (forthcoming). Assessing the reliability of textbook data in syntax: Adger's Core Syntax. *Journal of Linguistics*.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M. & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634

Textor, M. (2009). Devitt on the Epistemic Authority of Linguistic Intuitions. *Erkenntnis*, 71, 395–405

Wasow, T. & Arnold, J. (2005), Intuitions in linguistic argumentation. *Lingua*, 115, 1481–1496

Weinberg, J.M., Crowley, S., Gonnerman, C., Vanderwalker, I. & Swain, S. (2012). Intuition & Calibration. *Essays in Philosophy*, 13(1), 256-283

Weiskrantz, L. (1990). *Blindsight: A Case Study and Its Implications*. (Oxford University Press.)

Weskott, T. & Fanselow, G. (2011). On the Informativity of Different Measures of Linguistic Acceptability. *Language*, 87(2), 249-273

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain & Language*, 108, 40-55